1.5. Discontinuous Galerkin Methoden

1.5.1. Sobolev-Räume

Zu Beginn sollen einige wichtige Räume eingeführt werden. Funktionen in diesen Räumen besitzen wichtige Eigenschaften zur Charakterisierung von Lösungen partieller Differentialgleichungen. Es soll zunächst ein anderer Ableitungsbegriff definiert werden. Sei mit $L^{\mathrm{p}}(\Omega)$ der Lebesgue-Raum aller messbaren Funktionen auf Ω.

Definition 1.1 (Schwache Ableitung) *Definiere*

i) Sei X ein Banachraum, $S \subset \mathbb{R}^{\mathrm{n}}$, $f : S \to X$. Dann heißt für $1 \leq m \leq \infty$

$$C_{\mathrm{c}}^{\mathrm{m}}(S;X) := \{f \in C^{\mathrm{m}} \mid \mathit{supp}(f) \text{ ist kompakte Teilmenge von } S\} \tag{1.8}$$

der Raum aller m-mal stetig differenzierbarer Funktionen mit kompakten Träger.

ii) Für ein $f \in L^1_{loc}(\Omega)$, $D^\alpha_w f \in L^1_{loc}(\Omega)$ ist die schwache Ableitung

$$\int_\Omega D^\alpha_w \phi dx = (-1)^{|\alpha|} \int_\Omega f(x)\phi^{(\alpha)}(x)dx \qquad \forall \phi \in C_{\mathrm{c}}^{\mathrm{m}}(\Omega) \tag{1.9}$$

Falls also $v \in L^2(\Omega)$, kann v als Distribution in der Art

$$\forall \phi \in C_{\mathrm{c}}^{\infty}, v(\phi) = (v,\phi)_\Omega \tag{1.10}$$

interpretiert werden

Mithilfe dieses Ableitungsbegriffs lassen sich nun Sobolev-Räume definieren, in denen Lösungen schwacher Formulierungen von partiellen Differentialgleichungen leben.

Definition 1.2 (Sobolev-Räume) *Sei $1 \leq \mathrm{s} \leq \infty$, dann heißt der Raum*

$$W^{\mathrm{k,s}}(\Omega) := \{u \in L^{\mathrm{s}}(\Omega)\,; D^\alpha u \in L^{\mathrm{s}}(\Omega)\, \forall \alpha.\, |\alpha| \leq k\} \tag{1.11}$$

auf Ω mit den Normen

$$||u||^{\mathrm{s}}_{W^{\mathrm{k,s}}(\Omega)} = \sum_{|\alpha|\leq k} ||D^\alpha u||^{\mathrm{s}}_{L^{\mathrm{s}}(\Omega)}\,, \qquad 1 \leq s < \infty \tag{1.12}$$

$$||u||_{W^{\mathrm{k},\infty}(\Omega)} = \max_{(\alpha|\leq k} ||D^\alpha u||_{L^\infty(\Omega)} \tag{1.13}$$

und den Seminormen

$$|u|^{\mathrm{s}}_{W^{\mathrm{k,s}}(\Omega)} = \sum_{|\alpha|=k} ||D^\alpha u||^{\mathrm{s}}_{L^{\mathrm{s}}(\Omega)}\,, \qquad 1 \leq s < \infty \tag{1.14}$$

$$|u|_{W^{\mathrm{k},\infty}} = \max_{|\alpha|=k} ||D^\alpha||_{L^\infty(\Omega)} \tag{1.15}$$

Von Besonderer Bedeutung ist der Hilbertraum $H^{\mathrm{k}}(\Omega) = W^{\mathrm{k},2}(\Omega)$.

Für zeitabhängige Probleme lassen sich Bochner-Räume definieren. In diesem Kontext werden Funktionen in einem Sobolev-Raum zu jeden Zeitpunkt t betrachtet und das Problem nur räumlich variational formuliert. Somit ist zu jedem Zeitpunkt ein Problem $u \in H^{\mathrm{s}}(\Omega)$ zu lösen. Solche Funktionen $u(t)(x)$ leben in Bochner-Räumen.

Definition 1.3 (Bochner-Räume) *Sei X ein Banachraum und ein Funktion $u : [a, b] \to X$. Für $1 \leq s \leq \infty$ ist der Raum $L^s(a, b; X)$ der Raum der stark messbaren Funktionen (Bochner-Raum) mit den Normen*

$$||u||^{\mathrm{s}}_{L^s(a,b;X)} := \int_a^b ||u(t)||^{\mathrm{s}}_X \, dt < \infty, \qquad 1 \leq s < \infty \tag{1.16}$$

und

$$||u||^{\mathrm{s}}_{L^\infty(a,b;X)} := ess \sup_{t \in (a,b)} ||u(t)|| \tag{1.17}$$

und Seminormen

$$|u|^{\mathrm{s}}_{L^s(a,b;X)} := \int_a^b |u(t)|^{\mathrm{s}}_X \, dt < \infty, \qquad 1 \leq s < \infty \tag{1.18}$$

und

$$|u|^{\mathrm{s}}_{L^\infty(a,b;X)} := ess \sup_{t \in (a,b)} |u(t)| \tag{1.19}$$

Analog werden Sobolevräume $W^{\mathrm{k,s}}(a, b; X)$ von Funktione in X definiert.

In Lebesgue-Räumen ist eine punktweise Auswertung von Funktionen nicht sinnvoll, da dieser Äquivalenzrelation derart definiert, dass Funktionen als gleich angesehen werden können, sofern sie sich nur auf Lebesgue-Nullmengen unterscheiden. Die Zuordnung $L^2(\Omega) \ni f \mapsto f(0) \in \mathbb{R}$ ist nicht wohldefiniert [**?**]. Diese Erkenntnis ist die Motivation des Spursatzes.

Theorem 1.4 (Spursatz) *. Sei $\Omega \subset \mathbb{R}^n$ ein offenes und beschränktes Lipschitz-Gebiet. Dann haben Sobolev-Funktionen $u \in W^{1,\mathrm{p}}, 1 \leq p < \infty$ Randwerte in folgendem Sinn. Es gibt einen eindeutig bestimmten stetigen linearen Operator*

$$\gamma_0 : W^{1,\mathrm{p}} \to L^{\mathrm{p}}(\partial\Omega) \tag{1.20}$$

der auf $C^1\left(\overline{\Omega}\right)$ mit der klassischen Spur $u \mapsto u|_{\partial\Omega}$ übereinstimmt. Der Satz von Green lautet dann für Spuren:

$$\int_\Omega (\nabla v + v \nabla u) \, dx = \int_{\partial\Omega} \gamma_0^\Omega(u) \, \gamma_0^\Omega(v) \tag{1.21}$$

für $u \in W^{1,\mathrm{s}}(\Omega)$, $v \in W^{1,\mathrm{s}'}(\Omega)$ wobei $s' = \frac{s}{s-1}$.
Stetigkeit:

$$||\, u|_{\partial\Omega}||_{L^s(\partial\Omega)} = \left|\left|\gamma_0^\Omega(u)\right|\right|_{L^s(\partial\Omega)} \leq C \, ||u||_{W^{1,s}(\Omega)} \qquad u \in W^{1,\mathrm{s}}(\Omega) \tag{1.22}$$

Der Raum aller Spuren auf $\partial\Omega$ aller Funktionen $u \in H^1(\Omega)$ ist

$$H^{1/2}(\partial\Omega) := \left\{\gamma_0^\Omega u \mid u \in H^1(\Omega)\right\} \tag{1.23}$$

Definiere den Sobolev-Slobodetskii-Raum auf $\partial\Omega$ als

$$H^{\mathrm{k}-\frac{1}{2}}(\partial\Omega) = \left\{\gamma_0^\Omega \mid u \in H^{\mathrm{k}}(\Omega)\right\} \tag{1.24}$$

Ein Sobolev-Raum $H^{\mathrm{s}+\frac{1}{2}}(\Omega)$ mit fraktionalem Index entsteht dabei durch Interpolation zwischen den Räumen $H^{\mathrm{s}}(\Omega)$ und $H^{\mathrm{s}+1}(\Omega)$. Dabei ist $H^{\mathrm{s}+\frac{1}{2}}(\Omega)$ die Vervollständigung von $H^{\mathrm{s}+1}(\Omega)$ bezüglich der Norm $||v||^2_{H^{\mathrm{s}+\frac{1}{2}}(\Omega)} = \int_0^\infty t^{-2}K^2(v,t)dt$ mit dem Kernel $K(v,t)$. Dann gilt $H^{\mathrm{s}+1}(\Omega) \subset H^{\mathrm{s}+\frac{1}{2}}(\Omega) \subset H^{\mathrm{s}}(\Omega)$ und

$$\forall v \in H^{\mathrm{s}+1}(\Omega), ||v||_{H^{\mathrm{s}+\frac{1}{2}}(\Omega)} \leq C(\Omega)\, ||v||^{1/2}_{H^{\mathrm{s}+1}(\Omega)}\, ||v||^{1/2}_{H^{\mathrm{s}}(\Omega)} \tag{1.25}$$

Für den Begriff der schwachen Lösung soll beispielhaft die Poisson-Gleichung $-\Delta u = f$ in Ω, $u = g$ auf $\partial\Omega$ betrachtet werden.

Definition 1.5 (Schwache Lösung, Dirichlet-Problem) *Sei Ω ein beschränktes Lipschitz-Gebiet, $f \in L^2(\Omega)$ und $g \in L^2(\partial\Omega)$. Eine Abbildung $u : \Omega \to \mathbb{R}$ heißt schwache Lösung der Poisson-Gleichung, falls $u \in H^1(\Omega)$ mit $\gamma_0 u = g$*

$$\int_\Omega \nabla u \cdot \nabla\varphi = \int_\Omega f\varphi, \qquad \forall \varphi \in H_0^1(\Omega) \tag{1.26}$$

Es ist $H_0^1(\Omega)$ der Raum aller Funktionen $u \in H^1(\Omega)$ mit verschwindender Spur auf dem Rand. Der Lösungsbegriff efüllt die Kompatiblität:

i) Jede klassische Lösung $u \in H^1(\Omega)$ ist auch eine schwache Lösung

ii) Jede schwache Lösung mit hinreichender Regularität ist eine klassische Lösung

Analog kann die schwache Lösung für das Neumann-Problem formuliert werden. Zuletzt sollen einige wichtige Relationen erwähnt werden, die bei der Analysis der in Kapitel ?? zum Einsatz kommen.

Lemma 1.6 (Youngsche Ungleichung) *Sei $s, q \in (1, \infty)$ mit $\frac{1}{s} + \frac{1}{q} = 1$, $a, b \geq 0$. Dann ist*

$$ab \leq \frac{a^{\mathrm{s}}}{s} + \frac{b^{\mathrm{q}}}{q} \tag{1.27}$$

Im Spezialfall $s = q = 2$ und $\lambda > 0$ gilt

$$ab \leq \frac{1}{2\lambda}a^2 + \frac{\lambda}{2}b^2 \tag{1.28}$$

Lemma 1.7 *Multiplikative Spur-Ungleichung Für jedes $u \in H^{s}(K)$ für $s \geq 1$ existiert eine Konstante $C > 0$ unabhängig von h_K und u, so dass*

$$||\gamma_0 u||_{L^2(e)} \leq C\,|e|^{1/2}\,|K|^{-1/2}\left(||u||_{L^2(K)} + h_K\,||\nabla u||_{L^2(K)}\right), \qquad \forall e \subset \partial K \tag{1.29}$$

Falls $u \in \mathbb{P}_{\mathrm{N}}$, dann gilt

$$||u||_{L^2(e)} \leq C_t h_K^{-1/2}\,||u||_{L^2(K)}, \qquad \forall e \subset \partial K \tag{1.30}$$

$$||\nabla u \cdot \mathbf{n}_e||_{L^2(e)} \leq C_t h_K^{-1/2}\,||\nabla u||_{L^2(K)}, \qquad \forall e \subset \partial K \tag{1.31}$$

Lemma 1.8 *Inverse Spurungleichung Für eine Funktion $u \in \mathbb{P}_{\mathrm{N}}$ gilt*

$$\left|\left|\nabla^{\mathrm{j}} u\right|\right|_{L^2(K)} \leq C_{\mathrm{i}} h_K^{-\mathrm{j}}\,||u||_{L^2(K)} \tag{1.32}$$

1.5.2. Finite-Elemente-Räume

Der Kerngedanke der Finite-Elemente-Methoden ist die Unterteilung eines Gebiets Ω in eine kleinere Gebiete. Es soll nun also das finite Element definiert werden.

Definition 1.9 (Finite Elemente) *Sei $K \subset \mathbb{R}^{\mathrm{n}}$ eine kompakte Menge mit nichtleerem Inneren und stückweise glattem Rand, P ein endlich dimensionaler Raum von Funktionen auf K und $\mathcal{N} = \{N_1, \ldots, N_{\mathrm{k}}\}$ eine Basis für P'. Dann ist $(K, P, \mathcal{N})$ ein finites Element. Die Basis $\{\phi_1, \ldots, \phi_{\mathrm{k}}\}$ von P dual zu $\mathcal{N}$ (d.h. $N_{\mathrm{i}}(\phi_{\mathrm{j}}) = \delta_{\mathrm{i,j}}$) heißt nodale Basis zu P.*
Es werden Zerlegungen offener Mengen $\Omega \subset \mathbb{R}^{\mathrm{n}}$ betrachtet, wobei keine Kante eines Elements im Inneren eines anderen Elements liegen darf. Eine solche Zerlegung heißt dann Triangulation $\mathcal{T}_h$. Die Menge aller Kanten wird dann mit $\mathcal{F}_h$ bezeichnet

Bezeichne $h = \max_{K \in \mathcal{T}_h} h_K$, wobei h_K der Diameter des Elements K ist und ρ_K die größte zu bildende Kugel innerhalb eines Elements K ist. Mit $|K|$ und $|e|$ werden die Volumina von Elementen $K \in \mathcal{T}_h$ und Kanten $e \subset K$ bezeichnet. Es werden bezüglich der Triangulation folgende Annahmen gemacht:

i) Die Triangulation ist Form-regulär, d.h. es ist

$$\frac{h_K}{\rho_K} \leq C_R \;\forall K \in \mathcal{T}_h. \tag{1.33}$$

ii) Die Größe h_Γ erfüllt eine Äquivalenzbeziehung mit h_K, d.h.

$$C_T h_K \leq h_\Gamma \leq C_G h_K, \quad \forall K \in \mathcal{T}_h, \quad \forall \Gamma \in \mathcal{F}_h, \Gamma \subset \partial K \tag{1.34}$$

Im Speziellen Fall der Discontinuous-Finite-Elemente-Methoden, welche in Kapitel ?? noch näher beleuchtet werden, werden Lösungen betrachtet, an welche im Gegensatz zu den Lösungen in klassischen stetigen Finite-Elemente-verfahren nicht die Anforderungen der Stetigkeit an den Elementrändern gestellt werden. So werden sogenannte gebrochene Sobolev-Räume eingeführt.

Definition 1.10 (Polynom-Raum) *Sei X ein Banach-Raum. Der Raum der Polynome ist*

$$\mathbb{P}_{\mathrm{N}}(0,T;X) := \left\{ v \in C(a,b;X) ; v(t) = \sum_{\mathrm{i}=0}^{\mathrm{q}} \varphi_{\mathrm{i}} t^{\mathrm{i}}, \varphi_{\mathrm{i}} \in X, \mathrm{i} = 0,\dots,\mathrm{q}, t \in [0,T] \right\} \tag{1.35}$$

Definition 1.11 (Gebrochene Sobolev-Räume) *Für eine Zahl s wird der Sobolev-Raum*

$$H^{\mathrm{s}}(0,T;\Omega,\mathcal{T}_h) := \left\{ v \in L^2(\Omega) \;\middle|\; v|_K \in H^{\mathrm{s}}(0,T;K)\, \forall K \in \mathcal{T}_h \right\} \tag{1.36}$$

eingeführt, mit der Norm

$$||v||_{H^{\mathrm{s}}(\Omega,\mathcal{T}_h)} = \left(\sum_{K\in\mathcal{T}_h} ||v||^2_{H^{\mathrm{s}}(0,T;K)} \right)^{\frac{1}{2}} \tag{1.37}$$

und der Seminorm

$$|v|_{H^{\mathrm{s}}(\Omega,\mathcal{T}_h)} = \left(\sum_{K\in\mathcal{T}_h} |v|^2_{H^{\mathrm{s}}(0,T;K)} \right)^{\frac{1}{2}} \tag{1.38}$$

Definiere einen Unterraum von $H^{\mathrm{k}}(\Omega,\mathcal{T}_h)$, in dem approximierte Lösung gesucht wird:

$$S_{\mathrm{hp}} := \left\{ v \in L^2(\Omega) \;\middle|\; v|_K \in \mathbb{P}_{\mathrm{N}}(K)\, \forall K \in \mathcal{T}_h \right\} \subset H^{\mathrm{s}}(\Omega,\mathcal{T}_h) \tag{1.39}$$

Definition 1.12 (Referenzelement) *Sei $K, \hat{K} \in \mathbb{R}^{\mathrm{n}}$ sind affine-äquivalent, fall es eine invertierbare Abbildung $F_K : \hat{K} \to K$ gibt, so dass $F_K\left(\hat{K}\right)$ und $x = F_K(\hat{x}) = B_K\hat{x} + b_K \in K$, $\hat{x} \in \hat{K}$, wobei B_w eine nicht-singuläre Matrix mit*

$$||B|| = \sup_{0\neq x\in\mathbb{R}^{\mathrm{n}}} \frac{|Bx|}{|x|} \tag{1.40}$$

1.5.3. DG-Verfahren

Discontinuous Galerkin Verfahren bieten genau wie die klassischen finite Elemente Methoden ein breites Spektrum an Anwendungsmöglichkeiten. Eine vollumfängliche Darstellung übersteigt daher den Rahmen dieser Arbeit. Es sollen stattdessen nur einige wichtige Konzepte dargestellt werden, die hier zum Tragen kommen.
Allgemein können alle DG-Verfahren linearer Gleichungen in der Form

$$\mathcal{A}_h(u_h, v) = \mathcal{L}(v), \qquad \forall v \in S_{hp} \tag{1.41}$$

für ein $u_h \in S_{hp}$ geschrieben werden. Dabei heißt $\mathcal{A}_h : H^{\mathrm{s}}(\Omega,\mathcal{T}_h) \times \mathrm{s}(\Omega,\mathcal{T}_h) \to \mathbb{C}$ eine Sesquilinearform, da sie konjugiert linear im ersten und linear im zweiten Argument ist. Die rechte Seite wird dagegen durch eine Linearform $\mathcal{L} : H^{\mathrm{s}}(\Omega,\mathcal{T}_h) \to \mathbb{C}$ beschrieben. Bezüglich der Existenz einer Lösung zu Gleichung (1.41) existieren einige wichtige Sätze aus der Funktionalanalysis [**?**].

einige? ich zähle 1

Theorem 1.13 *(Darstellungssatz von Riesz.) Ist X ein Hilbertraum, so ist durch*

$$J(x)(y) := (y, x) \tag{1.42}$$

für alle y in X* und x in X ? Sollte hier hin

ein isometrischer konjugiert linearer Isomorphismus $J : X \to X'$ definiert.

Eine direkte Folge aus dem Rieszschen Darstellungssatz ist der Satz von lax-Milgram, der eine Aussage über die Invertierbarkeit eines zu $\mathcal{A}_h$ gehörenden linearen Operators A.

Lemma 1.14 *(Satz von Lax-Milgram.) Sei X ein Hilbertraum über einem Körper $\mathbb{K}$ und weiter sei $\mathcal{A} : X \times X \to \mathbb{K}$ sesquilinear. Es gebe Konstanten c_0 und C_0 mit $0 < c_0 \leq C_0 < \infty$, so dass für alle $x, y \in X$ gilt*

i) $|\mathcal{A}(x, y)| \leq C_0 \, ||x||_X \, ||y||_X$ *(Stetigkeit),*

ii) $\Re\mathcal{A}(x, x) \geq c_0 \, ||x||_X^2$ *(Koerzivität).*

Dann existiert genau eine Abbildung $A : X \to X$ mit

$$\mathcal{A}(y, x) = (y, Ax)_X \,, \qquad \forall x, y \in X \tag{1.43}$$

Ferner gilt: Es ist A ein linearer invertierbarer Operator mit

$$||A|| \leq C_0 \quad \text{und} \quad \left|\left|A^{-1}\right|\right| \leq \frac{1}{c_0} \tag{1.44}$$

Damit ist die Lösung des Problems (1.41) durch $u_h := A^{-1} J^{-1}(v) \, \mathcal{L}(v)$ gegeben. Für die Lösung gilt dann die Stabilitätsaussage $||u_h|| \leq \frac{1}{c_0} ||\mathcal{L}||$. Diese Aussagen unter den genannten Voraussetzungen garantieren Konvergenz und Eindeutigkeit der Lösung.
Bei Problemen höherer Ordnung kommen auch gemischte Verfahren zum Einsatz, bei denen durch die Einführung von Flüssen, also Ableitungen der gesuchten Lösung, ein System 1. Ordnung generiert wird. In diesem Fall sind dann mehrere gekoppelte Gleichungen zu lösen. Betrachte als Beispiel die stationäre Differentialgleichung zweiter Ordnung

$$Lu(\mathbf{x}) = \sum_{\mathrm{i,j}=1}^{\mathrm{n}} a_{\mathrm{j,i}} \partial_{\mathrm{i}} \partial_{\mathrm{j}} u(\mathbf{x}) + f(\mathbf{x}, u) = 0, \tag{1.45}$$

so kann durch $q_{\mathrm{j}}(\mathbf{x}) = \sum_{\mathrm{i}=1}^{\mathrm{n}} a_{\mathrm{j,i}} \partial_{\mathrm{i}} u(\mathbf{x})$, $j = 1, \ldots, n$ ein System 1. Ordnung

$$\sum_{\mathrm{i}=1}^{\mathrm{n}} \partial_{\mathrm{i}} q_{\mathrm{i}} + f(\mathbf{x}, u) = 0 \tag{1.46}$$

$$q_{\mathrm{j}} - \sum_{\mathrm{i}=1}^{\mathrm{n}} a_{\mathrm{j,i}} \partial_{\mathrm{i}} u(\mathbf{x}) = 0, \qquad j = 1, \ldots, n \tag{1.47}$$

generieren. Die Herleitung eines Finite-Elemente Verfahrens kann nun dadurch erfolgen, dass diese Gleichungen mit Testfunktionen $v \in H^{\mathrm{s}}(\Omega, \mathcal{T}_h)$ und $\mathbf{w} \in [H^{\mathrm{s}}(\Omega, \mathcal{T}_h)]^{\mathrm{n}}$ multipliziert, über ein Element K integriert und dann über alle Elemente $K \in \mathcal{T}_h$ integriert wird. Dies ergibt dann die allgemeine oder variationale Formulierung

$$\sum_{K \in \mathcal{T}_h} \int_K \left(-\sum_{\mathrm{i}=1}^{\mathrm{n}} q_{\mathrm{i}}(\mathbf{x}) \partial_{\mathrm{i}} \overline{v} + f(\mathbf{x}, u) \overline{v} \right) = -\sum_{K \in \mathcal{T}_h} \int_{\partial K} \sum_{\mathrm{i}}^{\mathrm{n}} n_{\mathrm{i}} q_{\mathrm{i}} \overline{v} \tag{1.48}$$

$$\sum_{K \in \mathcal{T}_h} \int_K \left(q_{\mathrm{j}}(\mathbf{x}) \overline{w}_{\mathrm{j}} - \sum_{\mathrm{i}=1}^{\mathrm{n}} a_{\mathrm{j,i}} \partial_{\mathrm{i}} u(\mathbf{x}) \cdot \overline{w}_{\mathrm{i}} \right) = 0, \qquad j = 1, \dots, n \tag{1.49}$$

Die Kernidee und der Unterschied zu klassischen Finite-Elemente-Methoden besteht nun darin, dass numerische Lösungen $u_h \in S_{hp}$ an den Elementkanten $e \subset \partial K$ unstetig sein dürfen. Dieser Umstand führt zu weitreichenden Konsequenzen. Über die Stetigkeitsbedingung wird üblicherweise der Fluss von einem Element in das Nachbarelement über die Oberfläche sichergestellt. Um das auszugleichen, wird ein numerischer Fluss $\hat{u}(u^1, u^2, \mathbf{n}), \hat{\mathbf{q}}(u^1, u^2, \mathbf{q}^1, \mathbf{q}^2, \mathbf{n})$ eingeführt, welcher den eigentlichen Fluss der Gleichung in der Form

$$\int_{\partial K} \mathbf{n} \cdot \mathbf{q} \overline{v} \approx \int_{\partial K} \mathbf{n} \cdot \hat{\mathbf{q}} \overline{v}, \qquad \int_{\partial K} \mathbf{n} \cdot u \overline{\mathbf{w}} \approx \int_{\partial K} \mathbf{n} \cdot \hat{u} \overline{\mathbf{w}} \tag{1.50}$$

approximieren. An einen numerischen Fluss $f(u, v, \mathbf{n})$ werden folgende Anforderungen gestellt:

i) f ist Lipschitz-stetig bezüglich u, v, d.h.

$$\left| f\left(u^1, v^1, \mathbf{n}\right) - f\left(u^2, v^2, \mathbf{n}\right) \right| \leq L_H \left(\left| u^1 - u^2 \right| + \left| v^1 - v^2 \right| \right) \tag{1.51}$$

ii) f ist konsistent, d.h. $f(u, u, \mathbf{n}) = u$

iii) f ist konservativ, d.h. $f(u, v, \mathbf{n}) = -f(v, u, -\mathbf{n})$

Für innere Kanten nehmen die Funktionen für jeweils zwei benachbarte Elemente K_1 und K_2 verschiedene Werte an. An dieser Stelle soll der Sprung und der Mittelwert auf einer Elementkante definiert werden:

$$\{u_h\} := \frac{1}{2} \left(u_h|_{\partial K_1} + u_h|_{\partial K_2} \right), \qquad [u_h] := \mathbf{n}_1 \, u_h|_{\partial K_1} + \mathbf{n}_2 \, u_h|_{\partial K_2} \tag{1.52}$$

Es ist zu beachten, dass die Zuordnung der Elemente sowohl Mittelwert als auch den Sprung bei Vertauschung invariant lassen. Für Kanten $\partial K \subset \partial \Omega$ werden Mittelwert und Sprung entsprechend definiert:

$$\{u_h\} := u_h|_{\partial K}, \qquad [u_h] := \mathbf{n} \, u_h|_{\partial K} \tag{1.53}$$

Die Wahl des numerischen Flusses ist nicht eindeutig, einige Beispiele von numerischen Flüssen aus bekannten DG-Methoden für elliptische Probleme sind in Tabelle **??** aufgelistet. Mit (1.50) kann das DG-Verfahren für ein $u \in S_{hp}$ und ein $\mathbf{q} \in [S_{hp}]^{\mathrm{n}}$

$$\sum_{K \in \mathcal{T}_h} \int_K \left(-\mathbf{q}_h \cdot \nabla \overline{v} + f(\mathbf{x}, u_h) \overline{v} \right) = -\sum_{K \in \mathcal{T}_h} \int_{\partial K} \mathbf{n} \cdot \hat{\mathbf{q}} \overline{v}, \qquad \forall v \in S_{hp} \tag{1.54}$$

$$\sum_{K \in \mathcal{T}_h} \int_K \left(\mathbf{q} \cdot \overline{\mathbf{w}} - A \nabla u \cdot \overline{\mathbf{w}} \right) = \sum_{K \in \mathcal{T}_h} \int_{\partial K} \mathbf{n} (u - \hat{u}) \overline{\mathbf{w}}, \qquad [S_{hp}]^{\mathrm{n}} \tag{1.55}$$

Tabelle 1.1.: Bekannte DG-Methoden und zugehörige Flüsse [?]. Mit $\alpha_{\mathrm{j}}(\varpi)$ und $\alpha_{\mathrm{r}}(\varphi)$ werden *penalty*-Funktionen zur Stabilisierung des Verfahrens bezeichnet.

Methode	$\hat{u}$	$\hat{q}$
Bassi-Rebay	$\{u_h\}$	$\{\mathbf{q}_h\}$
Brezzi et al.	$\{u_h\}$	$\{u_h\} - \alpha_r([u_h])$
LDG	$\{u_h\} - \beta \cdot [u_h]$	$\{\mathbf{q}_h\} + \beta \cdot [\mathbf{q}_h] - \alpha_{\mathrm{j}}([u_h])$
IP	$\{u_h\}$	$\{\nabla u_h\} - \alpha_{\mathrm{j}}([u_h])$
Bassi et al.	$\{u_h\}$	$\{\nabla u_h\} - \alpha_r([u_h])$
Baumann-Oden	$\{u_h\} + \mathbf{n} \cdot [u_h]$	$\{\nabla u_h\}$
NIPG	$\{u_h\} + \mathbf{n} \cdot [u_h]$	$\{\nabla u_h\} - \alpha_{\mathrm{j}}([u_h])$
Babuška-Zlámal	$(u_h\|_K)_{\partial K}$	$-\alpha_{\mathrm{j}}([u_h])$
Brezzi et al.	$(u_h\|_K)_{\partial K}$	$-\alpha_{\mathrm{r}}([u_h])$

angegeben werden. Wie die Ausarbeitungen in [?] zeigen, lassen sich alle lokalen Formulierungen (1.46)-(1.47) in eine primale Formulierung (1.41) überführen. Dazu werden sogenannte Lifting-Operatoren $L : H^{\mathrm{s}}(\Omega, \mathcal{T}_h) \to S_{hp}$ eingeführt, so dass

$$\mathbf{q}_h(u_h) = A(\nabla u_h - L(\hat{u} - u_h)) \tag{1.56}$$

folgt. Dies ist der durch die Einführung des numerischen Flusses $\hat{u}$ modifizierte Fluss, der im Problem für die exakte Lösung als $\mathbf{q} = A\nabla u$ definiert wurde. Diese Darstellung ist konsistent, wenn der numerische Fluss $\hat{u}$ konsistent ist, da $L(0) = 0$. Wird (1.56) nun in Gleichung (1.54) eingesetzt, kann das Verfahren in der Form (1.41) angegeben werden.

1.5.4. Fehlerabschätzungen

Ist durch Lemma 1.14 sichergestellt, dass das Verfahren (1.41) eine eindeutige Lösung besitzt, stellt sich anschließend die Frage, wie groß der Fehler zwischen der exakten Lösung $u \in H^{\mathrm{s}}(\Omega, \mathcal{T}_h)$ und der numerischen Lösung $u_h \in S_{hp}$ in Termen von der Polynomordnung N und der räumlichen Diskretisierung maximal sein darf. Zur Formalisierung dieser Fragestellung wird der lokale Projektionsoperator $\pi_{K,N} : L^2(K) \to P_N(K)$, so dass für jedes $u \in L^2(K)$

$$\int_K (\pi_{K,N} u - u)\, v = 0, \qquad \forall v \in P_N(K) \tag{1.57}$$

Auf dieser Basis kann der S_{hp}-Interpolations-Operator $\Pi_{hp} u|_K = \pi_{K,N}(u|_K)$ auf den Raum S_{hp} mit

$$\int_\Omega (\Pi_{hp} u - u)\, v = 0, \qquad \forall v \in S_{hp} \tag{1.58}$$

definiert werden.

Lemma 1.15 *Seien $N \leq 0$, $0 \leq q \leq \mu$ ganze Zahlen mit $\mu = \min(N+1, s)$, dann existiert eine Konstante C_A, so dass*

$$|\pi_{K,N} v - v|_{H^{\mathrm{q}}(K)} \leq C_A h_K^{\mu - \mathrm{q}} |v| . \qquad \forall v \in H^{\mathrm{s}}(K) \tag{1.59}$$

Eine a priori Fehlerabschätzung liefert dann die allgemeine Form

$$||e_h|| = ||u - u_h|| \leq Ch^{\mu} \tag{1.60}$$

Die genaue Bestimmung des Exponenten μ ist von dem jeweiligen Verfahren abhängig. Hierfür ist eine wichtige Eigenschaft die Konsistenz des Verfahrens, d.h. $u \in H^{\mathrm{s}}(\Omega)$ löst die Gleichung (1.41). Damit folgt für den Fehler $||e_h||$ die Galerkin-Orthogonalität

$$\mathcal{A}_h(e_h, v) = 0, \qquad \forall v \in S_{hp} \tag{1.61}$$

Die a posteriori Fehlerabschätzung soll die a priori Fehlerabschätzung durch ein numerisches Experiment bestätigen. Dazu kann für eine zu lösende Gleichung das Randwert- oder Anfangswertproblem so gewählt werden, dass dieses analytisch lösbar ist. Analog zu (1.60) wird der Fehler

$$||e_h|| = Ch^{\mathrm{EOC}} \tag{1.62}$$

mit der experimentellen Konvergenzordnung EOC betrachtet. Dieser lässt sich aus zwei numerischen Experimenten mit den Triangulationen $\mathcal{T}_{h_1}$ und $\mathcal{T}_{h_2}$ berechnen:

$$EOC = \frac{\log\left(||e_{h_1}|| \,/\, ||e_{h_2}||\right)}{\log\left(h_1/h_2\right)} \tag{1.63}$$

Eine globale Konvergenzordnung lässt sich aus einem Datensatz $(h_{\mathrm{n}}, e_{h_{\mathrm{n}}})$ durch eine Regressionsrechnung ermitteln.

2. Das DG-Verfahren

2.1. 2D-DG-Verfahren

2.1.1. Modellproblem

Sei Ω ein rechteckiges Gebiet mit den Kantenlängen L_x und L_y. Betrachtet wird die Liouville-von-Neumann-Gleichung für eine Skalare Funktion ρ und ihren vektoriellem Fluss $\mathbf{q}$

$$\rho_t - \mathrm{i}\nabla\mathbf{q} + \mathrm{i}B\rho = 0 \qquad \text{in } \Omega \times (t_0, t) \tag{2.1}$$

$$\mathbf{q} - A\nabla\rho = 0 \qquad \text{in } \Omega \times (t_0, t) \tag{2.2}$$

mit der 2×2-Matrix

$$A = \begin{pmatrix} 01 \\ 10 \end{pmatrix} \tag{2.3}$$

der skalaren Funktion $B(x,y) = V\left(x + \frac{1}{2}y\right) - V\left(x - \frac{1}{2}y\right)$. Dabei besteht $V(x,y) = V_0(x,y) + \Theta(t - t_0)\,U(x,y)$ aus dem Leitungsbandpotential V_0 und dem Hartree-Fock-Potential U. Letzteres wird zum Zeitpunkt t_0 eingeschaltet. Für ein $\mathbf{x} \in H$, ist die Matrix A beschränkt durch $-(\mathbf{x},\mathbf{x})_H \leq (A\mathbf{x},\mathbf{x})_H \leq (\mathbf{x},\mathbf{x})_H$ auf einem Hilbertraum H, entsprechend der minimalen und maximalen Eigenwerte $\lambda = \pm 1$. Alle Zeiten werden in Einheiten $\tau = \frac{V_0}{\hbar}$ und alle Längen in Einheiten $\gamma = \sqrt{\frac{mV_0}{\hbar^2}}$ gemessen.
Der Rand $\Gamma = \partial\Omega = \Gamma_D^0 \cup \Gamma_D^L \cup \Gamma_D^R$ wird in horizontale Kanten Γ_D^0 und vertikale Kanten links Γ_D^L und rechts Γ_D^R unterschieden. Es werden damit die Inflow-Randbedingungen

$$\left(\mathcal{F}^{-1}\right)^{+}(\rho) = \left(\mathcal{F}^{-1}\right)(\rho)_{k>0} = \hat{\rho}_D|_{k>0} \qquad \text{auf } \Gamma_D^L \tag{2.4}$$

$$\left(\mathcal{F}^{-1}\right)^{-}(\rho) = \left(\mathcal{F}^{-1}\right)(\rho)_{k<0} = \hat{\rho}_D|_{k<0} \qquad \text{auf } \Gamma_D^R. \tag{2.5}$$

vorgegeben. Der unitäre Operator $\mathcal{F} : L^2\left(\Gamma_D^{L/R}\right) \to L^2\left(\Gamma_D^{L/R}\right), \rho \mapsto \hat{\rho}$ überführt ρ von dem Ortsraum (x,y) in den Phasenraum (x,k). Dabei wird das Problem als über die Ausdehnung in y-Richtung als periodisch fortgesetzt interpretiert. Dies führt im diskretisierten Raum zur Anwendung der diskreten Fouriertransformation. Obwohl ρ_D im Phasenraum als Dirichlet-Randbedingung für einlaufende Wellen und als Neumann-Randbedingung für auslaufende Wellen definiert ist, wird in diesem Fall zunächst eine abstrakte Dirichlet-Randbedingung im Ortsraum gewählt. Zur Realisierung der Randbedingungen im Phasenraum kann der Operator in $\mathcal{F}^{-1} = (\mathcal{F}^{-1})^{+} + (\mathcal{F}^{-1})^{-}$ zerlegt werden. Die Operatoren $(\mathcal{F}^{-1})^{\pm}$ jeweils sind weder unitär noch invertierbar. Da $\hat{\rho}$ glatt ist, ist auch ρ hinreichend glatt, so dass das Problem die starke Lösung $\rho \in C^2\left(\overline{\Omega}\right)$ besitzt. Für den ~~vertikalen~~ Rand des Gebiets Ω soll die Randbedingung

horizontalen

$$u = 0 \qquad \text{auf } \Gamma_D^0 \tag{2.6}$$

verwendet gelten. Dies bewirkt allerdings, dass das Problem in y-Richtung stehende Zustände ausbildet und sich hin- und rücklaufende Wellen durch die Relfexion an Γ_D^0 kompensieren. Dieses Phänomen ist nicht physikalisch. Abhilfe kann zum einen durch die Einführung eines komplexen Absorbtions-Potentials geschehen. Dabei wird die Potentialfunktion B um einen Imaginärteil $-\mathrm{i}W(y)$ ergänzt. Dieser soll einen exponentiell abklingenden Abfall in Richtung im Randbereich ermöglichen. Eine mögliche Wahl ist die in [2] beschriebene Form eines quadratischen Potentials

$$W_\delta(y) = W_0 \Theta\left(|y| - (1-\delta)\frac{L_x}{2}\right)(|y| - y_0)^2 . \tag{2.7}$$

Eine andere, in Kapitel ?? beschrieben Methode ist die Einführung von Semi-Infiniten Elementen, die Randbedingungen $\lim\limits_{|y|\to\infty} \rho(x,y) = 0$ sicherstellen sollen.
Der Anfangszustand ist durch

$$\rho = \rho_{\mathrm{EQ}}, \qquad \Omega \times \{0\} \tag{2.8}$$

gegeben, wobei ρ_{EQ} das stationäre Gleichgewichts-problem

$$-\nabla\left(A\nabla\rho_{\mathrm{EQ}}\right) + B\rho_{\mathrm{EQ}} = 0, \qquad \text{in } \Omega \tag{2.9}$$

$$\rho_{\mathrm{EQ}} = \rho_{\mathrm{EQ,D}}, \qquad \text{auf } \Gamma_D \tag{2.10}$$

löst. Das Gebiet wird durch eine Finite-Elemente Diskretiesierung in Rechtecke unterteilt. Die resultierende Triangulation $\mathcal{T}_h$ soll die Bedingungen (1.61) in Kapitel (1.58) erfüllen.

2.1.2. Primale Formulierung

Das Ziel dieses Abschnittes ist es nun, eine schwache Formulierung für das Problem zu finden. Die Multiplikation mit der komplex Konjugierten der Testfunktionen $v \in H^2(\Omega, \mathcal{T}_h)$ und $\mathbf{w} \in \left[H^1(\Omega, \mathcal{T}_h)\right]^2$ für eine gegebene Triangulation $\mathcal{T}_h$ und anschließende Integration über ein Element $K \in \mathcal{T}_h$ sowie die Summation über alle Elemente K liefert

$$\sum_{K\in\mathcal{T}_h} \int_K \left(\frac{\partial\rho}{\partial t}\overline{v} + \mathrm{i}\mathbf{q}\cdot\nabla\overline{v} + \mathrm{i}B\rho\overline{v}\right) = \mathrm{i}\sum_{K\in\mathcal{T}_h}\int_{\partial K} \mathbf{n}\mathbf{q}\overline{v} \tag{2.11}$$

$$\sum_{K\in\mathcal{T}_h} \int_K \left(\mathbf{q}\overline{\mathbf{w}} - A\nabla\rho\cdot\overline{\mathbf{w}}\right) = \sum_{K\in\mathcal{T}_h}\int_{\partial K} \mathbf{n}\cdot A\overline{\mathbf{w}}\rho \tag{2.12}$$

Dabei wurde eine partielle Integration durchgeführt. Die Richtung des Normalenvektors $\mathbf{n}$ kann willkürlich gewählt werden. Nur für äußere Elemente wird die selbe Richtung wie der Normalenvektor auf $\partial\Omega$ gewählt. Die allgemeine Formulierung des Problems wird nun durch die Ersetzung von ρ und $\mathbf{q}$ durch die Flussfunktionen $\rho^* : H^1(\Omega, \mathcal{T}_h) \to H^1(\Omega, \mathcal{T}_h)$ und von

$\mathbf{q}^* : H^2(\Omega, \mathcal{T}_h) \times \left[H^1(\Omega, \mathcal{T}_h)\right]^2 \to \left[H^1(\Omega, \mathcal{T}_h)\right]^2$ in den Randintegralen ersetzt. Mit einer anschließenden weitere partiellen Integration in Gleichung (2.12) ergibt sich die allgemeine Formulierung wie folgt: Finde ein $\rho_h \in S_{hp}$, $\mathbf{q}_h \in [S_{hp}]^2$, so dass

$$\sum_{K\in\mathcal{T}_h} \int_K \left(\frac{\partial \rho_h}{\partial t}\overline{v} + \mathrm{i}\mathbf{q}_h \cdot \nabla\overline{v} + \mathrm{i}B\rho_h\overline{v}\right) = \mathrm{i} \sum_{K\in\mathcal{K}_h} \int_{\partial K} \mathbf{n}\mathrm{q}^*\overline{v} \qquad \forall v \in S_{hp} \tag{2.13}$$

$$\sum_{K\in\mathcal{T}_h} \int_K (\mathbf{q}_h\overline{\mathbf{w}} - A\nabla\rho_h \cdot \overline{\mathbf{w}}) = -\sum_{K\in\mathcal{T}_h} \int_{\partial K} A\cdot\mathbf{n}\cdot(\rho_h - \rho^*)\,\overline{\mathbf{w}} \qquad \forall \mathbf{w} \in [S_{hp}]^2 \tag{2.14}$$

wobei nach dem Galerkin-Ansatz nun ρ_h bzw. $\mathbf{q}_h$ aus dem gleichen Raum wie v bzw. $\mathbf{w}$ sein soll. Es wird der *internal penalty* Fluss

$$\mathbf{q}^* = \begin{cases} \{A\nabla\rho_h\} - \sigma[\rho_h] & \rho_h \in \Gamma_I \\ A\nabla\rho_h - \mathbf{n}\sigma(\rho_h - \rho_D) & \rho_h \in \Gamma_D \end{cases} \tag{2.15}$$

$$\rho^* = \begin{cases} \{\rho_h\} & \rho_h \in \Gamma_I \\ \rho_D & \rho_h \in \Gamma_D \end{cases} \tag{2.16}$$

gewählt. Der Strafparameter soll die Stetigkeit zwischen den einzelnen Elementen in einem schwachen Sinne sicherstellen. Bei entsprechender Wahl kann damit auch die Stabilität sichergestellt werden, wie im Abschnitt 2.1.4 gezeigt. Die Funktion $\rho_D \in H^{\frac{3}{2}}(\gamma_D)$ ist hier wieder nur über die Transformation $\mathcal{F}^{-1}\hat{\rho}_D$ definiert. Der Fluss ist sowohl konservativ, d.h. $\{\rho^*\} = \rho^*$, $\{\mathbf{q}^*\} = \mathbf{q}^*$ und $[\rho^*] = [\mathbf{q}^*] = 0$, als auch konsistent, da $\mathbf{q}^*(\rho) = A\nabla\rho|_\Gamma$ und $\rho^*(\rho) = \rho|_\Gamma$. Dies folgt unmittelbar aus der Konsistenz von Mittelwerten und dem Verschwinden von Sprüngen für ein $\rho \in H^2(\Omega)$. Die Konsistenz des Flusses führt zur Konsistenz des Verfahrens, wie in Abschnitt 2.1.3 explizit gezeigt wird.

Es soll nun die Sesquilinearform $\mathcal{B}_h$ der primalen Formulierung bestimmt werden. Mithilfe dieser können die fundamentalen Eigenschaften wie Konsistenz und der Kontinuitätscharakter untersucht werden. Für das numerische Verfahren in Finite-Elementen-Räumen können dann Eigenschaften wie Stetigkeit und Koerzivität als Voraussetzungen der Existenz und Eindeutigkeit der Lösung nach Lax-Milgram ?? nachgewiesen und entsprechende Bedingungen an den Strafparameter σ abgeleitet werden. Der Rest des Kapitels beinhaltet dann eine Fehlerabschätzung.

Mithilfe der Beziehung [**?**]

$$\sum_{K\in\mathcal{T}_h} \int_{\partial K} \mathbf{n}\cdot A\mathbf{u}\cdot v = \sum_{e\in\Gamma_I} \int_e (\{A\mathbf{u}\}[v] + \{v\}[K\mathbf{u}]) + \sum_{e\in\Gamma_D} \int_e \mathbf{n}\cdot A\mathbf{u}\cdot v \tag{2.17}$$

folgt mit der Kurzschreibweise $\int_{\Gamma_I} := \sum_{e\in\Gamma_I}\int_e$, $\int_{\gamma_D} := \sum_{e\in\Gamma_D}\int_e$ und unter Ausnutzung der Konservativität der numerischen Flusses

$$\begin{aligned} -\sum_{K\in\mathcal{T}_h} \int_{\partial K} \mathbf{n}\cdot A\overline{\mathbf{w}}\,(\rho_h - \rho^*) &= -\int_{\Gamma_I} \{A\overline{\mathbf{w}}\}[\rho_h] - \int_{\Gamma_D} \mathbf{n}\cdot A\overline{\mathbf{w}}\,(\rho_h - \rho_D) \\ &= -\sum_{K\in\mathcal{T}_h} \int_K (r(\rho_h)\,A\overline{\mathbf{w}} + s(\rho_h - \rho_{h,D})\,A\overline{\mathbf{w}}) \end{aligned} \tag{2.18}$$

Dabei wurden die Lifting-Operatoren $\tilde{r} : L^2(\Gamma_I) \to S_{hp}$, $\tilde{s} : L^2(\Gamma_D) \to S_{hp}$

$$\int_\Omega \tilde{r}(u)\,\mathbf{v} = \int_{\Gamma_I} [u]\,\{\mathbf{v}\} \tag{2.19}$$

$$\int_\Omega \tilde{s}(u)\,\mathbf{v} = \int_{\Gamma_D} u \cdot \mathbf{n}\mathbf{v} \tag{2.20}$$

eingeführt. Damit kann die Identität

$$\mathbf{q}_h = A\left(\nabla\rho_h - \tilde{r}(\rho_h) - \tilde{s}(\rho_h) + \tilde{s}(\rho_D)\right) \tag{2.21}$$

abgeleitet werden. Diese ist die numerische Approximation des Flusses $\mathbf{q} = A\nabla\rho$. Diese kann nun in Gleichung (2.13) eingesetzt werden, um das Verfahren

$$(\rho_{h,t}, v)_{\Omega,h} + \mathrm{i}\tilde{\mathcal{B}}_h(\rho_h, v) = \mathrm{i}\tilde{\mathcal{L}}(v) \tag{2.22}$$

zu definieren, mit der Sesquilinearform $\tilde{\mathcal{B}}_h : H^2(\Omega, \mathcal{T}_h) \times H^2(\Omega, \mathcal{T}_h) \to \mathbb{C}$.

$$\begin{aligned}\tilde{\mathcal{B}}_h(\rho_h, v) = &(A\nabla\rho_h, \nabla v)_{\Omega,h} - (\tilde{s}(\rho_h) + \tilde{r}(\rho_h), A\nabla v)_{\Omega,h} \\ &- (A\nabla\rho_h, \tilde{s}(v) + \tilde{r}(v))_{\Omega_h} + (B\rho_h, v)_{\Omega,h} + \int_{\Gamma_I\cup\Gamma_D} \sigma\,[u]\,[\overline{v}]\end{aligned} \tag{2.23}$$

und der Linearform $\tilde{\mathcal{L}} : H^2(\Omega, \mathcal{T}_h) \to \mathbb{C}$

$$\tilde{\mathcal{L}}(v) = \int_{\Gamma_D} (\sigma\overline{v} - \mathbf{n} \cdot A\nabla\overline{v})\,\rho_D \tag{2.24}$$

Allerdings wurden an dieser Stelle die Randbedingungen nicht berücksichtigt. Mit der Unterteilung $\Gamma_D = \Gamma_D^0 \cup \Gamma_D^L \cup \Gamma_D^R$ wird zwischen linken, rechten und vertikalen Rändern unterschieden. Mit der Zerlegung $\rho_D = \mathcal{F}^+\hat{\rho}_D + \mathcal{F}^-\hat{\rho}_D$ bzw. wird

$$\rho_D = \begin{cases}\mathcal{F}^+\hat{\rho}_D + \mathcal{F}^-\hat{\rho} & \text{auf } \Gamma_D^L \\ \mathcal{F}^+\hat{\rho} + \mathcal{F}^-\hat{\rho}_D & \text{auf } \Gamma_D^R \\ 0 & \text{auf } \Gamma_D^0\end{cases} \tag{2.25}$$

entsprechend den Randbedingungen in (2.4) und (2.5) gesetzt. Entsprechend analog ist die Zerlegung $\rho_h = \mathcal{F}^+\hat{\rho}_h + \mathcal{F}^-\hat{\rho}_h$ möglich. Dies kann aber nur unter der Voraussetzung geschehen, dass ρ_h auf $\Gamma_D^L \cup \Gamma_D^L$ stetig ist. Damit kann die Sesquilinearform

$$\begin{aligned}\mathcal{B}_h(\rho_h, v) = &(A\nabla\rho_h, \nabla v)_{\Omega,h} - (s(\rho_h), A\nabla v)_{\Omega,h} - (A\nabla\rho_h, r(v))_{\Omega_h} \\ &+ (B\rho_h, v)_{\Omega,h} + J_0(\rho_h, v)\end{aligned} \tag{2.26}$$

abgeleitet werden. Erneut wurden zwei Lifting-Operatoren eingeführt, die durch

$$\int_\Omega r(u)\mathbf{v} = \int_\Gamma \{\mathbf{v}\}\,[v] \tag{2.27}$$

$$\int_\Omega s(u)\mathbf{v} = \int_{\Gamma_I} \{\mathbf{v}\}\,[u] + \int_{\Gamma_D^L} \mathbf{n}\mathbf{v} \cdot \left(\mathcal{F}^{-1}\right)^+ \mathcal{F}(u) + \int_{\Gamma_D^R} \mathbf{n}\mathbf{v} \cdot \left(\mathcal{F}^{-1}\right)^- \mathcal{F}(u) \tag{2.28}$$

definiert sind. Mit $J_0\left(u,v\right)$ wird die *penalty*-Funktion

$$J_0\left(u,v\right) := \int_{\Gamma_I} \sigma\left[u\right]\left[\overline{v}\right] + \int_{\Gamma_D^L} \sigma\overline{v}\cdot\left(\mathcal{F}^{-1}\right)^{+}\mathcal{F}\left(u\right) + \int_{\Gamma_D^R} \sigma\overline{v}\cdot\left(\mathcal{F}^{-1}\right)^{-}\mathcal{F}\left(u\right) \tag{2.29}$$

bezeichnet. Die Sesquilinearform $\tilde{\mathcal{B}}_h$ ist symmetrisch in dem Sinne $\tilde{\mathcal{B}}_h\left(\rho_h,v\right) = \overline{\tilde{\mathcal{B}}_h\left(v,\rho_h\right)}$ und damit eine Hermitesche Form. Im Gegensatz dazu kann dies für die Sesquilinearform $\mathcal{B}_h$ nicht festgestellt werden. Die Randbedingungen zerstören offenbar die Hermitizität der allgemeinen Form. Dies steht in Übereinstimmung zu den Ergebnissen in Abschnitt ??, in der die Hermizität des Liouville-Operators bereits untersucht wurde.
Analog folgt für die neue Linearform

$$\mathcal{L}\left(v\right) = \int_{\Gamma_D^L} \left(\sigma\overline{v} - \mathbf{n}A\nabla\overline{v}\right)\left(\mathcal{F}^{-1}\right)^{+}\left(\hat{\rho_D}\right) + \int_{\Gamma_D^R} \left(\sigma\overline{v} - \mathbf{n}A\nabla\overline{v}\right)\left(\mathcal{F}^{-1}\right)^{-}\left(\hat{\rho_D}\right) \tag{2.30}$$

2.1.3. Konsistenz

Falls $\rho \in H^2\left(\Omega,\mathcal{T}_h\right)$ das Problem in Abschnitt 2.1.1 löst und die Gleichung

$$\left(\rho,v\right)_{\Omega,h} + \mathrm{i}\mathcal{B}_h\left(\rho,v\right) = \mathrm{i}\mathcal{L}\left(v\right) \qquad \forall v \in S_{hp} \tag{2.31}$$

gilt, ist die allgemeine Formulierung Konsistent zu dem Problem. Mit den Identitäten

$$\left(A\nabla\rho, r\left(v\right)\right)_{\Omega,h} = \int_{\Gamma_D} \overline{v}\cdot\mathbf{n}A\nabla\rho \tag{2.32}$$

$$\left(s\left(\rho\right), A\nabla v\right)_{\Omega,h} = \int_{\Gamma_D^L} \mathbf{n}A\nabla\overline{v}\cdot\left(\mathcal{F}^{-1}\right)^{+}\mathcal{F}\left(\rho\right) + \int_{\Gamma_D^R} \mathbf{n}A\nabla\overline{v}\cdot\left(\mathcal{F}^{-1}\right)^{-}\mathcal{F}\left(\rho\right) \tag{2.33}$$

$$J_0\left(\rho,v\right) = \int_{\Gamma_D^L} \sigma\overline{v}\cdot\left(\mathcal{F}^{-1}\right)^{+}\mathcal{F}\left(\rho\right) + \int_{\Gamma_D^R} \sigma\overline{v}\cdot\left(\mathcal{F}^{-1}\right)^{-}\mathcal{F}\left(\rho\right) \tag{2.34}$$

sowie (2.2) folgt sofort

$$\begin{aligned}
&\left(\rho_t,v\right)_{\Omega,h} + \mathrm{i}\mathcal{B}_h\left(\rho,v\right) - \mathrm{i}\mathcal{L}\left(v\right)\\
&= \left(\rho_t,v\right)_{\Omega,h} + \mathrm{i}\left(A\nabla\rho,\nabla v\right)_{\Omega,h} - \mathrm{i}\int_{\Gamma_D} \overline{v}\cdot\mathbf{n}\cdot A\nabla\rho\mathrm{i}\left(B\rho,v\right)_{\Omega,h}\\
&= \left(\rho_t,v\right)_{\Omega,h} + \mathrm{i}\left(A\nabla\rho,\nabla v\right)_{\Omega,h} - \mathrm{i}\int_{\Gamma_D\cup\Gamma_I} \overline{v}\cdot\mathbf{n}\cdot\mathbf{q} + \mathrm{i}\left(B\rho,v\right)_{\Omega,h} = 0
\end{aligned} \tag{2.35}$$

Im letzten Schritt wurde $\int_{\Gamma_I}\overline{v}\cdot\mathbf{n}\cdot\mathbf{q} = 0$ ausgenutzt. Auch an dem Verfahren (2.4) - (2.5) lässt sich aufgrund der Konsistenz des verwendeten numerischen Flusses sofort Konsistenz ablesen.

2.1.4. Existenz und Stabilität

In diesem Abschnitt sollen die Stetigkeit und Koerzivität der Sesquilinearform $\mathcal{B}_h$ untersucht werden. Für die Betrachtung der Koerzivität wird die Sesquilinearform

$$\begin{aligned}
\mathcal{B}_h\left(v,v\right) &= \left(A\nabla v,\nabla v\right)_{\Omega,h} - \left(s(v),A\nabla v\right)_{\Omega,h} - \left(A\nabla v, r(v)\right)_{\Omega_h} + \left(Bv,v\right)_{\Omega,h} + J_0\left(v,v\right)\\
&= \left(A\nabla v,\nabla v\right)_{\Omega,h} + \left(Bv,v\right)_{\Omega,h} + J_0\left(v,v\right) - \int_{\Gamma_I}\left(\left\{A\nabla\overline{v}\right\}\left[v\right] + \left\{A\nabla v\right\}\left[\overline{v}\right]\right)\\
&- \int_{\Gamma_D}\left\{A\nabla v\right\}\left[\overline{v}\right] - \int_{\Gamma_D^L}\mathbf{n}A\nabla\overline{v}\left(\mathcal{F}^{-1}\right)^{+}\mathcal{F}\left(v\right) - \int_{\Gamma_D^R}\mathbf{n}A\nabla\overline{v}\left(\mathcal{F}^{-1}\right)^{-}\mathcal{F}\left(v\right)
\end{aligned} \tag{2.36}$$

Betrachtet. Da $\mathcal{F}^{-1}$ unitär ist, ist er auch beschränkt, und da $\mathcal{F}^{-1} = \left(\mathcal{F}^{-1}\right)^{+} + \left(\mathcal{F}^{-1}\right)^{-}$, sind auch $\left(\mathcal{F}^{-1}\right)^{\pm}$ beschränkt. Außerdem ist $\left|\left|\left(\mathcal{F}^{-1}\right)^{+}\right|\right| = \left|\left|\left(\mathcal{F}^{-1}\right)^{-}\right|\right|$. Somit kann zuerst die Abschätzung

$$\begin{aligned}&\int_{\Gamma_D^L} \Re\left(\mathbf{n}A\nabla\overline{v}\left(\mathcal{F}^{-1}\right)^{+}\mathcal{F}\left(v\right)\right) + \int_{\Gamma_D^R} \Re\left(\mathbf{n}A\nabla\overline{v}\left(\mathcal{F}^{-1}\right)^{-}\mathcal{F}\left(v\right)\right)\\ &\leq \left|\left|\left(\mathcal{F}^{-1}\right)^{+}\right|\right| \left|\left|\mathcal{F}^{-1}\right|\right| \int_{\Gamma_D} \left(\Re v \Re \mathbf{n}A\nabla v + \Im v \Im \mathbf{n}A\nabla v\right)\end{aligned} \tag{2.37}$$

mithilfe der Relation

$$\Re\left(\overline{u}v\right) = \Re u \Re v + \Im u \Im v \tag{2.38}$$

durchführen. Ebenso lässt sich der Oberflächenanteil der Inneren Kanten in

$$\begin{aligned}\int_{\Gamma_I} \left([v]\left\{A\nabla\overline{v}\right\} + [\overline{v}]\left\{A\nabla v\right\}\right) &= \int_{\Gamma_I} \Re\left([v]\left\{A\nabla\overline{v}\right\}\right)\\ &= \int_{\Gamma_I} \left(\Re\left[v\right]\Re\left\{A\nabla v\right\} + \Im\left[v\right]\Im\left\{A\nabla v\right\}\right)\end{aligned} \tag{2.39}$$

zerlegen. Nun können beiden Oberflächenanteile mit der Konstanten
$C_F = \max\left(1, \left|\left|\left(\mathcal{F}^{-1}\right)^{+}\right|\right| \left|\left|\mathcal{F}^{-1}\right|\right|\right)$ zusammengeführt werden. Da $v \in S_{hp}$ und S_{hp} ein endlichdimensionaler Raum ist, kann hier die in Abschnitt 1.5.1 beschriebene Äquivalenz der L^2-Normen auf $K \in \mathcal{T}_h$ ausgenutzt werden. Es folgt also zunächst die Abschätzung für die Realteile

$$\int_{\Gamma_I \cup \Gamma_D} \Re\left[v\right]\Re\left\{A\nabla v\right\} \leq \epsilon \sum_{K \in \mathcal{T}_h} \left|\left|\Re\nabla v\right|\right|^2_{L^2(E)} + \frac{4C^2(N)}{\epsilon} \sum_{e \in \Gamma_h \cup \Gamma_D} \left|\left|\Re\left[v\right]\right|\right|^2_{L^2(e)} \tag{2.40}$$

Es wurde die multiplikative Spur-Ungleichung 1.7, die inverse Spurungleichung 1.8, die youngsche Ungleichung 1.6 sowie die Beschränktheit der Matrix A genutzt. Analog kann mit dem Imaginärteilen verfahren werden. Genauso kann die penalty-Funktion nach unten abgeschätzt werden:

$$J_0\left(v,v\right) \geq \frac{1}{C_F}\int_\Gamma \sigma\left|[v]\right|^2 \tag{2.41}$$

Mithilfe der Umschreibung des Skalarprodukts

$$\left(A\nabla v, \nabla v\right)_{\Omega,h} = \left|\left|\nabla\left(\beta\cdot v\right)\right|\right|^2_{\Omega,h} - \sum_{K\in\mathcal{T}_h}\left|\left|\nabla v\right|\right|^2_{L^2(K)} \tag{2.42}$$

mit $\beta = (1,1)^T$ folgt für die Sesquilinearform

$$\begin{aligned}\mathcal{B}_h(v,v) \geq{}& \left|\left|\nabla\left(\beta\cdot v\right)\right|\right|^2_{\Omega,h} - \left(C_F\epsilon + 1\right)\sum_{K\in\mathcal{T}_h}\left|\left|\nabla v\right|\right|^2_{L^2(K)}\\ &+ h^{-1}\sum_{e\in\Gamma_I\cup\Gamma_D}\left(\frac{\tilde{\sigma}}{C_F} - C_F\frac{4C^2(N)}{\epsilon}\right)\left|\left|\left[v\right]\right|\right|^2_{L^2(e)} + \left(Bv,v\right)_{\Omega,h}\end{aligned} \tag{2.43}$$

unter der Voraussetzung , dass o.B.d.A $h = \min(h_{E_1}, h_{E_2}) < 1$. Es wurde $\tilde{\sigma} = h\sigma$ gewählt. Aufgrund der Beschränktheit $-B_0 \leq B(x,y) \leq B_1$ lässt sich die Abschätzung

$$\left(B,v,v\right)_{\Omega,h} \geq -B_0\left|\left|v\right|\right|^2_{L^2(K)} \tag{2.44}$$

ableiten. Es bleibt die Norm über den Sprung abzuschätzen. Dazu wird wieder die multiplikative Spurungleichung sowie die Youngsche Ungleichung benutzt:

$$
\begin{aligned}
&\sum_{\Gamma_h \cup \Gamma_D} \left(\frac{\tilde{\sigma}}{C_F} - C_F \frac{4C^2(N)}{\epsilon} \right) || \, [v] \, ||^2_{L^2(e)} \\
&= \sum_{\Gamma_h \cup \Gamma_D} \left(\frac{\tilde{\sigma}}{C_F} - C_F \frac{4C^2(N)}{\epsilon} \right) \left(|| \, v|_{K_1} \, ||^2_{L^2(e)} + || \, v|_{K_2} \, ||^2_{L^2(e)} - 2 || \, v|_{K_1} \, ||_{L^2(e)} || \, v|_{K_2} \, ||_{L^2(e)} \right) \\
&\geq \sum_{\Gamma_h \cup \Gamma_D} \left(\frac{\tilde{\sigma}}{C_F} - C_F \frac{4C^2(N)}{\epsilon} \right) \left((1+\delta) \, || \, v|_{K_1} \, ||^2_{L^2(e)} + (1-\delta) \, || \, v|_{K_2} \, ||^2_{L^2(e)} \right) \\
&\geq \sum_{E \in \mathcal{E}_h} \frac{1-\delta}{4} \left(\frac{\tilde{\sigma}}{C_F} - C_F \frac{4C^2(N)}{\epsilon} \right) ||v||^2_{L^2(\partial E)}
\end{aligned}
\tag{2.45}
$$

Dabei wurde $\tilde{\sigma} > C_F^2 \frac{4C^2(N)}{\epsilon}$ und $\delta < 1$ gesetzt. Außerdem wurde im letzten Schritt verwendet, dass es sich bei der Triangulation $\mathcal{T}_h$ um ein Rechteckgitter handelt. Insgesamt resultiert damit die Form

$$
\mathcal{B}_h(v,v) \geq || \nabla \left(\beta v \right) ||^2_{\Omega,h} + \left(h^{-1} \sum_{K \in \mathcal{T}_h} \frac{1}{4} \left(\frac{\tilde{\sigma}}{C_F} - C_F \frac{4C^2(N)}{\epsilon} \right) (1-\delta) \, ||v||^2_{L^2(\partial K)} \right. \tag{2.46}
$$

$$
\left. - \left(\left(\epsilon C_F + 1 \right) C_I(N) + B_0 \right) ||v||^2_{L^2(K)} \right) \tag{2.47}
$$

Hierbei wurde außerdem die inverse Spurungleichung mit der Konstanten $C_I(N)$ genutzt. Falls mit einer Konstanten $C_\kappa > 1$

$$
\begin{aligned}
&\frac{1-\delta}{4} \left(\frac{\tilde{\sigma}}{C_F} - C_F \frac{4C^2(N)}{\epsilon} \right) \geq C_\kappa \left(\left(\epsilon C_F + 1 \right) C_I(N) + B_0 \right) \frac{||v||^2_{L^2(K)}}{||v||^2_{L^2(\partial K)}} \\
&\geq C_\kappa \frac{\left(\left(\epsilon C_F + 1 \right) C_I(N) + B_0 \right)}{4C(N)^2(N)} \geq C_\kappa \frac{\left(\left(\epsilon C_F + 1 \right) C_I(N) + B_0 \right)}{4C^2(N)}
\end{aligned}
\tag{2.48}
$$

gilt, folgt die Ungleichung

$$
\frac{1-\delta}{4} \left(\frac{\sigma}{C_F} - C_F \frac{4C^2(N)}{\epsilon} \right) ||v||^2_{L^2(\partial K)} \geq C_\kappa \left(\left(\epsilon C_F + 1 \right) C_I(N) + B_0 \right) ||v||^2_{L^2(K)} \tag{2.49}
$$

Damit ist mit der Norm

$$
|||v|||^2 := || \nabla \left(\beta v_h \right) ||^2_{\Omega,h} + h^{-1} \sum_{E \in \mathcal{E}_h} ||v_h||^2_E \tag{2.50}
$$

und der Konstanten $\kappa = \min \left(1, \left(C_\kappa - 1 \right) \left(\left(\epsilon C_F + 1 \right) C_I(N) + B_0 \right) \right)$ Koerzivität nach 1.14 erfüllt. Für die Sicherstellung der Stetigkeit wird die Sesquilinearform (2.26) betrachtet. Mit den bisherigen Überlegungen folgt sofort

$$
|\mathcal{B}_h \left(\rho_h, v \right)| \leq || \nabla \left(\beta \rho_h \right) ||_{\Omega,h} \, || \nabla \left(\beta \rho_h \right) ||_{\Omega,h} + || \nabla \rho_h ||_{\Omega,h} \, || \nabla v ||_{\Omega,h} + B_1 \, || \rho_h ||_{\Omega,h} \, ||v||_{\Omega,h} \tag{2.51}
$$

$$
+ \sum_{e \in \Gamma} || \nabla \rho_h ||_{L^2(e)} \, ||[v]||_{L^2(e)} + \frac{1}{C_F} \sum_{e \in \Gamma} || \nabla v ||_{L^2(e)} \, ||[\rho_h]||_{L^2(e)} + C_F \sum_{e \in \Gamma} \sigma \, ||[\rho_h]||_{L^2(e)} \, ||[v]||_{L^2(e)} \tag{2.52}
$$

Mithilfe von

$$\sum_{e\in\Gamma} ||[v]||^2_{L^2(e)} \leq 4C^2 h^{-1} \sum_{K\in\mathcal{T}_h} ||v||_{L^2(K)} \tag{2.53}$$

$$\sum_{e\in\Gamma} ||[\nabla v]||^2_{L^2(e)} \leq 4\tilde{C}^2 h^{-3} \sum_{K\in\mathcal{T}_h} ||v||_{L^2(K)} \tag{2.54}$$

Damit folgt die Stetigkeit mit der Konstanten $hC_c = \min\left(1, 4\tilde{C}^2 + 4C\tilde{C}\left(1+\frac{1}{C_F}\right) + C_F\sigma 4C^2 + B_1\right)$

A priori Fehlerabschätzung

Das Ziel dieses Abschnitts ist es, Voraussagen über den Fehler bei der numerischen Approximation der Lösung $\rho \in H^2(\Omega, \mathcal{T}_h)$ durch $\rho_h \in S_{hp}$ abzuschätzen. Die Vorarbeiten zur Koerzivität und Stetigkeit führen in natürlicher Weise auf eine Abschätzung in der mit (2.50) eingeführten Norm. Definiere den Fehler $e_h = \rho_h - \rho = \xi - \eta$ mit $\xi = \rho_h - \Pi_{hp}\rho \int S_{hp}$ und $\eta = \Pi_{hp}\rho - \rho \in H^2(\Omega, \mathcal{T}_h)$. Dabei beschreibt $\Pi_{hp}\rho$ die optimale $L^2(\Omega)$-Projektion von ρ auf den Raum S_{hp} ist, wie in (1.58) definiert. Mit der in 2.1.3 nachgewiesenen Konsistenz der Sesquilinearform folgt Galerkin-Orthogonalität (1.61) für den Fehler e_h

$$\mathcal{B}_h(\xi, v) = \mathcal{B}_h(\eta, v), \qquad v \in S_{hp} \tag{2.55}$$

Wähle nun $v = \xi$ und nutze Koerzivität und Stetigkeit von $\mathcal{B}_h$ aus, so dass

$$\kappa|||\xi|||^2 \leq \Re\mathcal{B}_h(\eta, \xi) \leq |\mathcal{B}_h(\eta, \xi)| \leq C_c|||\eta||| \cdot |||\xi||| \tag{2.56}$$

folgt. Dies impliziert

$$|||\xi||| \leq C_1|||\eta||| \leq C_2\, ||\eta||_{\Omega,h} \leq h^\mu\, |\rho|_{H^\mu(\Omega,\mathcal{T}_h)} \tag{2.57}$$

Im letzten Schritt wurde Lemma 1.15 benutzt. Unter Benutzung der Dreiecksgleichung folgt

$$|||e_h||| \leq |||\xi||| + |||\eta||| \leq C_3 h^\mu\, |u|_{H^\mu(\Omega,\mathcal{T}_h)} \tag{2.58}$$

mit $\mu = \min(N+1, s)$. Für hinreichend glatte Funktionen gilt

$$|||e_h||| \leq C_3 h^{N+1}\, |u|_{H^\mu(\Omega,\mathcal{T}_h)} \tag{2.59}$$

2.1.5. Implementierung

Das in 2.1.2 beschriebene Verfahren soll nun als numerisches Verfahren implementiert und auf die Form

$$H\rho = D\rho_D \tag{2.60}$$

gebracht werden. Mit der Anzahl der Elemente $k_{\max}$ in der Triangulation $\mathcal{T}_h$ und N_p der Anzahl der Knoten eines Elements $K \in \mathcal{T}_h$ sind H und D $k_{\max} \cdot N \times k_{\max} \cdot N$-Matrizen sowie ρ und ρ_D Vektoren der Dimension $k_{\max} \cdot N$. Für die Herleitung müssen sowohl Volumen- als auch Oberflächenintegrale ausgewertet und damit lokale Operatoren definiert werden. In diesem Kontext werden die Elemente $K \in \mathcal{T}_h$ durch eine lineare Abbildung $\gamma_{K\to\hat{K}} : \mathbb{R}^2 \to \mathbb{R}^2$, $[-1,1] \times [-1,1] \ni \hat{\mathbf{x}} := (r,s) \mapsto \mathbf{x} := (x,y)$,

$$\gamma_{K\to\hat{K}}(\hat{\mathbf{x}}) = B_K\hat{\mathbf{x}} + b_K \tag{2.61}$$

mit der Matrix

$$B_K = \begin{pmatrix} x_r\, x_s \\ y_r\, y_s \end{pmatrix} = \begin{pmatrix} \frac{h^{K,x}}{2} & 0 \\ 0 & \frac{h^{K,y}}{2} \end{pmatrix}, \qquad \left(B_K^{-1}\right)^T = \begin{pmatrix} r_x\, s_x \\ r_y\, s_y \end{pmatrix} = \begin{pmatrix} \frac{2}{h^{K,x}} & 0 \\ 0 & \frac{2}{h^{K,y}} \end{pmatrix} \tag{2.62}$$

auf ein Referenzelement $\hat{K}$ mit den Eckpunkten Punkten $(-1,-1)$, $(-1,1)$, $(1,1)$ und $(1,-1)$ abgebildet. Die Größen $h^{K,x}$ und $h^{K,y}$ sind die Volumina der vertikalen und horizontalen Kanten. Da die verwendeten Elemente Tensorproduktelemente sind, verschwinden die Nebendiagonalen. Die Abbildung ist für alle Elemente wegen der in ?? beschriebenen affinen Äquivalenz möglich und bewirkt eine erheblich Vereinfachung bei der Auswertung aller Integrale sowie eine Einsparung der Rechenzeit. Damit folgt für den Ableitungsoperator

$$\begin{pmatrix} \partial_x \\ \partial_y \end{pmatrix} = \left(B_K^T\right)^{-1} \begin{pmatrix} \partial_r \\ \partial_s \end{pmatrix} \tag{2.63}$$

und für die Jacobi-Determinante $J^K = \det\left|\left(B_K^T\right)^{-1}\right| = \frac{4}{h^{K,x}h^{K,y}} = \frac{4}{|K|}$. Die Diskretisierung des Problems erfolgt nun durch die Interpolation der Größen $\hat{\mathbf{q}} = \mathbf{q}\circ\gamma_{E\to\hat{E}}$ und $\hat{\rho} = \rho\circ\gamma_{K\to\hat{K}}$ in einer Basis aus lagrange-Polynomen $\{l_n\}_{n=1,\ldots,N_p}$ mit $l_n(\mathbf{x}_m) = \delta_{m,n}$ an den Punkten $\{\xi_n\}_{n=1,\ldots,N_p}$ auf dem Referenzelement $\hat{K}$. Die Testfunktion v wird auch aus dieser Basis gewählt. Damit ergeben sich Integrale, die von Produkten von Lagrange-Polynomen bzw. deren Ableitungen bestehen. Eine andere Möglichkeit für die Darstellung von $\mathbf{f}$ ist die Entwicklung in eine polynomiale Basis. Es ist also

$$\mathbf{f}(\mathbf{x},t) = \sum_{n=1}^{N_p} \mathbf{f}(\xi_{\mathbf{n}}, \mathbf{t})\,\ell_n(\mathbf{x}) = \sum_{n=1}^{N_p} \hat{f}_n \Psi_n(\mathbf{x}) \tag{2.64}$$

Die Gleichheit der letzten beiden Entwicklungen führt auf die Beziehung

$$\ell_{\mathrm{n}}(\mathbf{x}) = \sum_{n=1}^{N_p} \left(V^T\right)^{-1}_{m,n} \Psi(\mathbf{x}) \tag{2.65}$$

Damit kann jedes Integral über zwei Lagrange-Polynome in ein Integral in ein Integral über die Polynome Ψ_n umgeschrieben werden. Sind die $\{\Psi_n\}$ eine Basis aus Legendre-Polynomen, so kann der Fehler $||\mathbf{f} - \pi_{K,N}\mathbf{f}||_\infty$ in jeweils einer Richtung durch die Lösung des Problems $f(x) - \frac{d}{dx}P_{\mathrm{N}}(x) = 0$ mit $N_P = N+1$ gelöst werden [Quelle]. Die Wahl der Punkte $\{\xi_n\}_{n=1,\ldots,N_p}$ als Lösung dieses Problems sind die Legendre-Gauss-Lobatto Quadraturpunkte.

Volumenintegrale

Mit $\hat{v} = v \circ \gamma_{E \to \hat{E}} = \hat{l}_m$ und der Interpolation von ρ und $\mathbf{q}$ folgt sofort

$$\int_{K_1} (\rho_t v + \mathrm{i}\mathbf{q} \cdot \nabla v + \mathrm{i}B\rho v) \to J^{K_1} \left(M\rho_t^{K_1} + \mathrm{i}D_x^T M\mathbf{q}_x^{K_1} + \mathrm{i}D_y^T M\mathbf{q}_y^{E_1} + \mathrm{i}MFG_1^K \rho^{K_1} \right) \quad (2.66)$$

$$\int_{K_1} (\mathbf{q}v - A\nabla\rho \cdot v) \to \begin{pmatrix} J^{K_1} \left(M\mathbf{q}_x^{K_1} - MD_y\rho^{K_1} \right) \\ J^{K_1} \left(M\mathbf{q}_y^{K_1} - MD_x\rho^{K_1} \right) \end{pmatrix} \quad (2.67)$$

mit den Vektoren

$$\rho^K = \left(\rho^K(x_1), \dots, \rho^K(x_{N_p}) \right)^T \quad (2.68)$$

$$\mathbf{q}_x^K = \left(q_x^K(x_1), \dots, q_x^K(x_{N_p}) \right)^T \quad (2.69)$$

$$\mathbf{q}_y^K = \left(q_y^K(x_1), \dots, q_y^K(x_{N_p}) \right)^T \quad (2.70)$$

Der lokale Operator M mit den Elementen

$$M_{m,n} = \int_{\hat{K}} \hat{l}_m \hat{l}_n \quad (2.71)$$

wird als Massematrix bezeichnet. Diese kann genau wie in durch die Darstellung der Lagrange Polynome in der Basis der orthogonalen Legendre-Polynnome $\{\Psi\}_{m=1,\dots,N_p}$ berechnet werden:

$$M_{m,n} = \int_{\hat{E}} \hat{l}_m \hat{l}_n = \sum_{i,j=1}^{N_p} V_{m,i}^{-1} \left(V^T \right)_{j,n}^{-1} \int_{\hat{E}} \Psi_i \Psi_j = \sum_{i=1}^{N_p} V_{m,i}^{-1} \left(V^T \right)_{i,n}^{-1} \to M = \left(V^T V \right)^{-1} \quad (2.72)$$

mit der Vandermondematrix $V_{m,n} = \Psi_n(\xi_m)$. Die Operatoren D_r und D_s auf dem Referenzelement $\hat{K}$ sind diskretisierte Ableitungen auf dem Element $\hat{K}$. Sie kann als

$$(D_r)_{m,n} = \left. \frac{d\hat{l}_n}{dr} \right|_{\xi_m} \quad (2.73)$$

Mithilfe der Affinen Abbildung (2.61) können die Matrizen $D_x = r_x D_r + s_x D_s$ und $D_y = r_y D_r + s_y D_s$ für das physikalische Element definiert werden. Mithilfe diesen Matrizen kann die Verbindung zur in (2.66) auftretenden Steifigkeitsmatrix $(S_{x_\mathrm{i}})_{m,n} = \int_{\hat{K}} \hat{l}_m \partial_{x_\mathrm{i}} \hat{l}_n$ mit $S_{x_\mathrm{i}} = MD_{x_\mathrm{i}}$ mit $i = 1, 2$. Die Funktion $B(x, y)$ ist hochgradig nichtlinear. Für die Auswertung des Ausdrucks $\int_K B\rho v$ sollen zwei Möglichkeiten vorgestellt werden. Eine mögliche Variante wäre die Approximation des Integrals durch eine Gauss-Quadratur, so dass auf mit einem Satz aus N_q Punkten und Gewichten (x_n, w_n)

$$\int_{\hat{K}} (B \circ \gamma) \hat{\ell}_m \hat{\ell}_n \approx \sum_{k,q=1}^{N_p} \left(V^T \right)_{m,k}^{-1} (V)_{q,n}^{-1} \sum_{j=1}^{N_\mathrm{q}} B(\gamma(\hat{\mathbf{x}}_\mathrm{j})) \Psi_k(\mathbf{x}_\mathrm{j}) \Psi_q(\mathbf{x}_\mathrm{j}) w_j \quad (2.74)$$

Das führt dann zu der Matrix

$$G = \left(V^T \right)^{-1} \tilde{V}^T \mathrm{diag}\left(\{ B(\gamma(\hat{\mathbf{x}}_\mathrm{j})) \cdot w_j \} \right) \tilde{V} V^{-1} \quad (2.75)$$

wobei eine weitere Vandermonde-Matrix $\tilde{V}_{m,n} = \Psi_{\mathrm{n}}(\mathbf{x}_{\mathrm{m}})$ definiert wurde. Eine andere Möglichkeit ist die Interpolation von $B\rho$. Die hohe Nichtlinearität von B motiviert zur Einführung einer Filterfunktion in

$$(B\rho)(\mathbf{x}) = \sum_{n=1}^{N_p} \sigma\left(\frac{n-1}{N}\right) \hat{B\rho}_n \Psi_n(\mathbf{x}) \tag{2.76}$$

Eine oft verwendete Wahl ist $\sigma\left(\frac{n-1}{N}\right) = \exp\left(-\alpha\left(\frac{n-1}{N}\right)^{2s}\right)$. Umgeschrieben auf eine nodale Entwicklung führt diese Überlegung auf die Filtermatrix $F = V\Lambda V^{-1}$ mit $\Lambda_{m,m} = \sigma\left(\frac{i+j}{N}\right)$, $m = j + (N+1)i$, $0 \leq i, j \leq N$. Das führt dann auf die Matrix

$$G^K_{m,n} = \hat{B}(\xi^K_m)\delta_{m,n} \tag{2.77}$$

Oberflächenintegrale

Die aus den Oberflächenintegralen resultierenden lokalen Operatoren werden analog zu (2.19)-(2.20) als Lifting-Operatoren eingeführt, d.h. sie bilden den Vektor ρ^K, der ja eine Interpolation von ρ auf dem Element K darstellt, auf einen Oberflächenterm ab. Nach der in (2.14)-(2.15) getroffenen Wahl des numerischen Flusses gilt für die jeweiligen Integranden

$$\mathbf{n}\mathbf{q}^* = \begin{cases} n_x\{\rho_y\} + n_y\{\rho_x\} - \sigma\mathbf{n}[\rho], & q \in \Gamma_I \\ n_x\partial_y\rho_D + n_y\partial_x\rho_D - 2\sigma(\rho - \rho_D), & q \in \Gamma_D \end{cases} \tag{2.78}$$

$$A\mathbf{n}(\rho - \rho^*) = \begin{cases} \frac{1}{2}K[\rho], & \rho \in \Gamma_I \\ A\mathbf{n}(\rho - \rho_D), & \rho \in \Gamma_D \end{cases} \tag{2.79}$$

Ähnlich wie bei den Volumenintegralen werden die Oberflächenintegrale auf einer Kante e durch eine geeignete Parametrisierung auf ein Referenzelement abgebildet. Zunächst werden aber die Ableitungen in den Integranden durch die Koordinatentransformation $\gamma_{K\to\hat{K}}$ umgeschrieben und ρ und $\mathbf{q}$ entsprechend ?? interpoliert. Dann werden die Integrale mit der Abbildung $\gamma_e : \mathbb{R} \to \mathbb{R}^2$ und einem Parameter $t \in [-1,1]$,

$$\gamma_e(t) = \frac{1}{2}(\mathbf{N}_{e,2} - \mathbf{N}_{e,1})(t+1) + \mathbf{N}_{e,1} \tag{2.80}$$

mit $h^e = |\mathbf{N}_{e,2} - \mathbf{N}_{e,1}|$ parametrisiert. Die Eckpunkte der Kante werden mit $\mathbf{N}_{e,1}$ und $\mathbf{N}_{e,2}$. Die Jacobi-Determinante ist $J^e = \det|\partial_t\gamma_e| = \frac{h^e}{2}$. Damit folgt für die Integrale in (2.13)-(2.14)

$$\int_e \mathbf{n}_e \mathrm{q}^* v = \sum_{n=1}^{N_p} J^e \int_{\gamma_e} \left(\sum_{j=1}^{N_p} \left\{\tilde{\rho}(\xi_n)(D_n)_{j,n}\tilde{l}_j\right\}\tilde{l}_m - \sigma\mathbf{n}_e\left[\tilde{\rho}(\xi_n)\tilde{l}_n\right]\tilde{l}_m\right) \tag{2.81}$$

$$\int_e A\cdot\mathbf{n}_e\cdot(\rho - \rho^*)v = \frac{1}{2}\sum_{n=1}^{N_{p,e}} J^e \int_{\gamma_e} A\left[\tilde{\rho}(\eta_n)\tilde{l}_n\right]\tilde{l}_m \tag{2.82}$$

mit $D_n = n_yD_x + n_xD_y$. Dabei wurde die Beziehung

$$\frac{d\hat{l}_n}{dr} = \sum_{j=1}^{N_p} \left.\frac{d\hat{l}_n}{dr}\right|_{\xi_j} \hat{l}_j = \sum_{j=1}^{N_p} (D_r)_{j,n}\, l_j \tag{2.83}$$

ausgenutzt. Die Auswertung der Mittelwerte $\{\cdot\}$ und Sprünge $[\cdot]$ nach den Definitionen ?? führt zur Einführung der Lifting-Operatoren V^e und W^e, wobei V^e innere Funktionswerte $\tilde{\rho}^{K_1}$ und W^e äußere Funktionswerte $\tilde{\rho}^{K_2}$ auf die Kante $e \subset K_1$ abbildet. Damit folgt

$$\int_{\gamma_e} \left\{\tilde{\rho}\left(\xi_n\right)\left(D_n\right)_{j,n} \tilde{l}_j\right\} \tilde{l}_m = \frac{1}{2}\left(V^e_{m,j}\left(D_n^{K_1}\right)_{j,n} \tilde{\rho}^{K_1}\left(\xi_n\right) + W^e_{m,j}\left(D_n^{K_2}\right)_{j,n} \tilde{\rho}^{K_2}\left(\xi_n\right)\right) \tag{2.84}$$

$$\int_{\gamma_e} \left\{\tilde{\rho}\left(\xi_n\right) \tilde{l}_j\right\} \tilde{l}_m = \frac{1}{2}\left(V^e_{m,j}\tilde{\rho}^{K_1}\left(\xi_n\right) + W^e_{m,j}\tilde{\rho}^{K_2}\left(\xi_n\right)\right) \tag{2.85}$$

$$\int_{\gamma_e} \mathbf{n}\left[\tilde{\rho}\left(\xi_n\right) \tilde{l}_j\right] \tilde{l}_m = V^e_{m,j}\tilde{\rho}^{K_1}\left(\xi_n\right) - W^e_{m,j}\tilde{\rho}^{K_2}\left(\xi_n\right) \tag{2.86}$$

Eingesetzt in die Gleichungen ergibt sich insgesamt

$$\int_e \mathbf{n}_e \mathrm{q}^* v \rightarrow J^e \begin{cases} \frac{1}{2} V^e \left(D_n^{K_1} - 2\sigma I_{N_p}\right) \rho^{K_1} + \frac{1}{2} W^e \left(D_n^{K_2} + 2\sigma I_{N_p}\right) \rho^{K_2}, & e \in \Gamma_I \\ V^e D_n^{K_1} \rho^{K_1} - 2\sigma V^e \rho^{K_1} + 2\sigma V^e \rho_D^{K_1}, & e \in \Gamma_D \end{cases} \tag{2.87}$$

$$\int_e A \cdot \mathbf{n}_e \cdot \left(\rho - \rho^*\right) v \rightarrow J^e A \cdot \mathbf{n} \begin{cases} \frac{1}{2} V^e \rho^{K_1} - \frac{1}{2} W^e \rho^{K_2}, & e \in \Gamma_I \\ V^e \rho^{K_1} - V^e \rho_D^{K_1}, & e \in \Gamma_D \end{cases} \tag{2.88}$$

Die Matrizen V^e und W^e haben die Eigenschaft gemeinsam, dass Einträge mit Indizes zugehörig zu Lagrange-Polynomen und damit zu Punkten aus dem Inneren des Elements (da $l_m\left(x_n\right) = \delta_{ij}$) verschwinden. Die nicht verschwindenden Einträge dieser Matrizen entspricht damit dem Quadrat der Anzahl der Punkte N_{px} oder N_{py} auf einer Kante e. Technisch erzeugt das Integral über eine Kante einen eindimensionalen lokalen Operator M^{1D}, der entsprechend der Punktenummerierung bzw. der Nummerierung der Einträge in den Vektoren ρ^K in V^e und W^e eingebettet wird. Die Einbettung als Untermatrix erfolgt dann gemäß der in Abb. ?? gezeigten Nummerierung:

i) **<u>Linke Seite:</u>**

$$V^{e_l}_{m,n} = \sum_{i,j=0}^{N_{py}-1} \delta_{m,1+i\cdot N_{px}} \delta_{n,1+j\cdot N_{px}} \left(M^{1D,e}\right)^{m,n}_{i,j} \tag{2.89}$$

$$W^{e_l}_{m,n} = \sum_{i,j=0}^{N_{py}-1} \delta_{m,1+i\cdot N_{px}} \delta_{n,(j+1)\cdot N_{px}} \left(M^{1D,e}_{i,j}\right)^{m,n} \tag{2.90}$$

ii) **<u>Obere Seite:</u>**

$$V^{e_t}_{m,n} = \sum_{i,j=0}^{N_{px}-1} \delta_{m,1+i} \delta_{n,1+j} \left(M^{1D,e}\right)^{m,n}_{i,j} \tag{2.91}$$

$$W^{e_t}_{m,n} = \sum_{i,j=0}^{N_{px}-1} \delta_{m,1+i} \delta_{n,j+1+(N_{py}-1)\cdot N_{px}} \left(M^{1D,e}\right)^{m,n}_{i,j} \tag{2.92}$$

iii) **Rechte Seite:**

$$V_{m,n}^{e_r} = \sum_{i,j=0}^{N_{py}-1} \delta_{m,(1+i)\cdot N_{px}} \delta_{n,(1+j)\cdot N_{px}} \left(M^{1D,e}\right)_{i,j}^{m,n} \tag{2.93}$$

$$W_{m,n}^{e_r} = \sum_{i,j=0}^{N_{py}-1} \delta_{m,(1+i)\cdot N_{px}} \delta_{n,1+j\cdot N_{px}} \left(M^{1D,e}\right)_{i,j}^{m,n} \tag{2.94}$$

iv) **Untere Seite:**

$$V_{m,n}^{e_b} = -\sum_{i,j=0}^{N_{px}-1} \delta_{m,i+1+N_{px}\cdot(N_{py}-1)} \delta_{n,j+1+N_{px}\cdot(N_{py}-1)} \left(M^{1D,e}\right)_{i,j}^{m,n} \tag{2.95}$$

$$W_{m,n}^{e_b} = -\sum_{i,j=0}^{N_{px}-1} \delta_{m,i+1+N_{px}\cdot(N_{py}-1)} \delta_{n,j+1} \left(M^{1D,e}\right)_{i,j}^{m,n} \tag{2.96}$$

Die Matrix $M^{1D,e}$ ist dann definiert als

$$\left(M^{\mathrm{1D,e}}\right)_{j,m}^{m,n} = \int_{\gamma_e} \tilde{l}_m \tilde{l}_n \tag{2.97}$$

Bezüglich der Dirichlet-Randbedingungen muss beachtet werden, dass diese zwar in diesem Kontext als äußere Zustände am Rand interpretiert werden, aber links vom Rand $\partial\Omega$ gar kein Element und bis auf die Kanten $e \in \Gamma_D$ keine Punkte definiert sind. Es wird daher der Vektor

$$\left(\rho_D^{K_1}\right)_n = \begin{cases} \rho_D\left(\mathbf{x}_n\right), & \mathbf{x}_n \in e \subset \partial K_1 \cap \Gamma_D \\ 0, & \text{sonst} \end{cases} \tag{2.98}$$

Somit ergibt sich insgesamt für (2.81)-(2.82)

$$\begin{aligned} \mathrm{i} \sum_{e\in\partial K_1} \int_e \mathbf{n} \mathrm{q}^* v \to \frac{\mathrm{i}}{2} \sum_{e\in\partial K_1\cap\Gamma_I} & J^e \left(V^e \left(D_n^{K_1} - 2\sigma I_{N_p}\right) \rho^{K_1} + W^e \left(D_n^{K_2} + 2\sigma I_{N_p}\right) \rho^{K_2}\right) \\ + \mathrm{i} \sum_{e\in\partial K_1\cap\Gamma_D} & J^e \left(V^e D_n^{K_1} \rho^{K_1} - 2\sigma V^e \rho^{K_1} + 2\sigma V^e \rho_D^{K_1}\right) \end{aligned} \tag{2.99}$$

$$-\sum_{e\in\partial K_1} \int_e A \cdot \mathbf{n} \cdot (\rho - \rho^*)\, v \to -\frac{1}{2} \sum_{e\in\partial K_1\cap\Gamma_I} J^e K \mathbf{n} \left(V^e \rho^{K_1} - W^e \rho^{K_2}\right) - \sum_{e\in\partial K_1\cap\Gamma_D} J^e A \mathbf{n} \left(V^e \rho^{K_1} - V^e \rho_D^{E_1}\right) \tag{2.100}$$

Semi-infinite Elemente

Für die Realisierung der Randbedingung 2.6 existieren mehrere Möglichkeiten. Eine davon ist die Einführung von sogenannten infiniten Elementen. Hierbei gibt es wiederum zwei grundlegende Ansätze, den *mapping*-Ansatz und den Ansatz mit exponentiell abfallenden modalen

Funktionen.
In dieser Arbeit wird der *mapping*-Ansatz gewählt. Da, ebenso wie für die finite-Elemente-Formulierung, Polynomgrade beliebiger Ordnung N_y möglich sein sollen, wird ein Rechteck mit unendlich großer Ausdehnung in der Richtung y die allgemeine Abbildung $\mathbf{x} = \gamma_{K \to \hat{K}} : [-1,1] \times [-1,1] \to [\mathbf{N}_{1,x}, \mathbf{N}_{2,x}] \times [\mathbf{N}_{1,y}, \infty]$

$$y\,(s) = \sum_{n=0}^{N_y-1} \frac{\alpha_n s^n}{1-s} \tag{2.101}$$

$$x\,(r) = \frac{h_x}{2}\,(r+1) + N_{1,x} \tag{2.102}$$

vorgeschlagen. In der x-Richtung ist dies die übliche Abbildung auf das Referenzelement $\hat{K}$. wobei der N_{py}-te Punkt im unendlichen liegt und bei der Berechnung der lokalen Operatoren irrelevant ist, da ρ auf dem Rand verschwindet. Die Ableitung lautet dann

$$\frac{dy}{ds} = \sum_{n=1}^{N_y-1} \frac{n\alpha_n s^{n-1}}{1-s} - \sum_{n=0}^{N_y-1} \frac{\alpha_n s^n}{(1-s)^2} = \sum_{n=0}^{N_y-1} \frac{\beta_n(s)}{(1-s)^2}, \qquad \frac{dx}{dr} = \frac{h_x}{2} \tag{2.103}$$

mit

$$\beta_n(s) = \begin{cases} -\alpha_0 & n = 0 \\ \alpha_n s^{n-1}\,(n-(n+1)\,s) & n > 0 \end{cases} \tag{2.104}$$

Damit gilt für die Jacobi-Determinante

$$J = \left| \begin{pmatrix} \frac{h_x}{2} & 0 \\ 0 & \frac{\partial y}{\partial s} \end{pmatrix} \right| = \frac{h_x}{2}\frac{\partial y}{\partial s} \tag{2.105}$$

und den Gradienten

$$\begin{pmatrix} \partial_x \\ \partial_y \end{pmatrix} = \begin{pmatrix} \frac{2}{h_x} & 0 \\ 0 & \frac{\partial s}{\partial y} \end{pmatrix} \begin{pmatrix} \partial_r \\ \partial_s \end{pmatrix} \tag{2.106}$$

Die Koeffizienten α_n müssen anhand der Knotenpunkte des physikalischen Elements K und des Referenzelements $\hat{K}$ bestimmt werden. Die Wahl der Knotenpunkte $\{s\}_{j=1,\ldots,N_{py}}$ kann willkürlich festgelegt werden. Unter Beachtung derselben in ?? angestellten Überlegung bezüglich der Randbedingungen werden die Abstände der ersten $N_{py}-1$-Punkte genauso äquidistant gewählt wie die der finiten Elemente. Mit der Einführung einer weiteren Vandermondematrix

$$\mathcal{V}_{j,n} = \frac{s_j^n}{1-s_j} \tag{2.107}$$

ergeben sich die Koeffizienten $(\alpha_n)_{n=1\ldots N_{py}-1}$ zu

$$\alpha = \mathcal{V}^{-1}\mathbf{y} \tag{2.108}$$

mit $\alpha = (\alpha_1, \ldots, \alpha_{N_{px}-1})^T$ und $\mathbf{y} = (y(s_1), \ldots, y(s_{N_{px}}))^T$. Anschließend werden die Matrixelemente aus den vorherigen Sektionen für Infinite Elemente berechnet:

1. Massematrix: Die Massematrix ist nicht mehr so einfach in geschlossener Form anzugeben. Das bisherige Verfahren führt zunächst auf den Ausdruck

$$\tilde{M}^K_{m,n} = \int_{\tilde{E}} \tilde{l}_m(r,s)\tilde{l}_n(r,s)J^K(s) = \sum_{p,q=1}^{N_p} V^{-1}_{p,n} \left(V^T\right)^{-1}_{m,q} \int_{\tilde{K}} \Psi_p(r,s)\Psi_q(r,s)J^E(s) \quad (2.109)$$

Die zweidimensionalen Legendre-Polynome $\Psi_n(r,s)$ sind (ebenso wie die Lagrange-Polynome) aufgrund des Tensorproduktcharakters der Elemente mit $p = i + N_{px}j$ und $q = v + N_{px}w$ separierbar in $\Psi_p(r,s) = \psi_i(r)\psi_j(s)$, wobei die Polynome Ψ_i eindimensionale Legendre-Polynome darstellen. Mit dieser Darstellung lässt sich das Problem in s-Richtung einfach isolieren und das Integral mit einer Gauss-Legendre-Quadraturformel näherungsweise auswerten. Es folgt

$$\begin{aligned}\int_{\tilde{K}} \Psi_p\Psi_q J^K(s) &= \int_{-1}^{1} dr\psi_i(r)\psi_v(r) \int_{-1}^{1} ds\psi_j(s)\psi_w(s)J^K(s) \\ &\approx \delta_{i,v} \sum_{a,b=1}^{N_q} \tilde{V}^T_{j,a} w_a J^K(s_a)\delta_{a,b}\tilde{V}_{b,w} =: \delta_{i,v}\Phi_{j,w}\end{aligned} \quad (2.110)$$

wobei $\tilde{V}$ eine Vandermonde-Matrix an den Gauss-Legendre-Stützstellen $\{s_i\}_{i=1,\ldots,N_q}$ mit zugehörigen Gewichten $\{w_i\}_{i=1,\ldots,N_q}$ ausgewertet ist. Damit folgt insgesamt für die Massematrix des semi-infiniten Elements

$$\tilde{M}^K = \left(V^T\right)^{-1} \left(I_{N_{px}} \otimes \Phi^{(k)}\right) V^{-1} = \frac{h^{K,x}}{2}\tilde{M} \quad (2.111)$$

2. Steifigkeitsmatrix: Die Steifigkeitsmatrizen S_x und S_y werden genau wie im endlichen Fall über die Ableitungsmatrizen D_r und D_s gebildet:

$$(S_x)_{m,n} = \int_{\tilde{K}} \tilde{l}_m \frac{d\tilde{l}_n}{dr}\frac{dy}{ds} = \sum_{j=1}^{N_p} (D_r)_{j,n} \int_{\tilde{K}} \tilde{l}_m\tilde{l}_j \frac{dy}{ds} = \sum_{j=1}^{N_p} \tilde{M}_{m,j}(D_r)_{j,n} \rightarrow \tilde{S}_x = \tilde{M}D_r \quad (2.112)$$

$$(S_y)_{m,n} = \frac{h^{K,x}}{2} \int_{\tilde{K}} \tilde{l}_m \frac{d\tilde{l}_n}{ds} = \frac{h^K_x}{2} \sum_{j=1}^{N_p} (D_s)_{j,n} \int_{\tilde{K}} \tilde{l}_m\tilde{l}_j \rightarrow \frac{h^K_x}{2} MD_s \quad (2.113)$$

3. Oberflächenintegrale: Es gilt für innere Kanten $e \in \Gamma_h$

$$\begin{aligned}\int_e \mathbf{nq}^* v &= \int_e \left(n_x \{\rho_y\} v + n_y \{\rho_x\} v - \sigma\mathbf{n}[\rho] v\right) \\ &= \int_e \left(n_x \left\{\hat{\rho}_s \left(\frac{\partial y}{\partial s}\right)^{-1}\right\} \hat{v} + n_y \left\{\hat{\rho}_r \frac{2}{h^{x,e}}\right\} \hat{v} - \sigma\mathbf{n}[\hat{\rho}]\hat{v}\right)\end{aligned} \quad (2.114)$$

Bei den Oberflächenintegralen muss nun zwischen vertikalen und horizontalen Kanten unterschieden werden. Für die Integration über die horizontalen Kanten wird die Abbildung

$$\gamma_e(r) = (\mathbf{N}_2 - \mathbf{N}_1)\frac{r+1}{2} + \mathbf{N}_1 = \frac{h_x}{2}(r+1) + \mathbf{N}_1 \quad (2.115)$$

herangezogen. Damit folgt mit dem Flächenelement $do = \frac{h^{x,e}}{2} dr$

$$\begin{aligned}\int_e \mathbf{n}\mathbf{q}^* v &= \int_{\gamma_e} \left(n_y \left\{ \hat{\rho}_r \right\} \hat{v} - \sigma \frac{h^{x,e}}{2} \mathbf{n} \left[\hat{\rho} \right] \hat{v} \right) \\ &= \sum_{n=1}^{N_p} \int_{\gamma_e} \left(\sum_{j=1}^{N_p} n_y \left\{ \hat{\rho}\left(\xi_n\right) \left(D_r\right)_{j,n} \hat{l}_j \right\} \hat{l}_m - \sigma \frac{h^{x,e}}{2} \mathbf{n} \left[\hat{\rho}\left(\xi_n\right) \hat{l}_n \right] \hat{l}_m \right) \\ &\to \frac{J^e}{2} \left(V^e \left(D_n^{K_1} - 2\sigma I_{N_p} \right) \rho^{K_1} + W^e \left(D_n^{E_2} + 2\sigma I_{N_p} \right) \rho^{K_2} \right)\end{aligned} \tag{2.116}$$

Für die Integration über die vertikalen Kanten wird die Abbildung

$$x(s) = N_{1,x} \tag{2.117}$$

$$y(s) = \sum_{n=0}^{N_y-1} \frac{\alpha_n s^n}{1-s} \tag{2.118}$$

herangezogen. Damit folgt mit dem Flächenelement $do = \frac{\partial y}{\partial s} ds$

$$\int_e \mathbf{n}\mathbf{q}^* v = \int_{\gamma_e} \left(n_x \left\{ \hat{\rho}_s \right\} \hat{v} - \sigma \frac{\partial y}{\partial s} \mathbf{n} \left[\hat{\rho} \right] \hat{v} \right) = \sum_{n=1}^{N_p} \int_{\gamma_e} \left(\sum_{j=1}^{N_p} n_x \left\{ \hat{\rho}\left(\xi_n\right) \left(D_s\right)_{j,n} \hat{l}_j \right\} \hat{l}_m - \sigma \frac{\partial y}{\partial s} \mathbf{n} \left[\hat{\rho}\left(\xi_n\right) \hat{l}_n \right] \hat{l}_m \right) \tag{2.119}$$

$$\to \frac{1}{2} \left(V^e n_x D_s - 2\sigma \tilde{V}^e \right) \rho^{K_1} + \frac{1}{2} \left(W^e n_x D_s + 2\sigma \tilde{W}^e \right) \rho^{K_2} \tag{2.120}$$

Dabei besitzen die Matrizen $\tilde{V}^e$ und $\tilde{W}^e$ die selbe Belegung wie V^e und W^e mit der Matrix

$$\tilde{M}^{1D,e} = \left(V^T \right)^{-1} \left(\sum_{k=1}^{N_{py}-1} k \alpha_k \Phi^{(k)} \right) V^{-1} \tag{2.121}$$

Für Dirichlet-Kanten $e \in \Gamma_D$ folgt analog zu den vorherigen Rechnungen

$$\int_e \mathbf{n}\mathbf{q}^* v = \begin{cases} J^e \left(V^e D_n^{K_1} \rho^{E_1} - 2\sigma V^e \rho^{K_1} + 2\sigma V^e \rho_D^{K_1} \right) & e \in \Gamma_x \\ \left(V^e n_x D_s - 2\sigma \tilde{V}^e \right) \rho^{K_1} + 2\sigma \tilde{V}^e \rho_D^{K_1} & e \in \Gamma_y \end{cases} \tag{2.122}$$

Analog folgt für die anderen beiden Gleichungen für $e \subset \Gamma_h$

$$\int_e \frac{1}{2} K \left[\rho \right] v = \frac{1}{2} A \mathbf{n} \begin{cases} J^e \left(V^e \rho^{E_1} - W^e \rho^{K_2} \right) & e \in \Gamma_x \\ \tilde{V}^e \rho^{K_1} - \tilde{W}^e \rho^{K_2} & e \in \Gamma_y \end{cases} \tag{2.123}$$

und für $e \subset \Gamma_D$

$$\int_e K \mathbf{n} \left(\rho - \rho_D \right) v = A \mathbf{n} \begin{cases} J^e V^e \left(\rho^{K_1} - \rho_D^{K_1} \right) & e \in \Gamma_x \\ \tilde{V}^e \left(\rho^{K_1} - \rho_D^{K_1} \right) & e \in \Gamma_y \end{cases} \tag{2.124}$$

Gleichungssystem

Führt man nun die Rechnungen aus den Abschnitten ?? und ?? zusammen, ergibt sich ein Gleichungssystem in kompakter Form

$$\frac{\partial \rho^{K_1}}{\partial t} + \mathrm{i} A_1^{K_1} \rho^{K_1} + \mathrm{i} \sum_{e \in \partial K_1 \cap \Gamma_h} A_2^e \rho^{K_2(e)} = \mathrm{i} A_3^{K_1} \rho_D^{K_1} \tag{2.125}$$

$$A_4^{E_1} \mathbf{q}_x^{K_1} + A_5^{K_1} \rho^{K_1} + \sum_{e \in \partial K_1 \cap \Gamma_h} A_6^e \rho^{K_2(e)} = A_7^{K_1} \rho_D^{K_1} \tag{2.126}$$

Mit den Finite-Elemente-Matrizen

$$\begin{aligned} A_1^K &= \left(M^{-1} D_y^T M D_x + M^{-1} D_y^T M D_x + G \right) \\ &\quad - \frac{1}{2J^K} \sum_{e \in \partial K \cap \Gamma_h} J^e M^{-1} \left(V^e D_n^{K_1} - 2\sigma V^e + \left(D_n^{E_1} \right)^T V^e \right) \\ &\quad - \frac{1}{J^K} \sum_{e \in \partial K \cap \Gamma_D} J^e M^{-1} \left(V^e D_n^{K_1} + -2\sigma V^e + \left(D_n^{K_1} \right)^T V^e \right) \end{aligned} \tag{2.127}$$

$$A_2^e = -\frac{1}{2J^{K_1(e)}} J^e M^{-1} \left(W^e D_n^{E_2} + 2\sigma W^e - \left(D_n^{K_1} \right)^T W^e \right) \tag{2.128}$$

$$A_3^K = \frac{1}{J^K} \sum_{e \in \partial E \cap \Gamma_D} J^e M^{-1} \left(2\sigma V^e - \left(D_n^{K_1} \right)^T V^e \right) \tag{2.129}$$

$$A_4^K = J^K M \tag{2.130}$$

$$A_5^K = -J^{K_1} M D_y + \frac{1}{2} \sum_{e \in \partial K \cap \Gamma_h} J^e n_y V^e + \sum_{e \in \partial K \cap \Gamma_D} J^e n_y V^e \tag{2.131}$$

$$A_6^e = -\frac{1}{2} J^e n_y W^e \tag{2.132}$$

$$A_7^K = \sum_{e \in \partial K \cap \Gamma_D} J^e n_y V^e \tag{2.133}$$

und den IEM-Matrizen

$$\begin{aligned} A_1^K &= \frac{2}{h_x} \tilde{M}^{-1} \left(D_r^T M D_s + D_s^T M D_r \right) + G \\ &\quad - \frac{2}{h_x} \tilde{M}^{-1} \sum_{e \in \partial K_1 \cap \Gamma_h} \frac{1}{2} \begin{cases} J^e \left(V^e D_n^{K_1} - 2\sigma V^e + D_m^{K_1} V^e \right) & e \in \Gamma_x \\ V^e n_x D_s - 2\sigma \tilde{V}^e + D_m^{K_1} \tilde{V}^e & e \in \Gamma_y \end{cases} \\ &\quad - \frac{2}{h_x} \tilde{M}^{-1} \sum_{e \in \partial K_1 \cap \Gamma_D} \begin{cases} J^e \left(V^e D_n^{K_1} + -2\sigma V^e + D_m^{K_1} V^e \right) & e \in \Gamma_x \\ V^e n_x D_s + -2\sigma \tilde{V}^e + D_m^{K_1} \tilde{V}^e & e \in \Gamma_y \end{cases} \end{aligned} \tag{2.134}$$

$$A_2^e = -\frac{2}{h_x}\tilde{M}^{-1}\frac{1}{2}\begin{cases} J^e\left(W^e D_n^{K_2} + 2\sigma W^e - D_m^{K_1} W^e\right) & e \in \Gamma_x \\ W^e n_x D_s + 2\sigma \tilde{W}^e - D_m^{K_1}\tilde{W}^e & e \in \Gamma_y \end{cases} \tag{2.135}$$

$$A_3^K = \frac{2}{h_x}\tilde{M}^{-1}\begin{cases} J^e\left(2\sigma V^e - D_m^{K_1} V^e\right) & e \in \Gamma_x \\ 2\sigma\tilde{V}^e - D_m^{K_1}\tilde{V}^e & e \in \Gamma_y \end{cases} \tag{2.136}$$

$$A_4^K = \frac{h_x}{2}\tilde{M} \tag{2.137}$$

$$A_4^K = -\frac{h_x}{2} M D_s + \sum_{e \in \partial K_1 \cap \Gamma_h} \frac{1}{2} n_y \begin{cases} J^e V^e & e \in \Gamma_x \\ \tilde{V}^e & e \in \Gamma_y \end{cases} + \sum_{e \in \partial K_1 \cap \Gamma_D} n_y \begin{cases} J^e V^e & e \in \Gamma_x \\ \tilde{V}^e & e \in \Gamma_y \end{cases} \tag{2.138}$$

$$A_6^e = -\frac{1}{2} n_y \begin{cases} J^e W^e & e \in \Gamma_x \\ \tilde{W}^e & e \in \Gamma_y \end{cases} \tag{2.139}$$

$$A_7^K = n_y \begin{cases} J^e W^e & e \in \Gamma_x \\ \tilde{W}^e & e \in \Gamma_y \end{cases} \tag{2.140}$$

Die drei ursprünglichen Gleichungssysteme für $(\rho, \mathbf{q}_x, \mathbf{q}_y)$ wurden dabei zugunsten der interessierenden Lösung ρ aufgelöst, indem die Flusskomponenten $\mathbf{q}_x(\rho)$ und $\mathbf{q}_y(\rho)$ wieder in jetzt diskretisierte ursprüngliche Gleichung eingesetzt wurden. Zusätzlich ist das Gleichungssystem für die x-Komponente $\mathbf{q}_x$ aufgeführt, da diese die numerische Approximation von $\partial_y \rho$ darstellt und diese über ?? mit dem gesuchten Strom verknüpft ist. Mit der Einführung der Vektoren

$$\overline{\rho} = \left(\rho^{K_1}, \ldots, \rho^{K_{k_{\max}}}\right)^T \tag{2.141}$$

$$\mathbf{q}_x = \left(\mathbf{q}_x^{K_1}, \ldots, \mathbf{q}_x^{K_{k_{\max}}}\right)^T \tag{2.142}$$

folgt für das Gleichungssystem die kurze Form

$$\frac{\partial\overline{\rho}}{\partial t} + \mathrm{i}\overline{H}_1\overline{\rho} = \mathrm{i}\overline{D}_1\rho_D \tag{2.143}$$

$$\overline{H}_2\overline{\rho} + \overline{H}_3\mathbf{q}_x = \overline{\rho}_2\rho_D \tag{2.144}$$

Dabei sind H_i und D_i die Eingangs beschriebenen Systemmatrizen. Für den stationären Fall gilt

$$\begin{pmatrix} \mathrm{i}\overline{H}_1 & 0 \\ \overline{H}_2 & \overline{H}_3 \end{pmatrix}\begin{pmatrix} \rho \\ \mathbf{q}_x \end{pmatrix} = \begin{pmatrix} \mathrm{i}\overline{D}_1\rho_D \\ \overline{D}_2\rho_D \end{pmatrix} \tag{2.145}$$

Dieses Gleichungssystem ist noch nicht in der Lage, dass in ?? beschriebene Modellproblem zu lösen. Die Randbedingungen im Impulsraum müssen durch partielle Transformation mit einer diskreten Fouriertransformation formuliert und das Gleichungssystem entsprechend formuliert werden. Dazu wurden bereits in Abschnitt 2.1.1 einige Annahmen gemacht um sicherzustellen, dass eine solche bijektive Abbildung in den Phasenraum existiert.

Randbedingungen

Bei der Betrachtung des linken oder rechten Randes Γ_D^L oder Γ_D^R des Gebiets Ω können zwei Mengen an Knoten auf diesen Rändern identifiziert werden, die single-valued Punkte Γ_d und die double-valued Punkte Γ_d auf denen die Lösung ρ entweder einfach oder zweifach definiert ist. Es sei die Indexmengen

$$S = \{n \in \mathbb{N} \mid y_n \in \Gamma_S, n \leq N_y\} \tag{2.146}$$
$$D = \{n \in \mathbb{N} \mid y_n \in \Gamma_D, n \leq N_y\} \tag{2.147}$$

mit einer globalen Punktenummerierung $\{\mathbf{x}_1, \ldots, \mathbf{x}_{N_y}\}$. Dann wird die Fouriertransformierte

$$\hat{\rho}_k^m = \sum_{j \in S} \phi_{k,j}\rho_j + \sum_{j \in D} \phi_{k,j}\rho_j^m \tag{2.148}$$

mit den Matrixelementen

$$\phi_{k,j} = \frac{1}{\sqrt{N_y}} e^{2\pi \mathrm{i} \frac{k \cdot j}{N_y}} \tag{2.149}$$

und $m = \{0, 1\}$, $k = 1, \ldots, N_y$ eingeführt. Der Parameter m indiziert dann einen der beiden Funktionswerte auf Punkte in Γ_d. Es ergeben sich also zwei verschiedene Folgen $\hat{\rho}_k^m$ bzw. Transformierte. Die Annahme, die auch bereits in Abschnitt ?? getroffen wurden, ist die Stetigkeit von ρ auf Γ_d, woraus sich $\hat{\rho}_k^1 = \hat{\rho}_k^2 \forall k = 1, \ldots, N_y$ als Konsequenz ergibt. Seien nun die globalen Indexmengen s und d so aufgebaut, dass sie die Indizes aller diskreten Funktionswerte von ρ auf dem Rand Γ_D^L oder Γ_D^R in einer sortierten Reihenfolge enthalten. Damit sind auch doppelte Funktionswerte auf Γ_d enthalten. Die Vektoren

$$\rho = \left(\rho^{E_1}(\mathbf{x}_1), \ldots, \rho^{E_{N_y}}(\mathbf{x}_{N_y})\right)^T \tag{2.150}$$
$$\hat{\rho} = \left(\hat{\rho}(x_1, k_1), \ldots, \hat{\rho}(x_{N_y}, k_{N_y})\right)^T \tag{2.151}$$

beinhalten die Elemente ρ_j^m und $\hat{\rho}_k$ in einer ebenfalls sortierten Reihenfolge. Es soll nun eine invertierbare $k_{\max,y} \cdot N_{py} \times k_{\max,y} \cdot N_{py}$-Transformationsmatrix F konstruiert werden mit

$$\hat{\rho} = H\rho \tag{2.152}$$

Die Konstruktion von F ist nicht eindeutig, eine mögliche Vorschrift könnte

$$\begin{aligned} H(s, s) &= \phi(S, S) \\ H(s, d_1) &= \phi(S, D) \\ H(d_1, s) &= \phi(D, S) \\ H(d_1, d_1) &= \phi(D, D) \\ H(d_2, s) &= \phi(D, S) \\ H(d_2, d_2) &= \phi(D, D) \end{aligned}$$

sein. Mithilfe dieser Transformation und dem neu eingeführten Vektor

$$\rho = \left(\hat{\rho}^{K_1}, \rho^{K_2}, \ldots, \rho^{K_{k_{\max}-1}}, \hat{\rho}^{K_{k_{\max}}}\right) \tag{2.153}$$

kann ein neues Gleichungssystem eingeführt werden.

2.2. Die Charakteristiken-Methode

Man betrachte erneut das Modellproblem

$$\frac{\partial\rho}{\partial t} - \mathrm{i}\frac{\partial^2\rho}{\partial x\partial y} + \mathrm{i}B\rho = 0 \tag{2.154}$$

mit denselben Randbedingungen wie in ?? beschrieben. Das Gebiet Ω ist auch hier wieder rechteckig und es wird die selbe Triangulation $\mathcal{T}_h$ wie in den vorherigen Abschnitten verwendet. In einer festen Orthonormalbasis sind somit alle Richtungsvektoren unabhängig voneinander. Dadurch lässt sich die räumliche Diskretisierung in zwei unabhängigen Schritten durchführen. Die resultierenden Gleichungssysteme ergeben sich dann als Tensorprodukt der einzelnen Systeme aus den beiden Schritten. Sei also durch ein numerisches Verfahren eine Diskretisierung in y-Richtung vorgenommen, kann das lineare System gekoppelter eindimensionaler Differentialgleichungen in der Form

$$A\rho_t\left(x,t\right) + B\rho_x\left(x,t\right) + C\left(x,t\right)\rho\left(x,t\right) = 0 \tag{2.155}$$

geschrieben werden. Die Matrix B ist der diskretisierte Ableitungsoperator $-\mathrm{i}\partial_y$ zu den Randbedingungen $\rho|_{\Gamma_D^0} = 0$. Die gekoppelten Gleichungen können durch die Eigenwertzerlegung $A^{-1}B = UDU^{-1}$ entkoppelt werden. TODO: Eigenwerte
Dadurch transformiert sich Gleichung ?? in die Form

$$\mathbf{u}_t + D\mathbf{u}_x + G\left(x,t\right)u = 0 \tag{2.156}$$

mit $\mathbf{u} = U^{-1}\rho$ und $G\left(x,t\right) = U^{-1}C\left(x,t\right)U$. Für diese Gleichung im Eigenraum soll nun ein 1D-Finite-Elemente-Verfahren entwickelt werden. Dies erfolgt analog zu ?? und lautet entsprechend für ein $\mathbf{u} \in \left[S_{hp}\right]^n$

$$\int_K \left(\mathbf{u}_t\cdot\overline{\mathbf{v}} + D\mathbf{u}\cdot\overline{\mathbf{v}}_x + \tilde{G}\mathbf{u}\cdot\overline{\mathbf{v}}\right) = \int_{\partial K} n_x\left(D\mathbf{u} - \left(D\mathbf{u}\right)^*\right)\cdot\overline{\mathbf{v}}, \qquad \forall\mathbf{v}\left[S_{hp}\right]^n \tag{2.157}$$

Der Fluss wird (TODO:Begründung) zu

$$\left(D\mathbf{u}\right)^* = \begin{cases} D\left\{\mathbf{u}\right\} + \frac{1}{2}\left|D\right|\left[\mathbf{u}\right], & \text{auf } \Gamma_I \\ \frac{1}{2}D\left(u+u_D\right) + \frac{1}{2}\left|D\right|n_x\left(u-u_D\right) & \text{auf } \Gamma_D^L\cup\Gamma_D^R \end{cases} \tag{2.158}$$

Analog zu ?? wird das Element K mit der Abbildung

$$\gamma_{\hat{K}\to K}\left(r\right) = \frac{h^K}{2}\left(t+1\right) + N_1 \tag{2.159}$$

auf ein Referenzelement $\hat{K}$ abgebildet, wobei N_1 eines der linke Eckpunkte des Intervalls $\left[N_1,N_2\right]$ ist. Wähle als Testfunktion $\mathbf{v} = l_\mathrm{m}\hat{e}_\mathrm{i}$ mit $i = 1,\ldots,n$, wobei e_i der i-te Einheitsvektor ist. Läuft der Index i von 1 bis n, ergeben sich wieder n Gleichungen mit einer skalaren Testfunktion l_m, so dass

$$\frac{h^K}{2}\int_{\hat{K}}\left(\hat{\mathbf{u}}_t\hat{l}_m + \frac{2}{h^K}D\hat{\mathbf{u}}\partial_{\hat{x}}\hat{l}_m + \tilde{G}\hat{\mathbf{u}}\hat{l}_m\right) = \int_{\partial\hat{K}} n_x\left(D\hat{\mathbf{u}} - \left(D\hat{\mathbf{u}}\right)^*\right)\hat{l}_m \tag{2.160}$$

Eine Interpolation auf den Punkten $\{\tau_n\}_{n=1,\dots,N_p}$

$$\hat{\mathbf{u}}(r,t) = \sum_{n=1}^{N_p} \hat{\mathbf{u}}(\tau_n,t)\,\hat{l}_n(r,t) \tag{2.161}$$

liefert wieder die Massematrix M sowie die Ableitungsmatrix D_t mit

$$M_{m,n} = \int_{\hat{K}} \hat{l}_m \hat{l}_n \tag{2.162}$$

$$(D_t)_{m,n} = \left.\frac{d\hat{l}_n}{dt}\right|_{\tau_m} \tag{2.163}$$

Für linke Kanten folgt für das Oberflächenintegral

$$\int_{e_l} n_x \left(D\hat{\mathbf{u}} - (D\hat{\mathbf{u}})^*\right)\hat{l}_m = \sum_{n=1}^{N_p} \begin{cases} F^+\left((P_3)_{m,n}\,\mathbf{u}_n^{k-1} - (P_2)_{m,n}\,\mathbf{u}_n^k\right) & e_l \in \Gamma_I \\ F^+\,(P_2)_{m,n}\left(\mathbf{u}_n^1 - \mathbf{u}_{D,n}^1\right) & e_l \in \Gamma_D^L \end{cases} \tag{2.164}$$

und für rechte Kanten

$$\int_{e_r} n_x \left(D\hat{\mathbf{u}} - (D\hat{\mathbf{u}})^*\right)\hat{l}_m = \sum_{n=1}^{N_p} \begin{cases} F^-\left((P_1)_{m,n}\,\mathbf{u}_n^{k} - (P_4)_{m,n}\,\mathbf{u}_n^{k+1}\right) & e_l \in \Gamma_I \\ F^-\,(P_1)_{m,n}\left(\mathbf{u}_n^{k_{\max}} - \mathbf{u}_{D,n}^{k_{\max}}\right) & e_l \in \Gamma_D^R \end{cases} \tag{2.165}$$

Dabei wurden die beiden Flussfunktionen

$$F^+ = \frac{1}{2}\left(D + |D|\right) \tag{2.166}$$

$$F^- = \frac{1}{2}\left(D - |D|\right) \tag{2.167}$$

und die Matrizen

$$(P_1)_{m,n} = \delta_{n,N_p}\delta_{m,N_p}, \quad (P_2)_{m,n} = \delta_{n,1}\delta_{m,1}, \quad (P_3)_{m,n} = \delta_{n,N_p}\delta_{m,1}, \quad (P_4)_{m,n} = \delta_{n,1}\delta_{m,N_p} \tag{2.168}$$

eingeführt. Insgesamt ergibt sich für das Gleichungssystem

$$\frac{h^K}{2}\left(M \otimes I_{N_p}\right)\mathbf{u}_t^k + \left(\left((S_r)^T \otimes D\right) + G^k + \left(P_2 \otimes F^+\right)\right. \tag{2.169}$$

$$\left.- \left(P_1 \otimes F^-\right)\right)\mathbf{u}^k - \left(P_3 \otimes F^+\right)\mathbf{u}^{k-1} + \left(P_4 \otimes F^-\right)\mathbf{u}^{k+1} = 0, \qquad k = 2,\dots,k_{\max}-1 \tag{2.170}$$

$$\frac{h^1}{2}\left(M \otimes I_{N_p}\right)\mathbf{u}_t^k + \left(\left((S_r)^T \otimes D\right) + G^k + \left(P_2 \otimes F^+\right)\right. \tag{2.171}$$

$$\left.- \left(P_1 \otimes F^-\right)\right)\mathbf{u}^1 + \left(P_4 \otimes F^-\right)\mathbf{u}^2 = \left(P_2 \otimes F^+\right)\mathbf{u}_D^1, \qquad k = 1 \tag{2.172}$$

$$\frac{h^{k_{\max}}}{2}\left(M \otimes I_{N_p}\right)\mathbf{u}_t^k + \left(\left((S_r)^T \otimes D\right) + G^k + \left(P_2 \otimes F^+\right)\right. \tag{2.173}$$

$$\left.- \left(P_1 \otimes F^-\right)\right)\mathbf{u}^{k_{\max}} - \left(P_3 \otimes F^+\right)\mathbf{u}^{k_{\max}-1} = \left(P_1 \otimes F^-\right)\mathbf{u}_D^{k_{\max}}, \qquad k = k_{\max} \tag{2.174}$$

Eine etwas kompaktere Form ist durch die Umschreibung und Zusammenfassung durch

$$H\mathbf{u}_t + \Big[I^0_{k_{\max}} \otimes M^{-1}\left(\left(P_2 - P_1 + 2S_r^T\right) \otimes D + M^{-1}\left(P_1 + P_2\right) \otimes |D|\right) \tag{2.175}$$

$$- I^{-1}_{k_{\max}} \otimes M^{-1} P_3 \otimes (D + |D|) + I^1_{k_{\max}} \otimes M^{-1} P_4 \otimes (D - |D|) + G\Big]\,\mathbf{u} = \mathbf{L} \tag{2.176}$$

gegeben. Dabei ist

$$H = \operatorname{diag}\left(h^{K_1} I_{N_p}, \dots, h^{K_{k_{\max}} I_{N_p}}\right) \tag{2.177}$$

$$G = \operatorname{diag}\left(G^{K_1}, \dots, G^{K_{k_{\max}}}\right) \tag{2.178}$$

$$\tag{2.179}$$

und

$$\mathbf{L} = \left(\left(P_2 \otimes F^+\right) U^{-1} \rho_D, 0, \dots, 0, \left(P_1 \otimes F^-\right) U^{-1} \rho_D\right)^T \tag{2.180}$$

$$\mathbf{u} = \left(\mathbf{u}^{K_1}, \dots, \mathbf{u}^{K_{k_{\max}}}\right)^T \tag{2.181}$$

Summiere nun Gleichung ?? über alle Elemente, so dass

$$(\mathbf{u}_t, \mathbf{v})_{\Omega,h} + (D\mathbf{u} + G\mathbf{u}, \mathbf{v})_{\Omega,h} = \tag{2.182}$$

Eine Summation von Gleichung ?? über alle Elemente liefert die Gleichung

$$(\mathbf{u}_t, \mathbf{v})_{\Omega,h} + \mathcal{B}_h(\mathbf{u}, \mathbf{v}) = \mathcal{L}(\mathbf{v}) \tag{2.183}$$

mit der Sesquilinearform

$$\mathcal{B}_h(\mathbf{u}, \mathbf{v}_x) = (D\mathbf{u}, \mathbf{v})_{\Omega,h} + (G\mathbf{u}, \mathbf{v})_{\Omega,h} + \sum_{e\in\Gamma} \int_e \left(\frac{1}{2}\left[|D|\,\mathbf{u}\right] \cdot [\overline{\mathbf{v}}] - [D\mathbf{u}]\,\{\overline{\mathbf{v}}]\right) \tag{2.184}$$

und der Linearform

$$\mathcal{L}(\mathbf{v}) = \frac{1}{2} \sum_{e\in\Gamma_D} (|D| - nD)\,\mathbf{u}_D \cdot \overline{\mathbf{v}} \tag{2.185}$$

Dabei wurde erneut die Beziehung ?? ausgenutzt. Nun wird die komplex konjugierte Gleichung auf ?? hinzuaddiert mit $\mathbf{v} = \mathbf{u}$, so dass

$$\frac{1}{2}\frac{d}{dt}\,||\mathbf{u}||^2_{\Omega,h} + \sum_{K\in\mathcal{T}_h} \int_K \overline{\mathbf{u}} \cdot \left(G + G^\dagger\right) \mathbf{u} + \sum_{e\in\Gamma} \int_e \left[||D||\,\mathbf{u}\right] [\overline{\mathbf{u}}] = 0 \tag{2.186}$$

Damit ist gezeigt, dass das Verfahren stabil ist. Es wurde die Beziehung

$$(D\mathbf{u}, \mathbf{u}_x)_{\Omega,h} + (\mathbf{u}, D\mathbf{u}_x)_{\Omega,h} - \sum_{e\in\Gamma} \int_e \left([D\mathbf{u}]\,\{\overline{\mathbf{u}}] + [D\overline{\mathbf{u}}]\,\{\mathbf{u}]\right) = 0 \tag{2.187}$$

ausgenutzt. Auch mit der Form

$$\Re\mathcal{B}_h(\mathbf{u}, \mathbf{u}) = \Re\,(G\mathbf{u}, \mathbf{u})_{\Omega,h} + \frac{1}{2}\sum_{e\in\Gamma} \int_e \left[|D|\,\mathbf{u}\right] \cdot [\overline{\mathbf{u}}] \geq G_0\,||\mathbf{u}||^2_{\Omega,h} + \frac{1}{2}\sum_{e\in\Gamma} \int_e \left[|D|\,\mathbf{u}\right] \cdot [\overline{\mathbf{u}}] \tag{2.188}$$

$$\geq \sum_{K\in\mathcal{T}_h} \int_e \overline{\mathbf{u}} \cdot \left(I_n C h_K G_0 + \frac{1}{8}(1-\delta)\,|||D|||\right) \mathbf{u} \geq \kappa |||\mathbf{u}|||^2 \tag{2.189}$$

ist die Stabilität mit der Norm

$$|||\mathbf{u}|||^2 = \sum_{e \in \Gamma} \int_e ||\mathbf{u}||^2 \tag{2.190}$$

und der Konstanten $\kappa = \min_{j=1,\dots,n} \left(ChG_0 + \frac{1-\delta}{8} |D_{n,n}| \right)$ des stationärem Problems unter der Voraussetzung $G_0 \geq \frac{1-\delta}{8Ch} |||D|||_{\max}$ gegeben aus der Abschätzung $G_0 \left(\mathbf{u}, \mathbf{u}\right)_{\Omega,h} \leq \Re \left(G\mathbf{u}, \mathbf{u}\right)_{\Omega,h} \leq G_1 \left(\mathbf{u}, \mathbf{u}\right)_{\Omega,h}$. Für den letzten Schritt wurde wieder die Abschätzung ?? verwendet. Auch Stetigkeit lässt sich sofort an ?? ablesen. Anhand von Gleichung ?? lässt sich auch ablesen, warum ein Eigenwert $d_{\mathrm{j}} = 0$ zur Singularität des Verfahrens führen würde. Für einen solchen Eigenwert würde sich Gleichung ?? mit der entsprechenden Zeile $\mathbf{g}_j$ von G auf $\mathbf{g}_{\mathrm{j}}\mathbf{u} = 0$ reduzieren. Damit wäre $\mathbf{u}$ aber nicht eindeutig wählbar und das resultierende Gleichungssystem singulär. Die Diskretisierung in y-Richtung ist also symmetrisch um 0 zu wählen, so dass sich auch ein symmetrisches Spektrum ergibt.