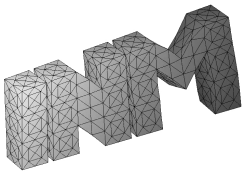


Masterarbeit

Eine Finite Elemente Methode für optimale Kontrollprobleme mit parabolischen Randwertaufgaben

vorgelegt der Fakultät für Technische Mathematik und Technische Physik
der Technischen Universität Graz
zur Erlangung des akademischen Grades
Diplom-Ingenieur (Dipl.-Ing.)

von

Martin Neumüller

Betreuung: Prof. Dr. O. Steinbach

Institut für Numerische Mathematik
Technische Universität Graz

2010

Masterarbeit:

Titel:	Eine Finite Elemente Methode für optimale Kontrollprobleme mit parabolischen Randwertaufgaben
Name, Vorname:	Neumüller, Martin
Matrikelnummer:	0530707
Lehrveranstaltung:	Masterarbeit
Institut:	Institut für Numerische Mathematik Technische Universität Graz
Betreuung:	Prof. Dr. O. Steinbach

Vorwort

An dieser Stelle möchte ich mich bei Herrn Prof. Dr. Steinbach bedanken, dass ich dieses interessante Thema bearbeiten durfte. Weiters bedanke ich mich für die gute Betreuung, da er trotz vollen Terminplans immer Zeit für mich gefunden hat.
Weiters möchte ich mich bei meinen Eltern bedanken, die mir den Weg bis zum Studium ermöglicht haben. Ferner bedanke ich mich bei meiner Freundin Kornelia.

Inhaltsverzeichnis

Einleitung

In dieser Arbeit wird ein optimales Kontrollproblem mit verteilter Steuerung betrachtet. Optimale Kontrollprobleme treten dann auf, wenn die Zielgröße eines bestimmten Vorgangs minimiert werden soll. Dabei wird der Vorgang durch eine partielle Differentialgleichung, nämlich der Zustandsgleichung beschrieben.
Ein Beispiel für solch ein Kontrollproblem ist gegeben, wenn ein Objekt zu einem gewissen Zeitpunkt mit Hilfe der Steuerung (Abkühlung beziehungsweise Erwärmung) auf eine gewünschte Temperaturverteilung gebracht werden soll. Man spricht dabei von einer verteilten Steuerung, wenn die Steuerung auf das ganze Objekt wirkt. Das heißt, dass das Objekt im Inneren erwärmt beziehungsweise abgekühlt wird. Die Temperaturverteilung wird mithilfe der Zustandsgleichung an die Steuerung gekoppelt. Ziel ist es nun, die gewünschte Temperaturverteilung bestmöglich zu erreichen und dabei die Kosten der Steuerung zu minimieren. Falls nur ein begrenztes Erwärmen beziehungsweise Abkühlen möglich ist, sind weitere Restriktionen an die Steuerung zu fordern. Diese werden durch eine weitere Nebenbedingung modelliert. Zur Theorie von optimalen Kontrollproblemen sei hier auf [14, 18, 31] verwiesen.
Zur näherungsweisen Berechnung der Lösung der Zustandsgleichung, welche eine parabolische partielle Differentialgleichung ist, kann zum Beispiel die Discontinuous Galerkin Methode im ganzen Raum–Zeit–Zylinder verwendet werden. Die Discontinuous Galerkin Methoden wurden anfangs für hyperbolische Probleme verwendet, siehe [8, 16]. Weiters wurden die Discontinuous Galerkin Methoden erfolgreich zur näherungsweisen Berechnung von elliptischen Problemen herangezogen, siehe [2, 12, 23]. In [29] wird für zeitabhängige Advektions–Diffusions–Probleme ein Ansatz mit der Discontinuous Galerkin Methode vorgestellt.
Die Anwendung der Discontinuous Galerkin Methode im ganzen Raum–Zeit–Zylinder erfordert für dreidimensionale Gebiete die Zerlegung des Raum–Zeit–Zylinders im vierdimensionalen Raum. In [5] wird eine Möglichkeit beschrieben, wie der Raum–Zeit–Zylinder in Pentatope zerteilt werden kann. Die Pentatope können dann mit dem Algoritmus von Freudenthal, siehe [6], weiter verfeinert werden. Eine andere Vorgehensweise wurde in der Masterprojektarbeit [21] beschrieben.

Diese Arbeit ist in vier Teile gegliedert. Im ersten Kapitel wird die Problemstellung eines optimalen Kontrollproblems beschrieben. Weiters werden die notwendigen Räume eingeführt, in denen die eindeutige Lösbarkeit der Zustandsgleichung gewährleistet ist. Ferner wird die eindeutige Lösbarkeit des optimalen Kontrollproblems gezeigt. Anschließend wird ein Optimalitätssystem hergeleitet, welches eine äquivalente Formulierung zum optimalen Kontrollproblem darstellt. Im zweiten Kapitel wird das Optimalitätsystem mittels der

Discontinuous Galerkin Methode diskretisiert. Weiters wird die eindeutige Lösbarkeit der diskreten Zustandsgleichung untersucht. Im dritten Kapitel wird die Zerlegung von vierdimensionalen Objekten behandelt. Dazu wird vorerst die Zerlegung eines Tetraeders untersucht und anschließend in ähnlicher Weise die Zerlegung eines Pentatops behandelt. Ferner wird eine mögliche Zerlegung eines Hyperwürfels angegeben und kurz auf die Visualisierung von vierdimensionalen Objekten eingegangen. Im vierten Kapitel werden numerische Beispiele für die Raumdimensionen $d = 1$ und $d = 3$ vorgestellt. Dabei ist jeweils die zu erwartende Konvergenzordnung zu beobachten.

1 Ein optimales Kontrollproblem

In diesm Kapitel wird die Problemstellung eines instationären Kontrollproblems mit verteilter Steuerung formuliert. Weiters wird für diese Aufgabenstellung die Existenz von Lösungen und deren Eindeutigkeit untersucht. Am Ende dieses Kapitels wird ein Optimalitätssystem hergeleitet, welches eine äquivalente Aufgabenstellung zum gegebenen instationären Kontrollproblem darstellt. Für einen Überblick zu optimalen Kontrollproblemen sei hier auf [31] verwiesen.

1.1 Modellproblem

Gegeben sei ein beschränktes Gebiet $\Omega \subset \mathbb{R}^d$ für $d = 1, 2, 3$ mit hinreichend glattem Rand $\Gamma := \partial\Omega$. Dieses Gebiet steht zum Beispiel für einen aufzuheizenden bzw. abzukühlenden Körper, der zum Zeitpunkt $T > 0$ eine gewünschte Temperaturverteilung $\bar{u} = \bar{u}(x)$ aufweisen soll. Dazu wird im Gebiet Ω die Steuerung $z = z(x,t)$ angelegt. Zu Beginn, also für $t = 0$, sei die Temperatur im Gebiet durch die Funktion $u_0 = u_0(x)$ gegeben. Ziel ist es nun, die gewünschte Temperaturverteilung $\bar{u}$ bestmöglich zu erreichen. Dabei sollen die Kosten der Steuerung z möglichst minimal sein. Dies führt auf die Minimierung des Funktionals

$$\mathcal{J}(u,z) := \frac{1}{2}\int_\Omega [u(x,T) - \bar{u}(x)]^2 \,\mathrm{d}x + \frac{1}{2}\varrho \int_0^T \int_\Omega [z(x,t)]^2 \,\mathrm{d}x\mathrm{d}t \tag{1.1}$$

unter der Nebenbedingung

$$\begin{aligned} \frac{\partial}{\partial t}u(x,t) - \Delta u(x,t) &= z(x,t) \qquad \text{für} \quad (x,t) \in Q := \Omega \times (0,T), \\ u(x,t) &= g(x,t) \qquad \text{für} \quad (x,t) \in \Sigma := \Gamma \times (0,T), \\ u(x,0) &= u_0(x) \qquad \text{für} \quad (x,0) \in \Sigma_0 := \Omega \times \{0\}\,. \end{aligned} \tag{1.2}$$

Dabei sind Dirichlet–Daten g auf dem Rand $\Gamma \times (0,T)$ vorgegeben und mit $\varrho \geq 0$ wird der Kostenkoeffizient der Steuerung bezeichnet. Ist die Steuerung zusätzlich nach unten mit z_a bzw. nach oben mit z_b beschränkt, so ist noch zusätzlich die Nebenbedingung

$$z_a(x,t) \leq z(x,t) \leq z_b(x,t) \quad \text{für } (x,t) \in Q \tag{1.3}$$

zu fordern. Im nächsten Abschnitt wird die eindeutige Lösbarkeit der Problemstellung (1.1)–(1.3) untersucht. Dazu werden geeignete Funktionenräume eingeführt, in denen die eindeutige Lösbarkeit der Zustandsgleichung (1.2) gewährleistet ist.

1.2 Eindeutige Lösbarkeit der Zustandsgleichung

In diesem Abschnitt werden die Räume eingeführt, die die Existenz von Lösungen der Zustandsgleichung (1.2) und deren Eindeutigkeit sicherstellen. Für ein weiteres Studium von diesen Räumen sei hier auf [11, 33, 35] verwiesen. Anschließend wird die eindeutige Lösbarkeit des optimalen Kontrollproblems (1.1)–(1.3) untersucht.

Definition 1.1. *Sei X ein Banachraum und $T > 0$. Für eine Funktion $v : (0,T) \to X$ sei*

$$\|v\|_{L_2(0,T;X)} := \left[\int_0^T \|v(\cdot\,,t)\|_X^2 \mathrm{d}t\right]^{1/2}.$$

Dann wird durch

$$L_2(0,T;X) := \left\{v : (0,T) \to X : \|v\|_{L_2(0,T;X)} < \infty\right\}$$

ein Banachraum erklärt, siehe [11, Satz 1.11].

Definition 1.2 (Verallgemeinerte schwache Zeitableitung)**.** *Sei $v \in L_2(0,T;X)$. Dann heißt $\frac{\partial}{\partial t}v \in L_2(0,T;X^*)$* verallgemeinerte schwache Zeitableitung*, falls*

$$\int_0^T v(t)\frac{\partial}{\partial t}\varphi(t)\mathrm{d}t = -\int_0^T \frac{\partial}{\partial t}v(t)\varphi(t)\mathrm{d}t \quad \textit{für alle } \varphi \in \mathcal{C}_0^\infty(0,T)$$

gilt.

Definition 1.3 (Gelfandscher Dreier)**.** *Sei V ein seperabler, reflexiver Banachraum und H ein seperabler Hilbertraum. Weiters sei V dicht in H eingebettet mit $\|v\|_H \le c\|v\|_V$ für alle $v \in V$ mit $c > 0$. Dann wird durch das Tripel*

$$V \subset H = H^* \subset V^*$$

ein Gelfandscher Dreier *definiert.*

Beispiel 1.1. *Für $V = H_0^1(\Omega)$ und $H = L_2(\Omega)$ ist durch*

$$H_0^1(\Omega) \subset L_2(\Omega) \subset H^{-1}(\Omega)$$

ein Gelfandscher Dreier gegeben.

Definition 1.4 (Sobolev–Raum)**.** *Sei $V \subset H \subset V^*$ ein Gelfandscher Dreier. Dann wird durch*

$$W_2^1(0,T;V,H) := \left\{v \in L_2(0,T;V) : \frac{\partial}{\partial t}v \in L_2(0,T;V^*)\right\}$$

ein Sobolev–Raum *erklärt. Mit der Norm*

$$\|v\|_{W_2^1(0,T;V,H)} := \left[\|v\|_{L_2(0,T;V)}^2 + \left\|\frac{\partial}{\partial t}v\right\|_{L_2(0,T;V^*)}^2\right]^{\frac{1}{2}},$$

wird der Raum $W_2^1(0,T;V,H)$ zu einem Banachraum, siehe [11, Satz 1.16].

Definition 1.5. *Sei X ein Banachraum und $[a,b] \subset \mathbb{R}$. Eine Funktion $v : [a,b] \to X$ heißt stetig, falls*

$$\lim_{\tau \to t} \|v(\tau) - v(t)\|_X = 0 \quad \text{für alle } t \in [a,b]$$

gilt.

Definition 1.6. *Sei X ein Banachraum. Dann wird für $[a,b] \subset \mathbb{R}$ mit $\mathcal{C}([a,b],X)$ der Raum aller stetigen Funktionen $v : [a,b] \to X$ mit der Norm*

$$\|v\|_{\mathcal{C}([a,b],X)} := \sup_{t \in [a,b]} \|v(t)\|_X$$

bezeichnet.

Satz 1.7. *Sei $v \in W_2^1(0,T;V,H)$. Dann kann v bis auf eine Menge vom Maß Null als Funktion aus $\mathcal{C}([0,T],H)$ angenommen werden. Weiters ist die Einbettung*

$$W_2^1(0,T;V,H) \hookrightarrow \mathcal{C}([0,T],H)$$

eine stetige Einbettung, das heißt

$$\sup_{t \in [0,T]} \|v(t)\|_H \leq c_T \|v\|_{W_2^1(0,T;V,H)} \quad \text{für alle } v \in W_2^1(0,T;V,H) \quad \text{mit } c_T > 0.$$

Beweis. Siehe zum Beispiel [11, Satz 1.17] oder [33, Theorem 25.5]. □

Satz 1.7 motiviert nun die folgenden Definitionen:

Definition 1.8 (Spuroperatoren)**.** *Für ein Gebiet Ω und eine Zeit $T > 0$ werden die folgenden Spuren im Raum–Zeit–Zylinder $Q = \Omega \times (0,T)$ definiert:*

$$\begin{aligned}
\gamma_{0,x}^{\text{int}} v(x,t) &:= \lim_{Q \ni (\tilde{x},\tilde{t}) \to (x,t) \in \Sigma} v(\tilde{x},\tilde{t}) && \text{für alle } (x,t) \in \Sigma, \\
\gamma_{1,x}^{\text{int}} v(x,t) &:= \lim_{Q \ni (\tilde{x},\tilde{t}) \to (x,t) \in \Sigma} \underline{n}_x \cdot \nabla_{\tilde{x}} v(\tilde{x},\tilde{t}) && \text{für alle } (x,t) \in \Sigma, \\
\gamma_{0,t}^{\text{int}} v(x,0) &:= \lim_{Q \ni (\tilde{x},\tilde{t}) \to (x,0) \in \Sigma_0} v(\tilde{x},\tilde{t}) && \text{für alle } (x,0) \in \Sigma_0, \\
\gamma_{T,t}^{\text{int}} v(x,T) &:= \lim_{Q \ni (\tilde{x},\tilde{t}) \to (x,T) \in \Sigma_T} v(\tilde{x},\tilde{t}) && \text{für alle } (x,T) \in \Sigma_T := \Omega \times \{T\}.
\end{aligned}$$

Satz 1.9 (Spursatz)**.** *Sei der Raum $W_2^1(0,T;V,H)$ gegeben. Die Spuroperatoren*

$$\begin{aligned}
\gamma_{0,t}^{\text{int}} &: W_2^1(0,T;V,H) \to H, \\
\gamma_{T,t}^{\text{int}} &: W_2^1(0,T;V,H) \to H
\end{aligned}$$

sind linear und beschränkt, das heißt es exisitert eine positive Konstante c_T, sodass

$$\begin{aligned}
\left\|\gamma_{0,t}^{\text{int}} v\right\|_H &\leq c_T \|v\|_{W_2^1(0,T;V,H)}, \\
\left\|\gamma_{T,t}^{\text{int}} v\right\|_H &\leq c_T \|v\|_{W_2^1(0,T;V,H)}
\end{aligned}$$

für alle $v \in W_2^1(0,T;V,H)$ gilt.

Beweis. Sei $v \in W_2^1(0,T;V,H)$. Die Behauptung folgt direkt aus Satz 1.7 durch

$$\left\|\gamma_{0,t}^{\text{int}} v\right\|_H \leq \sup_{t\in[0,T]} \|v(t)\|_H \leq c_T \|v\|_{W_2^1(0,T;V,H)},$$
$$\left\|\gamma_{T,t}^{\text{int}} v\right\|_H \leq \sup_{t\in[0,T]} \|v(t)\|_H \leq c_T \|v\|_{W_2^1(0,T;V,H)}.$$

□

In $W_2^1(0,T;V,H)$ ist weiters eine Formel der partiellen Integration erfüllt:

Satz 1.10 (Partielle Integration)**.** *Seien $u, v \in W_2^1(0,T;V,H)$. Dann gilt*

$$\int_0^T \left\langle \frac{\partial}{\partial t} u(\cdot\,,t), v(\cdot\,,t) \right\rangle_{V^*\times V} \mathrm{d}t = \left\langle \gamma_{T,t}^{\text{int}} u, \gamma_{T,t}^{\text{int}} v \right\rangle_H - \left\langle \gamma_{0,t}^{\text{int}} u, \gamma_{0,t}^{\text{int}} v \right\rangle_H - \int_0^T \left\langle \frac{\partial}{\partial t} v(\cdot\,,t), u(\cdot\,,t) \right\rangle_{V^*\times V} \mathrm{d}t.$$

Beweis. Siehe [11, Satz 1.17]. □

Satz 1.11 (Spursatz)**.** *Sei $\Omega \subset \mathbb{R}^d$ ein $C^{k-1,1}$–Gebiet und $T > 0$. Dann ist für $\frac{1}{2} < s \leq k$ der Operator*

$$\gamma_{0,x}^{\text{int}} : L_2(0,T;H^s(\Omega)) \to L_2(0,T;H^{s-1/2}(\Gamma))$$

ein linearer und beschränkter Operator, das heißt es gibt eine positive Konstante c_T, sodass

$$\left\|\gamma_{0,x}^{\text{int}} v\right\|_{L_2(0,T;H^{s-1/2}(\Gamma))} \leq c_T \|v\|_{L_2(0,T;H^s(\Omega))} \quad \textit{für alle } v \in L_2(0,T;H^s(\Omega))$$

gilt.

Beweis. Nach Voraussetzung ist der Spuroperator

$$\gamma_0^{\text{int}} : H^s(\Omega) \to H^{s-1/2}(\Gamma) \quad \text{mit} \quad \gamma_0^{\text{int}} v(x) := \lim_{\Omega\ni\tilde{x}\to x\in\Gamma} v(\tilde{x}) \quad \text{für alle } x \in \Gamma$$

ein linearer und beschränkter Operator, siehe [33, Theorem 8.7]. Also gibt es eine positive Konstante c_T mit

$$\left\|\gamma_0^{\text{int}} v\right\|_{H^{s-1/2}(\Gamma)} \leq c_T \|v\|_{H^s(\Omega)} \quad \text{für alle } v \in H^s(\Omega).$$

Für ein beliebiges $t \in (0,T)$ und $v \in L_2(0,T;H^s(\Omega))$ ist $v(\cdot\,,t) \in H^s(\Omega)$ und es gilt

$$\begin{aligned}
\gamma_{0,x}^{\text{int}} v(x,t) &= \lim_{Q\ni(\tilde{x},\tilde{t})\to(x,t)\in\Sigma} v(\tilde{x},\tilde{t}) \\
&= \lim_{Q\ni(\tilde{x},t)\to(x,t)\in\Sigma} v(\tilde{x},t) \\
&= \lim_{\Omega\ni\tilde{x}\to x\in\Gamma} v(\tilde{x},t) \\
&= \gamma_0^{\text{int}} v(x,t) \quad \text{für alle } x \in \Gamma.
\end{aligned}$$

Somit ist $\gamma_{0,x}^{\text{int}} v(\cdot\,,t) \in H^{s-1/2}(\Gamma)$ und man erhält

$$\left\|\gamma_{0,x}^{\text{int}} v(\cdot\,,t)\right\|_{H^{s-1/2}(\Gamma)} = \left\|\gamma_0^{\text{int}} v(\cdot\,,t)\right\|_{H^{s-1/2}(\Gamma)} \leq c_T \|v(\cdot\,,t)\|_{H^s(\Omega)}.$$

Mit

$$\begin{aligned}\left\|\gamma_{0,x}^{\text{int}} v\right\|_{L_2(0,T;H^{s-1/2}(\Gamma))} &= \left[\int_0^T \left\|\gamma_{0,x}^{\text{int}} v(\cdot\,,t)\right\|_{H^{s-1/2}(\Gamma)}^2 \mathrm{d}t\right]^{\frac{1}{2}} \\ &\leq c_T \left[\int_0^T \|v(\cdot\,,t)\|_{H^s(\Omega)}^2 \mathrm{d}t\right]^{\frac{1}{2}} = c_T \|v\|_{L_2(0,T;H^s(\Omega))},\end{aligned}$$

folgt die Behauptung des Satzes. □

Analog lässt sich der nächste Satz beweisen.

Satz 1.12 (Inverser Spursatz)**.** *Sei $\Omega \subset \mathbb{R}^d$ ein $C^{k-1,1}$–Gebiet und $T > 0$. Der Spuroperator $\gamma_{0,x}^{\text{int}} : L_2(0,T;H^s(\Omega)) \to L_2(0,T;H^{s-1/2}(\Gamma))$ besitzt für $\frac{1}{2} < s \leq k$ eine stetige Rechtsinverse*

$$\mathcal{E} : L_2(0,T;H^{s-1/2}(\Gamma)) \to L_2(0,T;H^s(\Omega))$$

mit $\gamma_{0,x}^{\text{int}} \mathcal{E} w = w$ für alle $w \in L_2(0,T;H^{s-1/2}(\Gamma))$ und

$$\|\mathcal{E} w\|_{L_2(0,T;H^s(\Omega))} \leq c_{IT} \|w\|_{L_2(0,T;H^{s-1/2}(\Gamma))} \quad \textit{für alle } w \in L_2(0,T;H^{s-1/2}(\Gamma)).$$

Beweis. Nach [33, Theorem 8.8] existiert laut den Voraussetzungen für den im Beweis von Satz 1.11 definierten Spuroperator $\gamma_0^{\text{int}} : H^s(\Omega) \to H^{s-1/2}(\Gamma)$, ein stetiger rechtsinverser Operator

$$\mathcal{E} : H^{s-1/2}(\Gamma) \to H^s(\Omega),$$

mit $\gamma_0^{\text{int}} \mathcal{E} g = g$ für alle $g \in H^{s-1/2}(\Gamma)$ und

$$\|\mathcal{E} g\|_{H^s(\Omega)} \leq c_{IT} \|g\|_{H^{s-1/2}(\Gamma)} \quad \text{für alle } g \in H^{s-1/2}(\Gamma).$$

Für ein beliebiges $t \in (0,T)$ und $w \in L_2(0,T;H^{s-1/2}(\Gamma))$ ist $w(\cdot\,,t) \in H^{s-1/2}(\Gamma)$. Somit ist

$$\mathcal{E} w(\cdot\,,t) \in H^s(\Omega)$$

und es gilt

$$\gamma_0^{\text{int}} \mathcal{E} w(\cdot\,,t) = w(\cdot\,,t) \quad \text{und} \quad \|\mathcal{E} w(\cdot\,,t)\|_{H^s(\Omega)} \leq c_{IT} \|w(\cdot\,,t)\|_{H^{s-1/2}(\Gamma)}.$$

Für $v(\cdot\,,t) \in H^s(\Omega)$ gilt $\gamma_{0,x}^{\text{int}} v(x,t) = \gamma_0^{\text{int}} v(x,t)$ für alle $x \in \Gamma$. Da weiters $t \in (0,T)$ beliebig war, folgt

$$\gamma_{0,x}^{\text{int}} \mathcal{E} w(x,t) = w(x,t) \quad \text{für alle } (x,t) \in \Sigma.$$

Da $w \in L_2(0,T;H^{s-1/2}(\Gamma)$ ist, gilt

$$\infty > \|w\|_{L_2(0,T;H^{s-1/2}(\Gamma))} = \left[\int_0^T \|w(\cdot\,,t)\|^2_{H^{s-1/2}(\Gamma)}\mathrm{d}t\right]^{\frac{1}{2}}$$
$$\geq \frac{1}{c_{IT}}\left[\int_0^T \|\mathcal{E}w(\cdot\,,t)\|^2_{H^s(\Omega)}\mathrm{d}t\right]^{\frac{1}{2}} = \frac{1}{c_{IT}}\|\mathcal{E}w\|_{L_2(0,T;H^s(\Omega))}.$$

Also ist $\mathcal{E}w \in L_2(0,T;H^s(\Omega))$ mit

$$\|\mathcal{E}w\|_{L_2(0,T;H^s(\Omega))} \leq c_{IT}\|w\|_{L_2(0,T;H^{s-1/2}(\Gamma))}.$$

□

Laut Satz 1.12 existiert für $g \in L_2(0,T;H^{1/2}(\Gamma))$ eine Fortsetzung $\mathcal{E}g \in L_2(0,T;H^1(\Omega))$. Es stellt sich nun die Frage, unter welchen Voraussetungen eine Fortsetzung in dem Raum $W_2^1(0,T;H^1(\Omega),L_2(\Omega)) \subset L_2(0,T;H^1(\Omega))$ existiert. Dazu wird das Bild von $\gamma_{0,x}^{\mathrm{int}}$ bezüglich $W_2^1(0,T;H^1(\Omega),L_2(\Omega))$, also

$$W_2^1(\Sigma) := \left\{\gamma_{0,x}^{\mathrm{int}}v : v \in W_2^1(0,T;H^1(\Omega),L_2(\Omega))\right\},$$

betrachtet. Dabei wird durch

$$\|g\|_{W_2^1(\Sigma)} := \inf_{\substack{v\in W_2^1(0,T;H^1(\Omega),L_2(\Omega))\\ \gamma_{0,x}^{\mathrm{int}}v=g}} \|v\|_{W_2^1(0,T;H^1(\Omega),L_2(\Omega))}$$

eine Norm auf $W_2^1(\Sigma)$ definiert. Somit ist per Definition der Spuroperator

$$\gamma_{0,x}^{\mathrm{int}} : W_2^1(0,T;H^1(\Omega),L_2(\Omega)) \to W_2^1(\Sigma)$$

surjektiv. Weiters gilt der folgende Satz:

Satz 1.13 (Spursatz)**.** *Der Spuroperator*

$$\gamma_{0,x}^{\mathrm{int}} : W_2^1(0,T;H^1(\Omega),L_2(\Omega)) \to W_2^1(\Sigma)$$

ist ein linearer und beschränkter Operator, das heißt es existiert eine Konstante $c_T > 0$, sodass

$$\left\|\gamma_{0,x}^{\mathrm{int}}v\right\|_{W_2^1(\Sigma)} \leq c_T\|v\|_{W_2^1(0,T;H^1(\Omega),L_2(\Omega))} \quad \textit{für alle } v \in W_2^1(0,T;H^1(\Omega),L_2(\Omega))$$

gilt.

Beweis. Sei $v \in W_2^1(0,T;H^1(\Omega),L_2(\Omega))$. Die Aussage des Satzes folgt direkt aus der Normdefinition von $W_2^1(\Sigma)$ durch

$$\left\|\gamma_{0,x}^{\text{int}} v\right\|_{W_2^1(\Sigma)} = \inf_{\substack{\tilde{v}\in W_2^1(0,T;H^1(\Omega),L_2(\Omega)) \\ \gamma_{0,x}^{\text{int}}\tilde{v}=\gamma_{0,x}^{\text{int}}v}} \|\tilde{v}\|_{W_2^1(0,T;H^1(\Omega),L_2(\Omega))} \leq \|v\|_{W_2^1(0,T;H^1(\Omega),L_2(\Omega))}$$

mit der Konstante $c_T = 1$. □

Satz 1.14 (Inverser Spursatz)**.** *Sei $g \in W_2^1(\Sigma)$. Dann existiert eine Fortsetzung*

$$\mathcal{E}g \in W_2^1(0,T;H^1(\Omega),L_2(\Omega))$$

mit $\gamma_{0,x}^{\text{int}}\mathcal{E}g = g$ und

$$\|\mathcal{E}g\|_{W_2^1(0,T;H^1(\Omega),L_2(\Omega))} \leq c_{IT}\|g\|_{W_2^1(\Sigma)} \quad \textit{mit } c_{IT} > 0.$$

Also ist $\mathcal{E} : W_2^1(\Sigma) \to W_2^1(0,T;H^1(\Omega),L_2(\Omega))$ ein linearer und beschränkter Operator.

Beweis. Sei $g \in W_2^1(\Sigma)$. Laut Definition gilt

$$\|g\|_{W_2^1(\Sigma)} = \inf_{\substack{v\in W_2^1(0,T;H^1(\Omega),L_2(\Omega)) \\ \gamma_{0,x}^{\text{int}}v=g}} \|v\|_{W_2^1(0,T;H^1(\Omega),L_2(\Omega))}.$$

Somit existiert eine Folge $(v_n)_{n\in\mathbb{N}} \subset W_2^1(0,T;H^1(\Omega),L_2(\Omega))$ mit $\gamma_{0,x}^{\text{int}}v_n = g$ und

$$\lim_{n\to\infty} \|v_n\|_{W_2^1(0,T;H^1(\Omega),L_2(\Omega))} = \inf_{\substack{v\in W_2^1(0,T;H^1(\Omega),L_2(\Omega)) \\ \gamma_{0,x}^{\text{int}}v=g}} \|v\|_{W_2^1(0,T;H^1(\Omega),L_2(\Omega))} = \|g\|_{W_2^1(\Sigma)}.$$

Sei nun $\varepsilon > 0$ unabhängig von g fest gewählt. Weil die Folge $\left(\|v_n\|_{W_2^1(0,T;H^1(\Omega),L_2(\Omega))}\right)_{n\in\mathbb{N}}$ konvergiert, gibt es ein $N \in \mathbb{N}$ mit

$$\left|\|v_N\|_{W_2^1(0,T;H^1(\Omega),L_2(\Omega))} - \|g\|_{W_2^1(\Sigma)}\right| \leq \varepsilon\|g\|_{W_2^1(\Sigma)}.$$

Daraus ergibt sich die Abschätzung

$$\|v_N\|_{W_2^1(0,T;H^1(\Omega),L_2(\Omega))} \leq (1+\varepsilon)\|g\|_{W_2^1(\Sigma)}.$$

Setzt man $\mathcal{E}g := v_N$, so ergibt sich

$$\gamma_{0,x}^{\text{int}}\mathcal{E}g = \gamma_{0,x}^{\text{int}}v_N = g.$$

Somit gilt die Behauptung des Satzes mit $c_{IT} = 1 + \varepsilon$.

□

Als motivierendes Beispiel für die Definition einer schwachen Lösung von parabolischen Differentialgleichungen wird nun das folgende instationäre Randwertproblem mit homogenen Dirichlet–Daten betrachtet:

$$\begin{aligned} \frac{\partial}{\partial t}u(x,t) - \Delta u(x,t) &= f(x,t) \qquad \text{für} \quad (x,t) \in Q, \\ u(x,t) &= 0 \qquad\quad \text{für} \quad (x,t) \in \Sigma, \\ u(x,0) &= u_0(x) \qquad \text{für} \quad (x,0) \in \Sigma_0. \end{aligned} \tag{1.4}$$

Wird nun angenommen, dass u eine klassische Lösung der Aufgabenstellung (1.4) ist und multipliziert man die Differentialgleichung (1.4) mit einer Testfunktion $v \in L_2(0,T;H_0^1(\Omega))$, so erhält man durch Integration über den Raum–Zeit–Zylinder $Q = \Omega \times (0,T)$

$$\int_0^T \int_\Omega \frac{\partial}{\partial t}u(x,t)v(x,t)\mathrm{d}x\mathrm{d}t - \int_0^T \int_\Omega \Delta u(x,t)v(x,t)\mathrm{d}x\mathrm{d}t = \int_0^T \int_\Omega f(x,t)v(x,t)\mathrm{d}x\mathrm{d}t.$$

Die partielle Integration bezüglich der Ortsvariablen x ergibt

$$\int_0^T \int_\Omega \frac{\partial}{\partial t}u(x,t)v(x,t)\mathrm{d}x\mathrm{d}t + \int_0^T \int_\Omega \nabla_x u(x,t)\cdot\nabla_x v(x,t)\mathrm{d}x\mathrm{d}t = \int_0^T \int_\Omega f(x,t)v(x,t)\mathrm{d}x\mathrm{d}t.$$

In der letzten Gleichung muss die Ableitung von u nach der Zeit nur als Funktional bezüglich $v \in L_2(0,T;H_0^1(\Omega))$ existieren. Dies motiviert nun die Definition einer schwachen Lösung der Differentialgleichung (1.4):

Definition 1.15 (Schwache Lösung)**.** *Eine Lösung $u \in W_2^1(0,T;H_0^1(\Omega),L_2(\Omega))$ heißt* schwache Lösung *der Aufgabenstellung (1.4), falls die Gleichung*

$$\begin{aligned} \int_0^T \left\langle \frac{\partial}{\partial t}u(\cdot\,,t), v(\cdot\,,t)\right\rangle_{H^{-1}(\Omega)\times H_0^1(\Omega)} \mathrm{d}t + \int_0^T \int_\Omega \nabla_x u(x,t)\cdot\nabla_x v(x,t)\mathrm{d}x\mathrm{d}t \\ = \int_0^T \int_\Omega f(x,t)v(x,t)\mathrm{d}x\mathrm{d}t \end{aligned} \tag{1.5}$$

für alle $v \in L_2(0,T;H_0^1(\Omega))$ und die Bedingung $u(x,0) = u_0(x)$ für fast alle $x \in \Omega$ erfüllt ist.

Für $t \in (0,T)$ sei der Operator $A : H_0^1(\Omega) \to H^{-1}(\Omega)$ definiert als

$$\langle Au(\cdot\,,t), v(\cdot\,,t)\rangle_{H^{-1}(\Omega)\times H_0^1(\Omega)} := \int_\Omega \nabla_x u(x,t)\cdot\nabla_x v(x,t)\mathrm{d}x$$

für alle $u(\cdot,t), v(\cdot,t) \in H_0^1(\Omega)$. Fasst man weiters die rechte Seite der Gleichung (1.5) als Funktional $F(t) : H_0^1(\Omega) \to \mathbb{R}$ auf mit

$$\langle F(t), v(\cdot\,,t)\rangle_{H^{-1}(\Omega)\times H_0^1(\Omega)} := \int_\Omega f(x,t)v(x,t)\mathrm{d}x,$$

so lässt sich die Gleichung (1.5) schreiben als

$$\int_0^T \left\langle \frac{\partial}{\partial t}u(\cdot\,,t) + Au(\cdot\,,t) - F(t), v(\cdot\,,t)\right\rangle_{H^{-1}(\Omega)\times H_0^1(\Omega)} \mathrm{d}t = 0 \quad \text{für alle } v \in L_2(0,T;H_0^1(\Omega)).$$

Somit ist die Definition einer schwachen Lösung äquivalent zu der Operatorschreibweise

$$\begin{aligned} \frac{\partial}{\partial t}u + Au &= F \qquad \text{in } L_2(0,T;H^{-1}(\Omega)),\\ \gamma_{0,t}^{\mathrm{int}}u &= u_0 \qquad \text{in } H = L_2(\Omega). \end{aligned} \tag{1.6}$$

Für allgemeine Sobolev–Räume $W_2^1(0,T;V,H)$ gilt für Evolutionsgleichungen der Gestalt (1.6) folgender Existenz- und Eindeutigkeitssatz:

Satz 1.16. *Sei $V \subset H \subset V^*$ ein Gelfandscher Dreier, $F \in L_2(0,T;V^*)$ und $u_0 \in H$. Weiters sei $A : L_2(0,T;V) \to L_2(0,T;V^*)$ ein beschränkter und V–elliptischer Operator, das heißt es existieren positive Konstanten c_1^A, c_2^A, sodass*

$$\begin{aligned} \langle Au, v\rangle_{V^*\times V} &\le c_2^A \|u\|_V \|v\|_V \quad \textit{für alle } u, v \in V,\\ \langle Au, u\rangle_{V^*\times V} &\ge c_1^A \|u\|_V^2 \qquad\quad \textit{für alle } u \in V \end{aligned}$$

gilt. Dann besitzt die Problemstellung

$$\begin{aligned} \frac{\partial}{\partial t}u + Au &= F \quad \textit{in } L_2(0,T;V^*),\\ \gamma_{0,t}^{\mathrm{int}}u &= u_0 \quad \textit{in } H, \end{aligned}$$

eine eindeutig bestimmte Lösung $u \in W_2^1(0,T;V,H)$ und es gilt die Stabilitätsabschätzung

$$\|u\|_{W_2^1(0,T;V,H)} \le c_s \left[\|F\|_{L_2(0,T;V^*)} + \|u_0\|_H\right] \quad \textit{mit } c_s > 0.$$

Beweis. Siehe zum Beispiel [11, 33, 35]. □

Mithilfe des allgemeinen Satzes 1.16 folgt die eindeutige Lösbarkeit eines inhomogenen Dirichlet–Randwertproblems durch Homogenisierung.

Satz 1.17. *Für $f \in L_2(0,T;H^{-1}(\Omega))$, $u_0 \in L_2(\Omega)$ und $g \in W_2^1(\Sigma)$ existiert genau eine schwache Lösung $u \in W_2^1(0,T;H^1(\Omega),L_2(\Omega))$ des instationären inhomogenen Dirichlet–Randwertproblems*

$$\begin{aligned} \frac{\partial}{\partial t}u(x,t) - \Delta u(x,t) &= f(x,t) \qquad \text{für} \quad (x,t) \in Q, \\ u(x,t) &= g(x,t) \qquad \text{für} \quad (x,t) \in \Sigma, \\ u(x,0) &= u_0(x) \qquad \text{für} \quad (x,0) \in \Sigma_0. \end{aligned} \tag{1.7}$$

Für diese Lösung gilt die Stabilitätsabschätzung

$$\|u\|_{W_2^1(0,T;H^1(\Omega),L_2(\Omega))} \le c_s \left[\|f\|_{L_2(0,T;H^{-1}(\Omega))} + \|g\|_{W_2^1(\Sigma)} + \|u_0\|_{L_2(\Omega)} \right] \quad \text{mit } c_s > 0. \tag{1.8}$$

Beweis. Da $g \in W_2^1(\Sigma)$ ist, gibt es wegen Satz 1.14 ein $u_g := \mathcal{E}g \in W_2^1(0,T;H^1(\Omega),L_2(\Omega))$ mit

$$g(x,t) = \gamma_{0,x}^{\text{int}} u_g(x,t) \quad \text{für alle } (x,t) \in \Sigma$$

und

$$\|\mathcal{E}g\|_{W_2^1(0,T;H^1(\Omega),L_2(\Omega))} \le c_{IT} \|g\|_{W_2^1(\Sigma)}.$$

Einsetzen von $u = u_H + u_g$ in die Variationsgleichung (1.5) liefert

$$\begin{aligned} &\int_0^T \left\langle \frac{\partial}{\partial t} u_H(\cdot,t), v(\cdot,t) \right\rangle_{H^{-1}(\Omega)\times H_0^1(\Omega)} \mathrm{d}t + \int_0^T \langle A u_H(\cdot,t), v(\cdot,t) \rangle_{H^{-1}(\Omega)\times H_0^1(\Omega)} \mathrm{d}t = \\ &\qquad = \int_0^T \left\langle f(\cdot,t) - \frac{\partial}{\partial t} u_g(\cdot,t) - A u_g(\cdot,t), v(\cdot,t) \right\rangle_{H^{-1}(\Omega)\times H_0^1(\Omega)} \mathrm{d}t =: F(v). \end{aligned} \tag{1.9}$$

Nach [28, Lemma 4.1 und Theorem 4.3] ist der Operator $A : H_0^1(\Omega) \to H^{-1}(\Omega)$ beschränkt und $H_0^1(\Omega)$–elliptisch, das heißt es gilt

$$\begin{aligned} \langle Au, v \rangle_{H^{-1}(\Omega)\times H_0^1(\Omega)} &\le c_2^A \|u\|_{H^1(\Omega)} \|v\|_{H^1(\Omega)} \quad \text{für alle } u, v \in H^1(\Omega), \\ \langle Au, u \rangle_{H^{-1}(\Omega)\times H_0^1(\Omega)} &\ge c_1^A \|u\|_{H^1(\Omega)}^2 \qquad\qquad \text{für alle } u \in H_0^1(\Omega) \end{aligned}$$

mit $c_1^A, c_2^A > 0$. Die Beschränktheit der linearen Abbildung $F : L_2(0,T;H_0^1(\Omega)) \to \mathbb{R}$ ergibt sich aus der Beschränktheit von A und durch die Anwendung der Cauchy–Schwarzschen Ungleichung:

$$\begin{aligned} |F(v)| &\le \left[\|f\|_{L_2(0,T;H^{-1}(\Omega))} + \left\| \frac{\partial}{\partial t} u_g \right\|_{L_2(0,T;H^{-1}(\Omega))} + c_2^A \|u_g\|_{L_2(0,T;H^1(\Omega))} \right] \|v\|_{L_2(0,T;H^1(\Omega))} \\ &\le \left[\|f\|_{L_2(0,T;H^{-1}(\Omega))} + \sqrt{2} \max\{1, c_2^A\} \|u_g\|_{W_2^1(0,T;H^1(\Omega),L_2(\Omega))} \right] \|v\|_{L_2(0,T;H^1(\Omega))} \\ &\le \left[\|f\|_{L_2(0,T;H^{-1}(\Omega))} + \sqrt{2} \max\{1, c_2^A\} c_{IT} \|g\|_{W_2^1(\Sigma)} \right] \|v\|_{L_2(0,T;H^1(\Omega))} \\ &\le c_2^F \left[\|f\|_{L_2(0,T;H^{-1}(\Omega))} + \|g\|_{W_2^1(\Sigma)} \right] \|v\|_{L_2(0,T;H^1(\Omega))}. \end{aligned}$$

Wegen Satz 1.9 ist weiters

$$\gamma_{0,t}^{\mathrm{int}} u_H = u_0 - \gamma_{0,t}^{\mathrm{int}} u_g \in L_2(\Omega).$$

Somit gibt es wegen Satz 1.16 eine eindeutig bestimmte schwache Lösung

$$u_H \in W_2^1(0,T; H_0^1(\Omega), L_2(\Omega))$$

mit $u_H(x,0) = u_0(x) - \gamma_{0,t}^{\mathrm{int}} u_g(x,0)$ für fast alle $x \in \Omega$. Diese erfüllt die Stabilitätsabschätzung

$$\begin{aligned}
\|u_H\|_{W_2^1(0,T;H_0^1(\Omega),L_2(\Omega))} &\le \tilde{c}_s \Big[\|F\|_{L_2(0,T;H^{-1}(\Omega))} + \big\|u_0 - \gamma_{0,t}^{\mathrm{int}} u_g\big\|_{L_2(\Omega)} \Big] \\
&\le \tilde{c}_s \Big[c_2^F \left(\|f\|_{L_2(0,T;H^{-1}(\Omega))} + \|g\|_{W_2^1(\Sigma)} \right) + \|u_0\|_{L_2(\Omega)} \\
&\qquad + c_T \|u_g\|_{W_2^1(0,T;H^1(\Omega),L_2(\Omega))} \Big] \\
&\le \tilde{c}_s \Big[c_2^F \left(\|f\|_{L_2(0,T;H^{-1}(\Omega))} + \|g\|_{W_2^1(\Sigma)} \right) + \|u_0\|_{L_2(\Omega)} \\
&\qquad + c_T c_{IT} \|g\|_{W_2^1(\Sigma)} \Big] \\
&\le \hat{c}_s \Big[\|f\|_{L_2(0,T;H^{-1}(\Omega))} + \|g\|_{W_2^1(\Sigma)} + \|u_0\|_{L_2(\Omega)} \Big].
\end{aligned}$$

Somit folgt mithilfe der Dreiecksungleichung die Stabilitätsabschätzung

$$\begin{aligned}
\|u\|_{W_2^1(0,T;H^1(\Omega),L_2(\Omega))} &= \|u_H + u_g\|_{W_2^1(0,T;H^1(\Omega),L_2(\Omega))} \\
&\le \|u_H\|_{W_2^1(0,T;H^1(\Omega),L_2(\Omega))} + \|u_g\|_{W_2^1(0,T;H^1(\Omega),L_2(\Omega))} \\
&\le \|u_H\|_{W_2^1(0,T;H^1(\Omega),L_2(\Omega))} + c_{IT} \|g\|_{W_2^1(\Sigma)} \\
&\le c_s \left[\|f\|_{L_2(0,T;H^{-1}(\Omega))} + \|g\|_{W_2^1(\Sigma)} + \|u_0\|_{L_2(\Omega)} \right].
\end{aligned}$$

□

1.3 Eindeutige Lösbarkeit des optimalen Kontrollproblems

Wegen der eindeutigen Lösbarkeit des instationären Dirichlet–Randwertproblems (1.7) (siehe Satz 1.17) kann zu jeder Steuerung $z \in L_2(0,T; L_2(\Omega)) = L_2(Q)$ der Zustand

$$u \in W_2^1(0,T; H^1(\Omega), L_2(\Omega))$$

bestimmt werden. Somit ist der Lösungsoperator

$$G : L_2(0,T; L_2(\Omega)) \times W_2^1(\Sigma) \times L_2(\Omega) \to W_2^1(0,T; H^1(\Omega), L_2(\Omega))$$

wohldefiniert und kann aufgrund der Linearität der Differentialgleichung (1.2) geschrieben werden als

$$G(z, g, u_0) = N(z) + W(g) + M(u_0).$$

Wegen der Stabilitätsabschätzung (1.8) sind die linearen Operatoren

$$\begin{aligned} N &: L_2(0,T;L_2(\Omega)) \to W_2^1(0,T;H^1(\Omega),L_2(\Omega)),\\ W &: W_2^1(\Sigma) \to W_2^1(0,T;H^1(\Omega),L_2(\Omega)),\\ M &: L_2(\Omega) \to W_2^1(0,T;H^1(\Omega),L_2(\Omega)) \end{aligned}$$

beschränkt. Das heißt, es existieren positive Konstanten c_2^N, c_2^W und c_2^M, sodass

$$\begin{aligned} \|N(f)\|_{W_2^1(0,T;H^1(\Omega),L_2(\Omega))} &\le c_2^N \|f\|_{L_2(0,T;L_2(\Omega))} && \text{für alle } f \in L_2(0,T;L_2(\Omega)),\\ \|W(g)\|_{W_2^1(0,T;H^1(\Omega),L_2(\Omega))} &\le c_2^W \|g\|_{W_2^1(\Sigma)} && \text{für alle } g \in W_2^1(\Sigma),\\ \|M(u_0)\|_{W_2^1(0,T;H^1(\Omega),L_2(\Omega))} &\le c_2^M \|u_0\|_{L_2(\Omega)} && \text{für alle } u_0 \in L_2(\Omega) \end{aligned}$$

gilt. Um nun die eindeutige Lösbarkeit des optimalen Kontrollproblems (1.1)–(1.3) zeigen zu können, wird mithilfe des Lösungsoperators G ein reduziertes Minimierungsproblem eingeführt. Definiert man den linearen und beschränkten Operator

$$S : L_2(0,T;L_2(\Omega)) \to L_2(\Sigma_T) = L_2(\Omega)$$

als $S := \gamma_{T,t}^{\text{int}} N(z)$, so kann das Kostenfunktional (1.1) geschrieben werden als

$$\begin{aligned} \mathcal{J}(u,z) &= \frac{1}{2}\left\|\gamma_{T,t}^{\text{int}} u - \bar{u}\right\|_{L_2(\Sigma_T)}^2 + \frac{1}{2}\varrho\|z\|_{L_2(Q)}^2\\ &= \frac{1}{2}\left\|\gamma_{T,t}^{\text{int}} G(z,g,u_0) - \bar{u}\right\|_{L_2(\Sigma_T)}^2 + \frac{1}{2}\varrho\|z\|_{L_2(Q)}^2\\ &= \frac{1}{2}\left\|\gamma_{T,t}^{\text{int}} N(z) + \gamma_{T,t}^{\text{int}} W(g) + \gamma_{T,t}^{\text{int}} M(u_0) - \bar{u}\right\|_{L_2(\Sigma_T)}^2 + \frac{1}{2}\varrho\|z\|_{L_2(Q)}^2\\ &= \frac{1}{2}\left\|S(z) - \left(\bar{u} - \gamma_{T,t}^{\text{int}} W(g) - \gamma_{T,t}^{\text{int}} M(u_0)\right)\right\|_{L_2(\Sigma_T)}^2 + \frac{1}{2}\varrho\|z\|_{L_2(Q)}^2\\ &=: \widetilde{\mathcal{J}}(z). \end{aligned}$$

Definiert man $\bar{w} := \bar{u} - \gamma_{T,t}^{\text{int}}\left(W(g) + M(u_0)\right) \in L_2(\Sigma_T)$ und die Menge der zulässigen Steuerungen als

$$Z_{ad} := \left\{z \in L_2(0,T;L_2(\Omega)) : z_a(x,t) \le z(x,t) \le z_b(x,t) \text{ für fast alle } (x,t) \in Q\right\},$$

so lässt sich das optimale Kontrollproblem (1.1)–(1.3) als folgendes reduziertes Minimierungsproblem schreiben: Gesucht ist die optimale Steuerung $\hat{z} \in Z_{ad}$, sodass

$$\widetilde{\mathcal{J}}(\hat{z}) = \min_{z \in Z_{ad}} \widetilde{\mathcal{J}}(z) = \min_{z \in Z_{ad}} \left\{\frac{1}{2}\|S(z) - \bar{w}\|_{L_2(\Sigma_T)}^2 + \frac{1}{2}\varrho\|z\|_{L_2(Q)}^2\right\} \tag{1.10}$$

gilt. In der Bakkalaureatsseminararbeit [19] wurde für Minimierungsprobleme der Gestalt (1.10) ein allgemeiner Existenz- und Eindeutigkeitssatz für Lösungen dieser Minimierungsprobleme bewiesen.

Satz 1.18. *Es seien $(Z, \langle\cdot,\cdot\rangle_Z)$ und $(U, \langle\cdot,\cdot\rangle_U)$ zwei reelle Hilberträume. Weiters sei $Z_{ad} \subseteq Z$ eine nichtleere, abgeschlossene und konvexe Menge. Ferner sei $S : Z \to U$ ein linearer und beschränkter Operator. Dann besitzt die Minimierung des Funktionals*

$$\mathcal{J}(z) := \frac{1}{2}\|S(z) - \bar{u}\|_U^2 + \frac{1}{2}\varrho\|z\|_Z^2$$

für jedes $\bar{u} \in U$ und $\varrho \geq 0$ eine optimale Lösung $\hat{z} \in Z_{ad}$. Das heißt, es gilt

$$\mathcal{J}(\hat{z}) = \min_{z \in Z_{ad}} \mathcal{J}(z).$$

Für $\varrho > 0$ ist diese eindeutig bestimmt.

Beweis. Siehe zum Beispiel [19, Satz 2.13, Korollar 2.14] oder [31, Satz 2.14, Satz 2.16]. □

Um die eindeutige Lösbarkeit des reduzierten Minimierungsproblems und somit die eindeutige Lösbarkeit der Aufgabenstellung (1.1)–(1.3) zeigen zu können, müssen nun die Voraussetzungen des Satzes 1.18 überprüft werden. Es bleiben somit nur mehr die notwendigen Eigenschaften der Menge Z_{ad} nachzuprüfen:

Lemma 1.19. *Die Menge der zulässigen Steuerungen Z_{ad} ist eine nichtleere, abgeschlossene und konvexe Teilmenge von $L_2(Q)$.*

Beweis. Analog wie der Beweis von [19, Lemma 2.18]. □

Da $S : L_2(0,T;L_2(\Omega)) \to L_2(\Sigma_T)$ ein linearer und beschränkter Operator ist, folgt mit Hilfe von Lemma 1.19 und Satz 1.18 für $\varrho \geq 0$ die Lösbarkeit sowie für $\varrho > 0$ die eindeutige Lösbarkeit der Aufgabenstellung (1.1)–(1.3).

1.4 Optimalitätssystem

In diesem Abschnitt wird für die Aufgabenstellung (1.1)–(1.3) eine äquivalente Formulierung hergeleitet. Dazu wird eine allgemeine Aussage über quadratische Optimierungsaufgaben in Hilberträumen verwendet, die in der Bakkalaureatsseminararbeit [19] bewiesen wurde.

Satz 1.20. *Es seien $(Z, \langle\cdot,\cdot\rangle_Z)$ und $(U, \langle\cdot,\cdot\rangle_U)$ zwei reelle Hilberträume. Weiters sei $Z_{ad} \subseteq Z$ eine nichtleere, abgeschlossene und konvexe Menge, $\bar{u} \in U$, $\varrho \geq 0$ und $S : Z \to U$ ein linearer und beschränkter Operator. Dann sind die folgenden zwei Aussagen äquivalent:*

$$i) \qquad f(\hat{z}) = \min_{z \in Z_{ad}} f(z) := \min_{z \in Z_{ad}} \left\{ \frac{1}{2}\|S(z) - \bar{u}\|_U^2 + \frac{1}{2}\varrho\|z\|_Z^2 \right\},$$

$$ii) \qquad \langle S^*\left(S(\hat{z}) - \bar{u}\right) + \varrho\hat{z}, z - \hat{z}\rangle_Z \geq 0 \quad \textit{für alle } z \in Z_{ad}.$$

Beweis. Siehe zum Beispiel [19, Satz 3.8] oder [31, Satz 2.22]. □

Damit Satz 1.20 auf das reduzierte Minimierungsproblem (1.10) angewendet werden kann, muss der adjungierte Operator $S^* : L_2(\Sigma_T) \to L_2(0,T;L_2(\Omega))$ von $S : L_2(0,T;L_2(\Omega)) \to L_2(\Sigma_T)$ bestimmt werden. Dazu wird das instationäre Randwertproblem

$$\begin{aligned} -\frac{\partial}{\partial t}p(x,t) - \Delta p(x,t) &= 0 && \text{für} \quad (x,t) \in Q, \\ p(x,t) &= 0 && \text{für} \quad (x,t) \in \Sigma, \\ p(x,T) &= w(x) && \text{für} \quad (x,T) \in \Sigma_T \end{aligned} \tag{1.11}$$

betrachtet. Eine schwache Lösung $p \in W_2^1(0,T;H_0^1(\Omega),L_2(\Omega))$ der Differentialgleichung (1.11) muss somit die Gleichung

$$-\int_0^T \left\langle \frac{\partial}{\partial t}p(\cdot\,,t), q(\cdot\,,t) \right\rangle_{H^{-1}(\Omega)\times H_0^1(\Omega)} \mathrm{d}t + \int_0^T \int_\Omega \nabla_x p(x,t) \cdot \nabla_x q(x,t) \mathrm{d}x\mathrm{d}t = 0 \tag{1.12}$$

für alle $q \in L_2(0,T;H_0^1(\Omega))$ und die Bedingung $p(x,T) = w(x)$ für fast alle $x \in \Omega$ erfüllen.

Satz 1.21. *Die Differentialgleichung (1.11) besitzt für $w \in L_2(\Sigma_T)$ genau eine schwache Lösung $p \in W_2^1(0,T;H_0^1(\Omega),L_2(\Omega))$ und es gilt die Stabilitätsabschätzung*

$$\|p\|_{W_2^1(0,T;H_0^1(\Omega),L_2(\Omega))} \le c\|w\|_{L_2(\Sigma_T)}.$$

Beweis. Betrachtet man die Transformation $t = T - \tau$ und definiert

$$\tilde{p}(x,\tau) := p(x,T-\tau) = p(x,t),$$

so ergibt sich für fast alle $(x,T) \in \Sigma_T$ die Bedingung

$$\tilde{p}(x,0) = p(x,T) = w(x).$$

Weiters gilt für das erste Integral der Gleichung (1.12) mit der Transformation $t = T - \tau$ die Beziehung

$$\begin{aligned} -\int_0^T \left\langle \frac{\partial}{\partial t}p(\cdot\,,t), q(\cdot\,,t) \right\rangle_{H^{-1}(\Omega)\times H_0^1(\Omega)} \mathrm{d}t &= \int_T^0 \left\langle -\frac{\partial}{\partial \tau}\tilde{p}(\cdot\,,\tau), q(\cdot\,,T-\tau) \right\rangle_{H^{-1}(\Omega)\times H_0^1(\Omega)} \mathrm{d}\tau \\ &= \int_0^T \left\langle \frac{\partial}{\partial \tau}\tilde{p}(\cdot\,,\tau), \tilde{q}(\cdot\,,\tau) \right\rangle_{H^{-1}(\Omega)\times H_0^1(\Omega)} \mathrm{d}\tau, \end{aligned}$$

wobei $\tilde{q}(x,\tau) := q(x, T-\tau)$. Analog lässt sich das zweite Integral umschreiben zu

$$\begin{aligned}\int_0^T \int_\Omega \nabla_x p(x,t) \cdot \nabla_x q(x,t) \mathrm{d}x\mathrm{d}t &= \int_T^0 \int_\Omega \nabla_x \tilde{p}(x,\tau) \cdot \nabla_x \tilde{q}(x,\tau) \mathrm{d}x(-\mathrm{d}\tau) \\ &= \int_0^T \int_\Omega \nabla_x \tilde{p}(x,\tau) \cdot \nabla_x \tilde{q}(x,\tau) \mathrm{d}x\mathrm{d}\tau.\end{aligned}$$

Somit lässt sich die Gleichung (1.12) schreiben als

$$\int_0^T \left\langle \frac{\partial}{\partial\tau}\tilde{p}(\cdot\,,\tau), \tilde{q}(\cdot\,,\tau) \right\rangle_{H^{-1}(\Omega)\times H_0^1(\Omega)} \mathrm{d}\tau + \int_0^T \int_\Omega \nabla_x \tilde{p}(x,\tau) \cdot \nabla_x \tilde{q}(x,\tau) \mathrm{d}x\mathrm{d}\tau = 0 \tag{1.13}$$

für alle $\tilde{q} \in L_2(0,T; H_0^1(\Omega))$. Weiters gilt die Bedingung

$$\tilde{p}(x,0) = w(x) \quad \text{für fast alle } x \in \Omega. \tag{1.14}$$

Nach Satz 1.16 existiert genau ein $\tilde{p} \in W_2^1(0,T; H_0^1(\Omega), L_2(\Omega))$ mit der Stabilitätsbedingung

$$\|\tilde{p}\|_{W_2^1(0,T;H^1(\Omega),L_2(\Omega))} \le c_s \|w\|_{L_2(\Omega)},$$

welches die Gleichung (1.13) und die Bedingung (1.14) erfüllt. Nach Rücktransformation von $\tilde{p} = \tilde{p}(x,\tau)$ mit $\tau = T - t$ ergibt sich die Aussage des Satzes für

$$p \in W_2^1(0,T; H_0^1(\Omega), L_2(\Omega)).$$

□

Der nächste Satz liefert eine Aussage über den adjungierten Operator S^* von S.

Satz 1.22. *Sei $z \in L_2(0,T; L_2(\Omega))$ und $u = N(z) \in W_2^1(0,T; H_0^1(\Omega), L_2(\Omega))$ die schwache Lösung der Differentialgleichung*

$$\begin{aligned}\frac{\partial}{\partial t}u(x,t) - \Delta u(x,t) &= z(x,t) \qquad \textit{für} \quad (x,t) \in Q, \\ u(x,t) &= 0 \qquad \textit{für} \quad (x,t) \in \Sigma, \\ u(x,0) &= 0 \qquad \textit{für} \quad (x,0) \in \Sigma_0.\end{aligned} \tag{1.15}$$

Somit ist $S(z) = \gamma_{T,t}^{\mathrm{int}} N(z) = \gamma_{T,t}^{\mathrm{int}} u$. Sei nun weiters $S^*(w) := p \in W_2^1(0,T; H_0^1(\Omega), L_2(\Omega))$ die schwache Lösung der Differentialgleichung

$$\begin{aligned}-\frac{\partial}{\partial t}p(x,t) - \Delta p(x,t) &= 0 \qquad \textit{für} \quad (x,t) \in Q, \\ p(x,t) &= 0 \qquad \textit{für} \quad (x,t) \in \Sigma, \\ p(x,T) &= w(x) \qquad \textit{für} \quad (x,T) \in \Sigma_T.\end{aligned} \tag{1.16}$$

Dann gilt

$$\langle S(z), w \rangle_{L_2(\Sigma_T)} = \langle z, S^*(w) \rangle_{L_2(Q)}.$$

Beweis. Da $u \in W_2^1(0,T;H_0^1(\Omega),L_2(\Omega))$ die schwache Lösung der Differentialgleichung (1.15) ist, erhält man durch Einsetzen von $v = p$ in die schwache Formulierung den Zusammenhang

$$\begin{aligned}\int_0^T \left\langle \frac{\partial}{\partial t}u(\cdot\,,t), p(\cdot\,,t)\right\rangle_{H^{-1}(\Omega)\times H_0^1(\Omega)} \mathrm{d}t + \int_0^T \int_\Omega \nabla_x u(x,t)\cdot\nabla_x p(x,t)\mathrm{d}x\mathrm{d}t \\ = \int_0^T \int_\Omega z(x,t)p(x,t)\mathrm{d}x\mathrm{d}t.\end{aligned} \tag{1.17}$$

Setzt man für $q = u$ in die schwache Formulierung (1.12) der Differentialgleichung (1.16) ein, so ergibt sich

$$-\int_0^T \left\langle \frac{\partial}{\partial t}p(\cdot\,,t), u(\cdot\,,t)\right\rangle_{H^{-1}(\Omega)\times H_0^1(\Omega)} \mathrm{d}t + \int_0^T \int_\Omega \nabla_x p(x,t)\cdot\nabla_x u(x,t)\mathrm{d}x\mathrm{d}t = 0.$$

Wendet man die in Satz 1.10 erklärte partielle Integration bezüglich der Zeit an, so erhält man weiters

$$\begin{aligned}\int_0^T \left\langle \frac{\partial}{\partial t}u(\cdot\,,t), p(\cdot\,,t)\right\rangle_{H^{-1}(\Omega)\times H_0^1(\Omega)} \mathrm{d}t - \left\langle \gamma_{T,t}^{\mathrm{int}}u, \gamma_{T,t}^{\mathrm{int}}p\right\rangle_{L_2(\Sigma_T)} \\ + \int_0^T \int_\Omega \nabla_x p(x,t)\cdot\nabla_x u(x,t)\mathrm{d}x\mathrm{d}t = 0.\end{aligned} \tag{1.18}$$

Durch die Subtraktion der Gleichung (1.18) von der Gleichung (1.17) und durch das Einsetzen der Beziehungen $S(z) = \gamma_{T,t}^{\mathrm{int}}u$ und $\gamma_{T,t}^{\mathrm{int}}p(x,t) = w(x)$ für fast alle $(x,t) \in \Sigma_T$, ergibt sich mit

$$\langle S(z), w\rangle_{L_2(\Sigma_T)} = \langle z, S^*(w)\rangle_{L_2(Q)}$$

die Behauptung des Satzes. □

Mithilfe von $\bar{w} = \bar{u} - \gamma_{T,t}^{\mathrm{int}}\left(W(g) + M(u_0)\right)$ wurde das optimale Kontrollproblem (1.1)–(1.3) geschrieben als reduziertes Minimierungsproblem: Gesucht ist die optimale Steuerung $\hat{z} \in Z_{ad}$, sodass $\hat{z}$ das quadratische Funktional

$$\widetilde{\mathcal{J}}(z) = \frac{1}{2}\|S(z) - \bar{w}\|_{L_2(\Sigma_T)}^2 + \frac{1}{2}\varrho\|z\|_{L_2(Q)}^2 \tag{1.19}$$

minimiert. Nach Satz 1.20 ist das reduzierte Minimierungsproblem (1.19) äquivalent zur Variationsungleichung

$$\left\langle S^*\left(S(\hat{z}) - \bar{w}\right) + \varrho\hat{z}, z - \hat{z}\right\rangle_{L_2(Q)} \geq 0 \quad \text{für alle } z \in Z_{ad}. \tag{1.20}$$

Weiters gilt die Beziehung

$$\begin{aligned} S(\hat{z}) - \bar{w} &= \gamma_{T,t}^{\text{int}} N(\hat{z}) - \bar{w} \\ &= \gamma_{T,t}^{\text{int}} N(\hat{z}) - \bar{u} + \gamma_{T,t}^{\text{int}} \left(W(g) + M(u_0)\right) \\ &= \gamma_{T,t}^{\text{int}} \left(N(\hat{z}) + W(g) + M(u_0)\right) - \bar{u} \\ &= \gamma_{T,t}^{\text{int}} G(\hat{z}, g, u_0) - \bar{u}. \end{aligned}$$

Definiert man nun den Zustand

$$\hat{u} := G(\hat{z}, g, u_0) \tag{1.21}$$

und den adjungierten Zustand

$$\begin{aligned} \hat{p} &:= S^*(S(\hat{z}) - \bar{u}) \\ &= S^*(\gamma_{T,t}^{\text{int}} G(\hat{z}, g, u_0) - \bar{u}) \\ &= S^*(\gamma_{T,t}^{\text{int}} \hat{u} - \bar{u}), \end{aligned} \tag{1.22}$$

so wird die Variationsungleichung (1.20) zu

$$\langle \hat{p} + \varrho\hat{z}, z - \hat{z} \rangle_{L_2(Q)} \geq 0 \quad \text{für alle } z \in Z_{ad}. \tag{1.23}$$

Wendet man die Definitionen des Operators G und die in Satz 1.22 bewiesene Darstellung des adjungierten Operators S^* für die Gleichungen (1.21)–(1.22) an, so ergibt sich mit der Variationsungleichung (1.23) ein zum optimalen Kontrollproblem (1.1)–(1.3) äquivalentes Optimalitätssystem:
Gesucht sind der optimale Zustand $\hat{u} \in W_2^1(0,T; H^1(\Omega), L_2(\Omega))$, der optimale adjungierte Zustand $\hat{p} \in W_2^1(0,T; H_0^1(\Omega), L_2(\Omega))$ und die optimale Steuerung $\hat{z} \in L_2(0,T; L_2(\Omega))$ als schwache Lösungen der Differentialgleichungen

$$\begin{aligned} \frac{\partial}{\partial t}\hat{u}(x,t) - \Delta\hat{u}(x,t) &= \hat{z}(x,t) && \text{für} \quad (x,t) \in Q, \\ \hat{u}(x,t) &= g(x,t) && \text{für} \quad (x,t) \in \Sigma, \\ \hat{u}(x,0) &= u_0(x) && \text{für} \quad (x,0) \in \Sigma_0, \end{aligned} \tag{1.24}$$

und

$$\begin{aligned} -\frac{\partial}{\partial t}\hat{p}(x,t) - \Delta\hat{p}(x,t) &= 0 && \text{für} \quad (x,t) \in Q, \\ \hat{p}(x,t) &= 0 && \text{für} \quad (x,t) \in \Sigma, \\ \hat{p}(x,T) &= \gamma_{T,t}^{\text{int}}\hat{u}(x,T) - \bar{u}(x) && \text{für} \quad (x,T) \in \Sigma_T, \end{aligned} \tag{1.25}$$

und der Variationsungleichung

$$\langle \hat{p} + \varrho\hat{z}, z - \hat{z} \rangle_{L_2(Q)} \geq 0 \qquad \text{für alle } z \in Z_{ad}. \tag{1.26}$$

Sind keine Restriktionen an die Steuerung vorgegeben, also $Z_{ad} = L_2(Q)$, so wird die Variationsungleichung (1.26) zur Variationsgleichung

$$\langle \hat{p} + \varrho\hat{z}, z\rangle_{L_2(Q)} = 0 \quad \text{für alle } z \in L_2(Q).$$

Die Schwierigkeit im Lösen des Optimalitätssystems liegt darin, dass der Zustand $\hat{u}$ zum Zeitpunkt $t = 0$ und der adjungierte Zustand $\hat{p}$ zum Zeitpunkt $t = T$ gegeben sind. Im nächsten Kapitel wird für das Optimalitätssystem eine mögliche Diskretisierung mittels der Discontinuous Galerkin Methode hergeleitet. Dabei wird der gesamte Raum–Zeit–Zylinder diskretisiert.

2 Diskretisierung

In diesem Kapitel wird das Optimalitätssystem (1.24)–(1.26), welches in Kapitel 1 hergeleitet wurde, für den Fall $Z_{ad} = L_2(Q)$ diskretisiert. Dazu wird eine diskrete Variationsformulierung im ganzen Raum–Zeit–Zylinder $Q = \Omega \times (0,T)$ hergeleitet. In [7] wird der Raum–Zeit–Zylinder mithilfe der Finiten Differenzen Methode diskretisiert. Hier wird für den elliptischen Anteil ein Ansatz mittels der Discontinuous Galerkin Methode wie in [23] vorgenommen. Für den zeitlichen Anteil wird ebenfalls die Discontinuous Galerkin Methode wie in [30] verwendet. Für ein weiteres Studium der Discontinuous Galerkin Methode sei hier auf [2, 3, 17, 23] verwiesen.

2.1 Grundlagen

In diesem Abschnitt werden die zur Diskretisierung benötigten Begriffe eingeführt. Dabei wurden die Bezeichnungen wie in [28] verwendet.
Betrachtet wird der beschränkte Raum–Zeit–Zylinder $Q = \Omega \times (0,T) \subset \mathbb{R}^{d+1}$, $d = 1,2,3$. Dabei wird angenommen, dass der Raum–Zeit–Zylinder Q entweder polygonal ($d = 1$), polyhedral ($d = 2$) oder polychoral ($d = 3$) berandet sei. Sei nun eine Folge $\{\mathcal{T}_N\}_{N\in\mathbb{N}}$ von Unterteilungen

$$\overline{Q} = \overline{\mathcal{T}}_N = \bigcup_{k=1}^{N} \overline{\tau}_k$$

in Finite Elemente τ_k gegeben. Dabei werden die einfachsten Finiten Elemente betrachtet. Für $d = 1$ sind diese durch Dreiecke, für $d = 2$ durch Tetraeder und für $d = 3$ durch Pentatope gegeben.

Definition 2.1 (Zulässige Zerlegung)**.** *Eine Zerlegung $\mathcal{T}_N$ heißt* zulässig*, falls zwei benachbarte Elemente entweder eine Kante ($d = 1,2,3$), ein Dreieck ($d = 2,3$) oder einen Tetraeder ($d = 3$) gemeinsam haben. Zwei Elemente heißen* benachbart*, wenn der Durchschnitt der Abschlüsse dieser Elemente eine k–dimensionale Mannigfaltigkeit mit $k \leq d$ ergibt.*

Hier muss eine Zerlegung $\mathcal{T}_N$ im Allgemeinen nicht zulässig sein, das heißt, es sind hängende Knoten erlaubt, siehe Abbildung 2.1. Die Überlappung von je zwei Elementen ist jedoch nicht erlaubt.

Definition 2.2 (Inneres Element, Randelement)**.** *Sei $\mathcal{T}_N$ eine Zerlegung. Weiters seien $\tau_k, \tau_\ell \in \mathcal{T}_N$ zwei benachbarte Elemente. Dann wird mit*

$$\Gamma_{k\ell} := \overline{\tau}_k \cap \overline{\tau}_\ell \quad \textit{für } k < \ell$$

ein inneres Element *bezeichnet, falls* $\Gamma_{k\ell}$ *eine* d*–dimensionale Mannigfaltigkeit bezüglich* $\mathbb{R}^{d+1}$ *ist. Die Menge aller inneren Elemente der Triangulierung* $\mathcal{T}_N$ *wird mit* $\mathcal{I}_N$ *bezeichnet. Falls der Durchschnitt*

$$\Gamma_k := \overline{\tau}_k \cap \partial Q \neq \emptyset$$

nichtleer ist, so wird mit Γ_k *ein* Randelement *bezeichnet. Weiters beschreibt* $\mathcal{R}_N$ *die Menge aller Randelemente der Triangulierung* $\mathcal{T}_N$.

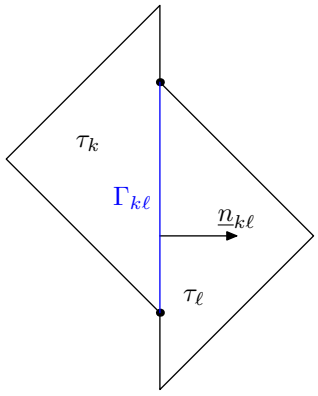


Abbildung 2.1: Ein inneres Element $\Gamma_{k\ell}$ mit Normalenvektor $\underline{n}_{k\ell}$ für $d = 1$.

Jedes innere Element $\Gamma_{k\ell} \in \mathcal{I}_N$ besitzt einen Normalenvektor $\underline{n}_{k\ell}$, dessen Richtung nicht eindeutig bestimmt ist. Hier wird die Richtung eindeutig festgelegt, indem der Normalenvektor $\underline{n}_{k\ell}$ für $k < \ell$ nach außen bezüglich des Elements τ_k zeigt, siehe Abbildung 2.1. Deshalb wird in den folgenden Abschnitten für ein inneres Element $\Gamma_{k\ell} \in \mathcal{I}_N$ die Bedingung $k < \ell$ gefordert. Für Randelemente $\Gamma_k \in \mathcal{R}_N$ wird für den Normalenvektor $\underline{n}_k$ der nach außen gerichtete Normalenvektor bezüglich des Randes ∂Q verwendet.

Definition 2.3 (Volumen, lokale Maschenweite, Durchmesser, Radius)**.** *Sei* $\tau_k \in \mathcal{T}_N$ *ein Finites Element. Dann wird mit*

$$\Delta_k := \int_{\tau_k} \mathrm{d}x\mathrm{d}t$$

das Volumen *und mit*

$$h_k := \Delta_k^{1/(d+1)}$$

die lokale Maschenweite *des Finiten Elements* τ_k *definiert. Weiters bezeichnet*

$$d_k := \sup_{(x_1,t_1),(x_2,t_2)\in\tau_k} |(x_1,t_1) - (x_2,t_2)|$$

den Durchmesser *des Finiten Elements* τ_k*. Der* Radius r_k *des Finiten Elements* τ_k *ist definiert durch den Radius der größten in* τ_k *enthaltenen Kugel*

$$B_{r_k}(x_k,t_k) := \left\{(x,t) \in \mathbb{R}^{d+1} : |(x,t) - (x_k,t_k)| < r_k\right\} \subset \tau_k.$$

Definition 2.4 (Formregularität)**.** *Die Finiten Elemente τ_k einer Unterteilung $\mathcal{T}_N$ heißen* formregulär*, falls eine Konstante $c_F > 0$ existiert, sodass*

$$d_k \leq c_F r_k \quad \textit{für alle } k = 1, \ldots, N$$

unabhängig von N gilt.

Definition 2.5 (Globale Maschenweite, global gleichmäßig)**.** *Für eine Zerlegung $\mathcal{T}_N$ ist die* globale Maschenweite *definiert durch*

$$h := h_{\max} := \max_{k=1,\ldots,N} h_k.$$

Sei weiterhin die minimale Maschenweite gegeben durch

$$h_{\min} := \min_{k=1,\ldots,N} h_k.$$

Dann heißt die Familie von Unterteilungen $\mathcal{T}_N$ global gleichmäßig*, falls eine von N unabhängige Konstante $c_G \geq 1$ existiert, sodass für alle Triangulierungen $\mathcal{T}_N$*

$$\frac{h_{\max}}{h_{\min}} \leq c_G$$

gilt.

Für innere Elemente lässt sich der Sprung und der Mittelwert definieren:

Definition 2.6 (Sprung, Mittelwert)**.** *Sei $e = \Gamma_{k\ell} \in \mathcal{I}_N$ ein inneres Element der Unterteilung $\mathcal{T}_N$. Dann wird mit*

$$[v]_e(x,t) := v_{|\tau_k}(x,t) - v_{|\tau_\ell}(x,t) \quad \textit{für alle } (x,t) \in e = \Gamma_{k\ell} \textit{ mit } k < \ell$$

der Sprung und mit

$$\langle v \rangle_e(x,t) := \frac{1}{2}\left[v_{|\tau_k}(x,t) + v_{|\tau_\ell}(x,t)\right] \quad \textit{für alle } (x,t) \in e = \Gamma_{k\ell}$$

der Mittelwert bezüglich des inneren Elements e bezeichnet.

Lemma 2.7. *Sei $e \in \mathcal{I}_N$ ein inneres Element der Unterteilung $\mathcal{T}_N$. Für zwei Funktionen u, v gilt dann die Beziehung*

$$[uv]_e(x,t) = [u]_e(x,t)\langle v \rangle_e(x,t) + \langle u \rangle_e(x,t)[v]_e(x,t) \quad \textit{für alle } (x,t) \in e. \tag{2.1}$$

Beweis. Nach Definition 2.6 folgt die Behauptung für $e = \Gamma_{k\ell}$ mit

$$\begin{aligned}
[u]_e\langle v \rangle_e + \langle u \rangle_e[v]_e &= \big(u_{|\tau_k} - u_{|\tau_\ell}\big)\frac{1}{2}\big(v_{|\tau_k} + v_{|\tau_\ell}\big) + \frac{1}{2}\big(u_{|\tau_k} + u_{|\tau_\ell}\big)\big(v_{|\tau_k} - v_{|\tau_\ell}\big) \\
&= \frac{1}{2}\big(uv_{|\tau_k} - u_{|\tau_\ell}v_{|\tau_k} + u_{|\tau_k}v_{|\tau_\ell} - uv_{|\tau_\ell} \\
&\qquad\qquad + uv_{|\tau_k} + u_{|\tau_\ell}v_{|\tau_k} - u_{|\tau_k}v_{|\tau_\ell} - uv_{|\tau_\ell}\big) \\
&= uv_{|\tau_k} - uv_{|\tau_\ell} = [uv]_e.
\end{aligned}$$

□

Weiters wird ein anisotroper Sobolev–Raum eingeführt. Dazu sei $\tau \subset \mathbb{R}^{d+1}$ für $d = 1, 2, 3$ ein beschränktes Gebiet. Dann wird für alle $s_x, s_t \in \mathbb{N}_0$ durch

$$H^{s_x,s_t}(\tau) := \{v \in L_2(\tau) : D^{\alpha_t} D^{\alpha_x} v \in L_2(\tau) \text{ für alle } |\alpha_x| \leq s_x, \alpha_t \leq s_t\}$$

ein anisotroper Sobolev–Raum erklärt. Dabei ist α_x ein Multiindex der Dimension d und $\alpha_t \in \mathbb{N}_0$. Der Raum $H^{s_x,s_t}(\tau)$ wird versehen mit der Norm

$$\|u\|_{H^{s_x,s_t}(\tau)} := \left[\sum_{\substack{\alpha_t \leq s_t \\ |\alpha_x| \leq s_x}} \|D^{\alpha_t} D^{\alpha_x} u\|^2_{L_2(\tau)} \right]^{\frac{1}{2}}.$$

Für jede Unterteilung $\mathcal{T}_N$ lässt sich nun ein stückweiser anisotroper Sobolev–Raum definieren:

Definition 2.8. *Für das Gebiet Q sei eine Unterteilung $\mathcal{T}_N$ in Finite Elemente $\tau_k \in \mathcal{T}_N$ gegeben. Dann wird für $s_x, s_t \in \mathbb{N}_0$ mit*

$$H^{s_x,s_t}(\mathcal{T}_N) := \left\{v \in L_2(Q) : v_{|\tau_k} \in H^{s_x,s_t}(\tau_k) \text{ für alle } \tau_k \in \mathcal{T}_N\right\}$$

ein stückweiser Sobolev–Raum *definiert. Weiters ist für alle $v \in H^{s_x,s_t}(\mathcal{T}_N)$ durch*

$$\|v\|_{H^{s_x,s_t}(\mathcal{T}_N)} := \left[\sum_{k=1}^{N} \|v\|^2_{H^{s_x,s_t}(\tau_k)} \right]^{\frac{1}{2}}$$

eine Norm in $H^{s_x,s_t}(\mathcal{T}_N)$ gegeben.

2.2 Variationsformulierungen

In diesem Abschnitt wird für das Optimalitätssystem (1.24)–(1.26) für den Fall $Z_{ad} = L_2(Q)$ eine diskrete Variationsformulierung mittels der Discontinuous Galerkin Methode hergeleitet. Dazu wird vorerst eine allgemeine Form der Zustandsgleichung diskretisiert. Anschließend wird in ähnlicher Weise eine diskrete Variationsformulierung für die adjungierte Zustandsgleichung hergeleitet.

2.2.1 Zustandsgleichung

Betrachtet wird das instationäre Dirichlet–Randwertproblem

$$\begin{aligned} \frac{\partial}{\partial t} u(x,t) - \Delta u(x,t) &= f(x,t) && \text{für} \quad (x,t) \in Q, \\ u(x,t) &= g(x,t) && \text{für} \quad (x,t) \in \Sigma, \\ u(x,0) &= u_0(x) && \text{für} \quad (x,0) \in \Sigma_0. \end{aligned} \tag{2.2}$$

Im Folgenden wird nun eine Variationsformulierung für die Aufgabenstellung (2.2) hergeleitet. Dazu wird zunächst angenommen, dass die Lösung u von (2.2) eine klassische Lösung ist. Das heißt, die Lösung u und deren Ableitungen sind stetig. Somit gelten für die Sprünge an einem inneren Elements $e \in \mathcal{I}_N$ die Beziehungen

$$[u]_e(x,t) = 0 \quad \text{und} \quad [\underline{n}_e \cdot \nabla_x u]_e(x,t) = 0 \quad \text{für alle } (x,t) \in e.$$

Für die Variationsformulierung wird die Differentialgleichung in (2.2) mit einer Testfunktion $v \in H^{s_x,s_t}(\mathcal{T}_N)$ multipliziert und über den Raum–Zeit–Zylinder Q integriert:

$$\int_Q \frac{\partial}{\partial t} u(x,t)v(x,t)\mathrm{d}q - \int_Q \Delta u(x,t)v(x,t)\mathrm{d}q = \int_Q f(x,t)v(x,t)\mathrm{d}q. \tag{2.3}$$

Dabei ist $\mathrm{d}q := \mathrm{d}x\mathrm{d}t$. Teilt man das erste Integral auf der linken Seite der Gleichung (2.3) auf die Unterteilung $\mathcal{T}_N$ auf, so erhält man

$$\int_Q \frac{\partial}{\partial t} u(x,t)v(x,t)\mathrm{d}q = \sum_{k=1}^{N} \int_{\tau_k} \frac{\partial}{\partial t} u(x,t)v(x,t)\mathrm{d}q.$$

Die partielle Integration bezüglich der Zeit t liefert

$$= -\sum_{k=1}^{N} \int_{\tau_k} u(x,t)\frac{\partial}{\partial t} v(x,t)\mathrm{d}q + \sum_{k=1}^{N} \int_{\partial\tau_k} n_t(x,t)u(x,t)v(x,t)\mathrm{d}s_q.$$

Das Aufteilen der Summation der Ränder $\partial\tau_k$ auf die inneren Elemente und auf die Randelemente ergibt

$$= -\sum_{k=1}^{N} \int_{\tau_k} u(x,t)\frac{\partial}{\partial t} v(x,t)\mathrm{d}q + \sum_{e\in\mathcal{I}_N} \int_e n_t(x,t)[uv]_e(x,t)\mathrm{d}s_q + \int_{\Sigma_0\cup\Sigma_T} n_t(x,t)u(x,t)v(x,t)\mathrm{d}s_q.$$

Dabei ist das Randintegral bezüglich dem Rand Σ gleich Null, da hier das Gebiet Ω als konstant vorausgesetzt wird. Für den Ausdruck bezüglich der inneren Elemente $e = \Gamma_{k\ell} \in \mathcal{I}_N$ gilt wegen Lemma 2.7 die Beziehung

$$\int_e n_t(x,t)[uv]_e(x,t)\mathrm{d}s_q = \int_e n_t(x,t)[u]_e(x,t)\langle v\rangle_e(x,t)\mathrm{d}s_q + \int_e n_t(x,t)\langle u\rangle_e(x,t)[v]_e(x,t)\mathrm{d}s_q.$$

Da die klassische Lösung u stetig ist, gilt weiters

$$= \int_e n_t(x,t)\langle u\rangle_e(x,t)[v]_e(x,t)\mathrm{d}s_q$$
$$= \int_{\Gamma_{k\ell}} n_t(x,t)u^{\mathrm{up}}_{k\ell}(x,t)[v]_e(x,t)\mathrm{d}s_q.$$

Dabei ist der Upwind $u^{\mathrm{up}}_{k\ell}$ für ein inneres Element $\Gamma_{k\ell} \in \mathcal{I}_N$ mit $k < \ell$ definiert als

$$u^{\mathrm{up}}_{k\ell}(x,t) := \begin{cases} u_{|\tau_k}(x,t) & \text{für } n_t \geq 0, \\ u_{|\tau_\ell}(x,t) & \text{für } n_t < 0 \end{cases} \quad \text{für alle } (x,t) \in \Gamma_{k\ell} \text{ mit } k < \ell.$$

Unter Berücksichtigung der Anfangsbedingung $u(x,0) = u_0(x)$ kann die Gleichung (2.3) somit umgeschrieben werden zu:

$$\begin{aligned} &-\sum_{k=1}^{N}\int_{\tau_k} u(x,t)\frac{\partial}{\partial t}v(x,t)\mathrm{d}q - \int_Q \Delta u(x,t)v(x,t)\mathrm{d}q + \int_{\Sigma_T} u(x,t)v(x,t)\mathrm{d}s_q \\ &+ \sum_{e=\Gamma_{k\ell}\in\mathcal{I}_N}\int_{\Gamma_{k\ell}} n_t(x,t)u^{\mathrm{up}}_{k\ell}(x,t)[v]_e(x,t)\mathrm{d}s_q = \int_Q f(x,t)v(x,t)\mathrm{d}q + \int_{\Sigma_0} u_0(x,t)v(x,t)\mathrm{d}s_q. \end{aligned} \tag{2.4}$$

In ähnlicher Weise wird nun das zweite Integral auf der linken Seite von (2.4) behandelt. Dazu wird der Raum–Zeit–Zylinder Q auf die Unterteilung $\mathcal{T}_N$ aufgeteilt

$$-\int_Q \Delta u(x,t)v(x,t)\mathrm{d}q = -\sum_{k=1}^{N}\int_{\tau_k} \Delta u(x,t)v(x,t)\mathrm{d}q.$$

Die partielle Integration nach dem Ort x liefert weiters

$$= \sum_{k=1}^{N}\int_{\tau_k} \nabla_x u(x,t)\cdot\nabla_x v(x,t)\mathrm{d}q - \sum_{k=1}^{N}\int_{\partial\tau_k} \underline{n}_x\cdot\nabla_x u(x,t)v(x,t)\mathrm{d}s_q.$$

Das Aufteilen der Summation in innere Elemente und Randelemente ergibt

$$= \sum_{k=1}^{N}\int_{\tau_k} \nabla_x u(x,t)\cdot\nabla_x v(x,t)\mathrm{d}q - \sum_{e\in\mathcal{I}_N}\int_e [\underline{n}_x\cdot\nabla_x uv]_e(x,t)\mathrm{d}s_q$$
$$- \int_\Sigma \underline{n}_x\cdot\nabla_x u(x,t)v(x,t)\mathrm{d}s_q.$$

Für den Ausdruck über die inneren Kanten $e \in \mathcal{I}_N$ gilt wegen Lemma 2.7 folgende Beziehung

$$\int_e [\underline{n}_x \cdot \nabla_x uv]_e(x,t)\mathrm{d}s_q = \int_e [\underline{n}_x \cdot \nabla_x u]_e(x,t)\langle v\rangle_e(x,t)\mathrm{d}s_q + \int_e \langle \underline{n}_x \cdot \nabla_x u\rangle_e(x,t)[v]_e(x,t)\mathrm{d}s_q.$$

Da alle Ableitungen der Lösung u stetig sind, ergibt sich

$$= \int_e \langle \underline{n}_x \cdot \nabla_x u\rangle_e(x,t)[v]_e(x,t)\mathrm{d}s_q.$$

Somit gilt die Beziehung

$$\begin{aligned} -\int_Q \Delta u(x,t)v(x,t)\mathrm{d}q &= \sum_{k=1}^{N} \int_{\tau_k} \nabla_x u(x,t)\cdot\nabla_x v(x,t)\mathrm{d}q \\ &\quad - \sum_{e\in\mathcal{I}_N} \int_e \langle \underline{n}_x \cdot \nabla_x u\rangle_e(x,t)[v]_e(x,t)\mathrm{d}s_q - \int_\Sigma \underline{n}_x \cdot \nabla_x u(x,t)v(x,t)\mathrm{d}s_q. \end{aligned} \tag{2.5}$$

Um die Symmetrie und die Stabilität der resultierenden Variationsformulierung zu ermöglichen, werden weitere Gleichungen zur Gleichung (2.4) addiert. Da die Lösung u stetig ist, gilt für ein beliebiges $\varepsilon \in \mathbb{R}$

$$\varepsilon \sum_{e\in\mathcal{I}_N} \int_e [u]_e(x,t)\langle \underline{n}_x \cdot \nabla_x v\rangle_e(x,t)\mathrm{d}s_q = 0. \tag{2.6}$$

Am Rand Σ ergibt sich wegen der Dirichlet–Randbedingung die Beziehung

$$\varepsilon \sum_{\substack{e\in\mathcal{R}_N \\ e\subseteq\Sigma}} \int_e u(x,t)\underline{n}_x \cdot \nabla_x v(x,t)\mathrm{d}s_q = \varepsilon \sum_{\substack{e\in\mathcal{R}_N \\ e\subseteq\Sigma}} \int_e g(x,t)\underline{n}_x \cdot \nabla_x v(x,t)\mathrm{d}s_q. \tag{2.7}$$

Um die Stabilität gewährleisten zu können, wird die Bilinearform

$$\begin{aligned} J^{\sigma,\beta}(u,v) := \sum_{e\in\mathcal{I}_N} \frac{\sigma}{|e|^\beta} \int_e \left(\underline{n}_x(x,t)\cdot\underline{n}_x(x,t)\right)[u]_e(x,t)[v]_e(x,t)\mathrm{d}s_q \\ + \sum_{\substack{e\in\mathcal{R}_N \\ e\subseteq\Sigma}} \frac{\sigma}{|e|^\beta} \int_e u(x,t)v(x,t)\mathrm{d}s_q \end{aligned}$$

eingeführt. Dabei sei σ auf jedem inneren Element und auf jedem Randelement als konstant vorgegeben. Da u eine klassische Lösung der Aufgabenstellung (2.2) ist, gilt

$$J^{\sigma,\beta}(u,v) = \sum_{\substack{e\in\mathcal{R}_N \\ e\subseteq\Sigma}} \frac{\sigma}{|e|^\beta} \int_e g(x,t)v(x,t)\mathrm{d}s_q. \tag{2.8}$$

Durch das Einsetzen der Beziehung (2.5) in die Gleichung (2.4) und durch die Addition der Gleichungen (2.6)–(2.8) ergibt sich die folgende Variationsformulierung: Gesucht ist $u \in H^{s_x,s_t}(\mathcal{T}_N)$, sodass

$$\begin{aligned}
&-\sum_{k=1}^{N}\int_{\tau_k} u(x,t)\frac{\partial}{\partial t}v(x,t)\mathrm{d}q + \int_{\Sigma_T} u(x,t)v(x,t)\mathrm{d}s_q + \sum_{k=1}^{N}\int_{\tau_k}\nabla_x u(x,t)\cdot\nabla_x v(x,t)\mathrm{d}q\\
&-\sum_{e\in\mathcal{I}_N}\int_e \Big[\langle \underline{n}_x\cdot\nabla_x u\rangle_e(x,t)[v]_e(x,t) - \varepsilon[u]_e(x,t)\langle\underline{n}_x\cdot\nabla_x v\rangle_e(x,t)\Big]\mathrm{d}s_q\\
&-\sum_{\substack{e\in\mathcal{R}_N\\ e\subseteq\Sigma}}\int_e \Big[\underline{n}_x\cdot\nabla_x u(x,t)v(x,t) - \varepsilon u(x,t)\underline{n}_x\cdot\nabla_x v(x,t)\Big]\mathrm{d}s_q\\
&+\sum_{e=\Gamma_{k\ell}\in\mathcal{I}_N}\int_{\Gamma_{k\ell}} n_t(x,t)u^{\mathrm{up}}_{k\ell}(x,t)[v]_e(x,t)\mathrm{d}s_q + J^{\sigma,\beta}(u,v)\\
&=\int_Q f(x,t)v(x,t)\mathrm{d}q + \int_{\Sigma_0} u_0(x,t)v(x,t)\mathrm{d}s_q + \varepsilon\sum_{\substack{e\in\mathcal{R}_N\\ e\subseteq\Sigma}}\int_e g(x,t)\underline{n}_x\cdot\nabla_x v(x,t)\mathrm{d}s_q\\
&+\sum_{\substack{e\in\mathcal{R}_N\\ e\subseteq\Sigma}}\frac{\sigma}{|e|^\beta}\int_e g(x,t)v(x,t)\mathrm{d}s_q
\end{aligned} \tag{2.9}$$

für alle $v \in H^{s_x,s_t}(\mathcal{T}_N)$ erfüllt ist. Falls die Lösung u der Variationsformulierung (2.9) in $H^{2,1}(\mathcal{T}_N)$ liegt, so ist laut Konstruktion u auch eine schwache Lösung der entsprechenden Variationsformulierung (1.5), siehe Definition 1.15. Definiert man die Bilinearformen

$$\begin{aligned}
a_\varepsilon(u,v) :=& \sum_{k=1}^{N}\int_{\tau_k}\nabla_x u(x,t)\cdot\nabla_x v(x,t)\mathrm{d}q\\
&-\sum_{e\in\mathcal{I}_N}\int_e \Big[\langle \underline{n}_x\cdot\nabla_x u\rangle_e(x,t)[v]_e(x,t) - \varepsilon[u]_e(x,t)\langle\underline{n}_x\cdot\nabla_x v\rangle_e(x,t)\Big]\mathrm{d}s_q\\
&-\sum_{\substack{e\in\mathcal{R}_N\\ e\subseteq\Sigma}}\int_e \Big[\underline{n}_x\cdot\nabla_x u(x,t)v(x,t) - \varepsilon u(x,t)\underline{n}_x\cdot\nabla_x v(x,t)\Big]\mathrm{d}s_q + J^{\sigma,\beta}(u,v),\\
b(u,v) :=& \sum_{k=1}^{N}\int_{\tau_k} u(x,t)\frac{\partial}{\partial t}v(x,t)\mathrm{d}q,\\
c_0(u,v) :=& \int_{\Sigma_0} u(x,t)v(x,t)\mathrm{d}s_q,\\
c_T(u,v) :=& \int_{\Sigma_T} u(x,t)v(x,t)\mathrm{d}s_q,
\end{aligned}$$

$$d_0(u,v) := \sum_{e=\Gamma_{k\ell}\in\mathcal{I}_N} \int_{\Gamma_{k\ell}} n_t(x,t) u_{k\ell}^{\mathrm{up}}(x,t)[v]_e(x,t)\mathrm{d}s_q$$

und die Linearform

$$G_\varepsilon(v) := \varepsilon \sum_{\substack{e\in\mathcal{R}_N\\ e\subseteq\Sigma}} \int_e g(x,t)\underline{n}_x\cdot\nabla_x v(x,t)\mathrm{d}s_q + \sum_{\substack{e\in\mathcal{R}_N\\ e\subseteq\Sigma}} \frac{\sigma}{|e|^\beta}\int_e g(x,t)v(x,t)\mathrm{d}s_q,$$

so lässt sich die Variationsformulierung (2.9) schreiben als

$$-b(u,v)+a_\varepsilon(u,v)+c_T(u,v)+d_0(u,v) = \langle f,v\rangle_Q + c_0(u_0,v) + G_\varepsilon(v). \tag{2.10}$$

2.2.2 Elliptizität

Schränkt man die Variationsformulierung (2.10) auf einen diskreten Teilraum ein, so ergibt sich eine diskrete Variationsformulierung. Ein solcher Teilraum ist zum Beispiel der Raum der stückweise stetigen Polynome der Ordnung $r\in\mathbb{N}$:

$$\mathcal{D}_r(\mathcal{T}_N) := \left\{v : v_{|\tau_k}\in\mathbb{P}_r(\tau_k) \text{ für alle } \tau_k\in\mathcal{T}_N\right\}.$$

Hierbei bezeichnet $\mathbb{P}_r(\tau_k)$ den Raum der Polynome auf τ_k vom Grad kleiner oder gleich $r\in\mathbb{N}$. Im weiteren wird auf dem diskreten Teilraum $\mathcal{D}_r(\mathcal{T}_N)$ bezüglich der Energienorm

$$\begin{aligned}\|u\|_{\mathrm{DG}}^2 &:= \sum_{k=1}^N \|\nabla_x u\|_{L_2(\tau_k)}^2 + J^{\sigma,\beta}(u,u) + \frac{1}{2}\left[\|u\|_{L_2(\Sigma_0\cup\Sigma_T)}^2 + \sum_{e\in\mathcal{I}_N}\left\|\sqrt{|n_t|}[u]_e\right\|_{L_2(e)}^2\right]\\ &= \sum_{k=1}^N \|\nabla_x u\|_{L_2(\tau_k)}^2 + \sum_{e\in\mathcal{I}_N}\frac{\sigma}{|e|^\beta}\left\|\sqrt{|\underline{n}_x|}[u]_e\right\|_{L_2(e)}^2 + \sum_{\substack{e\in\mathcal{R}_N\\ e\subseteq\Sigma}}\frac{\sigma}{|e|^\beta}\|u\|_{L_2(e)}^2\\ &\quad + \frac{1}{2}\left[\|u\|_{L_2(\Sigma_0\cup\Sigma_T)}^2 + \sum_{e\in\mathcal{I}_N}\left\|\sqrt{|n_t|}[u]_e\right\|_{L_2(e)}^2\right]\end{aligned}$$

die Elliptizität der Bilinearform

$$a_{\mathrm{DG}}(u,v) := -b(u,v)+c_T(u,v)+d_0(u,v)+a_\varepsilon(u,v)$$

untersucht. Dazu werden die folgenden Lemmata benötigt:

Lemma 2.9. *Für ein inneres Element $e=\Gamma_{k\ell}\in\mathcal{I}_N, k<\ell$, mit dem Normalvektor $\underline{n}$ und $u\in\mathcal{D}_r(\mathcal{T}_N)$ gilt für den Upwind $u_{k\ell}^{\mathrm{up}}$ die Beziehung*

$$u_{k\ell}^{\mathrm{up}}(x,t) = \langle u\rangle_e(x,t) + \frac{1}{2}\mathrm{sign}(n_t)[u]_e(x,t) \quad \textit{für alle } (x,t)\in e.$$

Beweis. Die Behauptung ergibt sich durch

$$\begin{aligned}\langle u\rangle_e(x,t) + \frac{1}{2}\mathrm{sign}(n_t)[u]_e(x,t) &= \frac{1}{2}\left[u_{|\tau_k}(x,t) + u_{|\tau_\ell}(x,t) + \mathrm{sign}(n_t)\left(u_{|\tau_k}(x,t) - u_{|\tau_\ell}(x,t)\right)\right] \\ &= \begin{cases} u_{|\tau_k}(x,t) & \text{für } n_t \geq 0, \\ u_{|\tau_\ell}(x,t) & \text{für } n_t < 0 \end{cases} \\ &= u^{\mathrm{up}}_{k\ell}(x,t).\end{aligned}$$

□

Lemma 2.10. *Sei $e = \Gamma_{k\ell} \in \mathcal{I}_N$ ein inneres Element. Für $u \in \mathcal{D}_r(\mathcal{T}_N)$ gilt*

$$\left[u^2\right]_e(x,t) = 2\langle u\rangle_e(x,t)[u]_e(x,t) \quad \textit{für alle } (x,t) \in e.$$

Beweis. Das Einsetzen der Definitionen für den Sprung und den Mittelwert ergibt die Behauptung mit

$$\begin{aligned}2\langle u\rangle_e(x,t)[u]_e(x,t) &= \left(u_{|\tau_k}(x,t) + u_{|\tau_\ell}(x,t)\right)\left(u_{|\tau_k}(x,t) - u_{|\tau_\ell}(x,t)\right) \\ &= \left(u_{|\tau_k}(x,t)\right)^2 - \left(u_{|\tau_\ell}(x,t)\right)^2 \\ &= \left[u^2\right]_e(x,t).\end{aligned}$$

□

Mit Hilfe der obigen Lemmata lässt sich der folgende Satz zeigen:

Satz 2.11. *Für $u \in \mathcal{D}_r(\mathcal{T}_N)$ gilt*

$$-b(u,u) + c_T(u,u) + d_0(u,u) = \frac{1}{2}\left[\|u\|^2_{L_2(\Sigma_0\cup\Sigma_T)} + \sum_{e\in\mathcal{I}_N}\left\|\sqrt{|n_t|}[u]_e\right\|^2_{L_2(e)}\right].$$

Beweis. Das Einsetzen der Definitionen der Bilinearformen ergibt

$$\begin{aligned}-b(u,u) + c_T(u,u) + d_0(u,u) = &-\sum_{k=1}^{N}\int_{\tau_k} u(x,t)\frac{\partial}{\partial t}u(x,t)\mathrm{d}q + \int_{\Sigma_T}(u(x,t))^2\mathrm{d}s_q \\ &+ \sum_{e=\Gamma_{k\ell}\in\mathcal{I}_N}\int_e n_t u^{\mathrm{up}}_{k\ell}(x,t)[u]_e(x,t)\mathrm{d}s_q.\end{aligned}$$

Mit der Beziehung $u\frac{\partial}{\partial t}u = \frac{1}{2}\frac{\partial}{\partial t}\left(u^2\right)$ und mit Hilfe von Lemma 2.9 erhält man weiters

$$\begin{aligned}= &-\frac{1}{2}\sum_{k=1}^{N}\int_{\tau_k}\frac{\partial}{\partial t}\left[(u(x,t))^2\right]\mathrm{d}q + \int_{\Sigma_T}(u(x,t))^2\mathrm{d}s_q \\ &+ \sum_{e=\Gamma_{k\ell}\in\mathcal{I}_N}\int_e\left[n_t\langle u\rangle_e(x,t)[u]_e(x,t) + \frac{1}{2}|n_t|\,[u]^2_e(x,t)\right]\mathrm{d}s_q.\end{aligned}$$

Das Anwenden des Satzes von Gauß und Lemma 2.10 liefert

$$= -\frac{1}{2}\sum_{k=1}^{N}\int\limits_{\partial\tau_k} n_t\left[(u(x,t))^2\right]\mathrm{d}s_q + \int\limits_{\Sigma_T}(u(x,t))^2\mathrm{d}s_q$$
$$+\frac{1}{2}\sum_{e=\Gamma_{k\ell}\in\mathcal{I}_N}\int\limits_e \left[n_t\left[u^2\right]_e(x,t) + |n_t|\,[u]_e^2(x,t)\right]\mathrm{d}s_q.$$

Das Aufteilen der Summation der Ränder $\partial\tau_k$ auf die inneren Elemente und auf die Randelemente ergibt dann schließlich die Behauptung des Satzes

$$= \frac{1}{2}\left[\|u\|^2_{L_2(\Sigma_0\cup\Sigma_T)} + \sum_{e\in\mathcal{I}_N}\left\|\sqrt{|n_t|}[u]_e\right\|^2_{L_2(e)}\right].$$

□

Um nun die Elliptizität der Bilinearform $a_{\mathrm{DG}}(\cdot,\cdot)$ zeigen zu können, muss die Bilinearform $a_\varepsilon(\cdot,\cdot)$ noch geeignet abgeschätzt werden. In der Masterseminararbeit [20] beziehungsweise in [23] ist dies für den rein elliptischen Fall durchgeführt.
Betrachtet man unter Anwendung von Satz 2.11 für ein $u \in \mathcal{D}_r(\mathcal{T}_N)$ die Bilinearform $a_{\mathrm{DG}}(u,u)$, so ergibt sich:

$$\begin{aligned}
a_{\mathrm{DG}}(u,u) &= a_\varepsilon(u,u) - b(u,u) + c_T(u,u) + d_0(u,u)\\
&= \sum_{k=1}^{N}\int\limits_{\tau_k}\nabla_x u(x,t)\cdot\nabla_x u(x,t)\mathrm{d}q\\
&\quad - \sum_{e\in\mathcal{I}_N}\int\limits_e\Big[\langle\underline{n}_x\cdot\nabla_x u\rangle_e(x,t)[u]_e(x,t) - \varepsilon[u]_e(x,t)\langle\underline{n}_x\cdot\nabla_x u\rangle_e(x,t)\Big]\mathrm{d}s_q\\
&\quad - \sum_{\substack{e\in\mathcal{R}_N\\ e\subseteq\Sigma}}\int\limits_e\Big[\underline{n}_x\cdot\nabla_x u(x,t)u(x,t) - \varepsilon u(x,t)\underline{n}_x\cdot\nabla_x u(x,t)\Big]\mathrm{d}s_q + J^{\sigma,\beta}(u,u)\\
&\quad + \frac{1}{2}\left[\|u\|^2_{L_2(\Sigma_0\cup\Sigma_T)} + \sum_{e\in\mathcal{I}_N}\left\|\sqrt{|n_t|}[u]_e\right\|^2_{L_2(e)}\right]\\
&= \sum_{k=1}^{N}\|\nabla_x u\|^2_{L_2(\tau_k)} + (\varepsilon-1)\sum_{e\in\mathcal{I}_N}\int\limits_e\langle\underline{n}_x\cdot\nabla_x u\rangle_e(x,t)[u]_e(x,t)\mathrm{d}s_q\\
&\quad + (\varepsilon-1)\sum_{\substack{e\in\mathcal{R}_N\\ e\subseteq\Sigma}}\int\limits_e\underline{n}_x\cdot\nabla_x u(x,t)u(x,t)\mathrm{d}s_q + J^{\sigma,\beta}(u,u)\\
&\quad + \frac{1}{2}\left[\|u\|^2_{L_2(\Sigma_0\cup\Sigma_T)} + \sum_{e\in\mathcal{I}_N}\left\|\sqrt{|n_t|}[u]_e\right\|^2_{L_2(e)}\right]
\end{aligned}$$

$$= \|u\|_{\mathrm{DG}}^2 + (\varepsilon - 1) \sum_{e \in \mathcal{I}_N} \int_e \langle \underline{n}_x \cdot \nabla_x u \rangle_e(x,t)[u]_e(x,t) \mathrm{d}s_q$$
$$+ (\varepsilon - 1) \sum_{\substack{e \in \mathcal{R}_N \\ e \subseteq \Sigma}} \int_e \underline{n}_x \cdot \nabla_x u(x,t) u(x,t) \mathrm{d}s_q.$$

Für $\varepsilon = 1$ erhält man daraus sofort die Elliptizität bezüglich der Energienorm $\|\cdot\|_{\mathrm{DG}}$. Für $\varepsilon \in \{-1, 0\}$ müssen die Ausdrücke

$$\sum_{e \in \mathcal{I}_N} \int_e \langle \underline{n}_x \cdot \nabla_x u \rangle_e(x,t)[u]_e(x,t) \mathrm{d}s_q \quad \text{und} \quad \sum_{\substack{e \in \mathcal{R}_N \\ e \subseteq \Sigma}} \int_e \underline{n}_x \cdot \nabla_x u(x,t) u(x,t) \mathrm{d}s_q$$

geeignet nach oben abgeschätzt werden. Dazu wird die Youngsche Ungleichung benötigt.

Lemma 2.12 (Youngsche Ungleichung)**.** *Seien $p, q \in \mathbb{R}$ mit $\frac{1}{p} + \frac{1}{q} = 1$ und $p, q > 1$. Weiters seien $a, b \in \mathbb{R}$ mit $a, b \geq 0$. Dann gilt*

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}.$$

Beweis. Siehe zum Beispiel [1, Theorem 2.15]. □

Für ein $\delta > 0$ und $p = q = 2$ folgt aus der Youngschen Ungleichung der Spezialfall

$$ab = \left(a\sqrt{\delta}\right)\left(\frac{b}{\sqrt{\delta}}\right) \leq \frac{\delta}{2}a^2 + \frac{1}{2\delta}b^2 \quad \text{für alle } a, b \in \mathbb{R}. \tag{2.11}$$

Weiters werden die folgenden zwei Lemmata benötigt:

Lemma 2.13. *Sei $\tau_\ell \in \mathcal{T}_N$ und $\psi \in \mathbb{P}_r(\tau_\ell)$ mit $r \in \mathbb{N}$. Dann gilt für $e \subseteq \partial\tau_\ell$*

$$\|n_e \nabla\psi\|_{L_2(e)} \leq c_t h_\ell^{-\frac{1}{2}} \|\nabla\psi\|_{L_2(\tau_\ell)}.$$

Beweis. Mit Hilfe von [26, Theorem 4.76] und einem Skalierungsargument lässt sich die Behauptung für $d = 2$ für Dreiecke und Quadrate zeigen. Für den allgemeinen d–Simplex sei hier auf [32] verwiesen. □

Lemma 2.14. *Sei e ein inneres Element oder ein Randelement, das heißt $e \in \mathcal{I}_N \cup \mathcal{R}_N$. Weiters sei τ_ℓ ein an e angrenzendes Element. Dann gilt*

$$|e| \leq c_E h_\ell^{d-1} \leq c_E h^{d-1}.$$

Beweis. Aufgrund der Formregularität der Unterteilung $\mathcal{T}_N$ gilt

$$|e| \leq d_\ell^{d-1} \leq c_F^{d-1} r_\ell^{d-1}.$$

Da die Kugel mit Radius r_ℓ im Element τ_ℓ enthalten ist, folgt die Behauptung des Lemmas mit

$$\leq c_F^{d-1} c_B h_\ell^{d-1} = c_E h_\ell^{d-1} \leq c_E h^{d-1}.$$

□

Lemma 2.15. *Sei $e = \Gamma_{k\ell} \in \mathcal{I}_N$ ein inneres Element. Weiters sei $\beta(d-1) \geq 1$. Dann gilt für $u \in \mathcal{D}_r(\mathcal{T}_N)$:*

$$\int_e \langle \underline{n}_x \cdot \nabla_x u \rangle_e(x,t)[u]_e(x,t)\mathrm{d}s_q \leq c_t c_E^{\frac{\beta}{2}} \left(\|\nabla_x u\|^2_{L_2(\tau_k)} + \|\nabla_x u\|^2_{L_2(\tau_\ell)} \right)^{\frac{1}{2}} \frac{1}{|e|^{\frac{\beta}{2}}} \|[u]_e\|_{L_2(e)}.$$

Beweis. Sei $e = \Gamma_{k\ell} \in \mathcal{I}_N$ ein inneres Element. Laut Definition 2.6 gilt:

$$\|\langle \underline{n}_x \cdot \nabla_x u \rangle_e\|_{L_2(e)} = \left\| \frac{1}{2} \left(\underline{n}_x \cdot \nabla_x u_{|\tau_k} + \underline{n}_x \cdot \nabla_x u_{|\tau_\ell} \right) \right\|_{L_2(e)}.$$

Mit Hilfe der Dreiecksungleichung und Lemma 2.13 erhält man:

$$\begin{aligned} &\leq \frac{1}{2} \left\| \underline{n}_x \cdot \nabla_x u_{|\tau_k} \right\|_{L_2(e)} + \left\| \underline{n}_x \cdot \nabla_x u_{|\tau_\ell} \right\|_{L_2(e)} \\ &\leq \frac{c_t}{2} \left(h_k^{-\frac{1}{2}} \|\nabla_x u\|_{L_2(\tau_k)} + h_\ell^{-\frac{1}{2}} \|\nabla_x u\|_{L_2(\tau_\ell)} \right). \\ &\leq \frac{c_t h^{-\frac{1}{2}}}{2} \left(\|\nabla_x u\|_{L_2(\tau_k)} + \|\nabla_x u\|_{L_2(\tau_\ell)} \right). \end{aligned}$$

Weiters erhält man durch die Anwendung der Cauchy–Schwarzschen Ungleichung

$$\int_e \langle \underline{n}_x \cdot \nabla_x u \rangle_e(x,t)[u]_e(x,t)\mathrm{d}s_q \leq \|\langle \underline{n}_x \cdot \nabla_x u \rangle_e\|_{L_2(e)} \|[u]_e\|_{L_2(e)}.$$

Mit Hilfe der obigen Abschätzung erhält man

$$\begin{aligned} &\leq \frac{c_t h^{-\frac{1}{2}}}{2} \left(\|\nabla_x u\|_{L_2(\tau_k)} + \|\nabla_x u\|_{L_2(\tau_\ell)} \right) \|[u]_e\|_{L_2(e)} \\ &= \frac{c_t \, |e|^{\frac{\beta}{2}} \, h^{-\frac{1}{2}}}{2} \left(\|\nabla_x u\|_{L_2(\tau_k)} + \|\nabla_x u\|_{L_2(\tau_\ell)} \right) \frac{1}{|e|^{\frac{\beta}{2}}} \|[u]_e\|_{L_2(e)}. \end{aligned}$$

Die Anwendung von Lemma 2.14 liefert:

$$\leq \frac{c_t c_E^{\frac{\beta}{2}} h^{\frac{\beta}{2}(d-1)-\frac{1}{2}}}{2} \left(\|\nabla_x u\|_{L_2(\tau_k)} + \|\nabla_x u\|_{L_2(\tau_\ell)} \right) \frac{1}{|e|^{\frac{\beta}{2}}} \|[u]_e\|_{L_2(e)}.$$

Mit der Voraussetzung $\beta(d-1) \geq 1$ und der Annahme $h \leq 1$ erhält man weiters

$$\leq \frac{c_t c_E^{\frac{\beta}{2}}}{2} \left(\|\nabla_x u\|_{L_2(\tau_k)} + \|\nabla_x u\|_{L_2(\tau_\ell)} \right) \frac{1}{|e|^{\frac{\beta}{2}}} \|[u]_e\|_{L_2(e)}.$$

Die Behauptung folgt nun durch das Anwenden der Cauchy–Schwarzschen Ungleichung für Summen:

$$\leq c_t c_E^{\frac{\beta}{2}} \left(\|\nabla_x u\|^2_{L_2(\tau_k)} + \|\nabla_x u\|^2_{L_2(\tau_\ell)} \right)^{\frac{1}{2}} \frac{1}{|e|^{\frac{\beta}{2}}} \|[u]_e\|_{L_2(e)}.$$

□

Analog zum Lemma 2.15 lässt sich das nächste Lemma zeigen:

Lemma 2.16. *Ist $e \in \mathcal{R}_N$ ein Randelement mit dem benachbarten Element $\tau_k \in \mathcal{T}_N$ und sei weiters $\beta(d-1) \geq 1$. Dann gilt für $u \in \mathcal{D}_r(\mathcal{T}_N)$ die Abschätzung*

$$\int_e \underline{n}_x \cdot \nabla_x u(x,t) u(x,t) \mathrm{d}s_q \leq c_t c_E^{\frac{\beta}{2}} \|\nabla_x u\|_{L_2(\tau_k)} \frac{1}{|e|^{\frac{\beta}{2}}} \|u\|_{L_2(e)}.$$

Beweis. Sei $e \in \mathcal{R}_N$ ein Randelement und $\tau_k \in \mathcal{T}_N$ das an e angrenzende Element. Dann gilt

$$\int_e \underline{n}_x \cdot \nabla_x u(x,t) u(x,t) \mathrm{d}s_x \leq \|\underline{n}_x \cdot \nabla_x u\|_{L_2(e)} \|u\|_{L_2(e)}.$$

Die Anwendung von Lemma 2.13 liefert

$$\leq c_t h_k^{-\frac{1}{2}} \|\nabla_x u\|_{L_2(\tau_k)} \|u\|_{L_2(e)}.$$

Weiters folgt mit Lemma 2.14 die Abschätzung

$$\leq c_t c_E^{\frac{\beta}{2}} h^{-\frac{1}{2}+\frac{\beta}{2}(d-1)} \|\nabla_x u\|_{L_2(\tau_k)} \frac{1}{|e|^{\frac{\beta}{2}}} \|u\|_{L_2(e)}.$$

Die Behauptung folgt aus der Voraussetzung $\beta(d-1) \geq 1$ und der Annahme, dass $h \leq 1$ gilt:

$$\leq c_t c_E^{\frac{\beta}{2}} \|\nabla_x u\|_{L_2(\tau_k)} \frac{1}{|e|^{\frac{\beta}{2}}} \|u\|_{L_2(e)}.$$

□

Mit Hilfe von Lemma 2.15 und Lemma 2.16 lassen sich nun die folgenden Summen abschätzen.

Lemma 2.17. *Sei* $\sigma^* = \inf_{(x,t)\in Q} \sigma(x,t) > 0$ *und* $n_0 \in \mathbb{N}$ *sei die maximale Anzahl an Nachbarn eines Elements* $\tau_k \in \mathcal{T}_N$. *Weiters sei* $\beta(d-1) \geq 1$ *und* $u \in \mathcal{D}_r(\mathcal{T}_N)$, *dann gilt für alle* $\delta > 0$ *die Ungleichung*

$$\sum_{e\in\mathcal{I}_N} \int_e \langle \underline{n}_x \cdot \nabla_x u \rangle_e(x,t)[u]_e(x,t)\mathrm{d}s_q + \sum_{\substack{e\in\mathcal{R}_N \\ e\subseteq\Sigma}} \int_e \underline{n}_x \cdot \nabla_x u(x,t)u(x,t)\mathrm{d}s_q$$

$$\leq \frac{\delta}{2}\sum_{k=1}^{N} \|\nabla_x u\|^2_{L_2(\tau_k)} + \frac{n_0 c_t^2 c_E^\beta}{2\delta\sigma^*}\left[\sum_{e\in\mathcal{I}_N} \frac{\sigma}{|e|^\beta}\|[u]_e\|^2_{L_2(e)} + \sum_{\substack{e\in\mathcal{R}_N \\ e\subseteq\Sigma}} \frac{\sigma}{|e|^\beta}\|u\|^2_{L_2(e)}\right].$$

Beweis. Die Anwendung von Lemma 2.15 und Lemma 2.16 auf die einzelnen Summanden ergibt:

$$\sum_{e\in\mathcal{I}_N} \int_e \langle \underline{n}_x \cdot \nabla_x u \rangle_e(x,t)[u]_e(x,t)\mathrm{d}s_q + \sum_{\substack{e\in\mathcal{R}_N \\ e\subseteq\Sigma}} \int_e \underline{n}_x \cdot \nabla_x u(x,t)u(x,t)\mathrm{d}s_q$$

$$\leq \sum_{e\in\mathcal{I}_N} c_t c_E^{\frac{\beta}{2}} \left(\|\nabla_x u\|^2_{L_2(\tau_k)} + \|\nabla_x u\|^2_{L_2(\tau_\ell)}\right)^{\frac{1}{2}} \frac{1}{|e|^{\frac{\beta}{2}}}\|[u]_e\|_{L_2(e)}$$

$$+ \sum_{\substack{e\in\mathcal{R}_N \\ e\subseteq\Sigma}} c_t c_E^{\frac{\beta}{2}} \|\nabla_x u\|_{L_2(\tau_k)} \frac{1}{|e|^{\frac{\beta}{2}}}\|u\|_{L_2(e)}.$$

Mit Hilfe der Cauchy–Schwarzschen Ungleichung für Summen erhält man

$$\leq \left[\sum_{e\in\mathcal{I}_N} c_t^2 c_E^\beta \left(\|\nabla_x u\|^2_{L_2(\tau_k)} + \|\nabla_x u\|^2_{L_2(\tau_\ell)}\right) + \sum_{\substack{e\in\mathcal{R}_N \\ e\subseteq\Sigma}} c_t^2 c_E^\beta \|\nabla_x u\|^2_{L_2(\tau_k)}\right]^{\frac{1}{2}} \cdot$$

$$\cdot \left[\sum_{e\in\mathcal{I}_N} \frac{1}{|e|^\beta}\|[u]_e\|^2_{L_2(e)} + \sum_{\substack{e\in\mathcal{R}_N \\ e\subseteq\Sigma}} \frac{1}{|e|^\beta}\|u\|^2_{L_2(e)}\right]^{\frac{1}{2}}.$$

Da jedes Element maximal n_0 Nachbarn besitzt, gilt

$$\leq \left[n_0 c_t^2 c_E^\beta \sum_{k=1}^{N} \|\nabla_x u\|^2_{L_2(\tau_k)}\right]^{\frac{1}{2}} \left[\sum_{e\in\mathcal{I}_N} \frac{1}{|e|^\beta}\|[u]_e\|^2_{L_2(e)} + \sum_{\substack{e\in\mathcal{R}_N \\ e\subseteq\Sigma}} \frac{1}{|e|^\beta}\|u\|^2_{L_2(e)}\right]^{\frac{1}{2}}$$

$$= \left[\sum_{k=1}^{N} \|\nabla u\|^2_{L_2(\tau_k)}\right]^{\frac{1}{2}} \sqrt{n_0} c_t c_E^{\frac{\beta}{2}} \left[\sum_{e\in\mathcal{I}_N} \frac{1}{|e|^\beta}\|[u]_e\|^2_{L_2(e)} + \sum_{\substack{e\in\mathcal{R}_N \\ e\subseteq\Sigma}} \frac{1}{|e|^\beta}\|u\|^2_{L_2(e)}\right]^{\frac{1}{2}}.$$

Für ein $\delta > 0$ erhält man mit dem Spezialfall der Youngschen Ungleichung (2.11) die Ungleichung

$$\leq \frac{\delta}{2} \sum_{k=1}^{N} \|\nabla_x u\|^2_{L_2(\tau_k)} + \frac{n_0 c_t^2 c_E^\beta}{2\delta} \left[\sum_{e \in \mathcal{I}_N} \frac{1}{|e|^\beta} \|[u]_e\|^2_{L_2(e)} + \sum_{\substack{e \in \mathcal{R}_N \\ e \subseteq \Sigma}} \frac{1}{|e|^\beta} \|u\|^2_{L_2(e)} \right].$$

Mit $\sigma^* = \inf_{(x,t) \in Q} \sigma(x) > 0$ ergibt sich schließlich die Behauptung des Lemmas,

$$\leq \frac{\delta}{2} \sum_{k=1}^{N} \|\nabla_x u\|^2_{L_2(\tau_k)} + \frac{n_0 c_t^2 c_E^\beta}{2\delta\sigma^*} \left[\sum_{e \in \mathcal{I}_N} \frac{\sigma}{|e|^\beta} \|[u]_e\|^2_{L_2(e)} + \sum_{\substack{e \in \mathcal{R}_N \\ e \subseteq \Sigma}} \frac{\sigma}{|e|^\beta} \|u\|^2_{L_2(e)} \right].$$

□

Mit Hilfe von Lemma 2.17 lässt sich die Bilinearform $a_\varepsilon(\cdot,\cdot)$ geeignet nach unten abschätzen.

Satz 2.18. *Sei $\varepsilon \in \{-1, 0, 1\}$. Falls der Parameter σ groß genug ist, gilt*

$$a_\varepsilon(u,u) \geq c_1^{a_\varepsilon} \left[\sum_{k=1}^{N} \|\nabla u\|^2_{L_2(\tau_k)} + J^{\sigma,\beta}(u,u) \right] \quad \textit{für alle } u \in \mathcal{D}_r(\mathcal{T}_N).$$

Beweis. Für ein $u \in \mathcal{D}_r(\mathcal{T}_N)$ gilt für die Bilinearform $\mathrm{a}_\varepsilon(\cdot,\cdot)$ die Darstellung

$$\begin{aligned} \mathrm{a}_\varepsilon(u,u) = &\sum_{k=1}^{N} \|\nabla_x u\|^2_{L_2(\tau_k)} + (\varepsilon - 1) \sum_{e \in \mathcal{I}_N} \int_e \langle \underline{n}_x \cdot \nabla_x u \rangle_e(x,t) [u]_e(x,t) \mathrm{d}s_q \\ &+ (\varepsilon - 1) \sum_{\substack{e \in \mathcal{R}_N \\ e \subseteq \Sigma}} \int_e \underline{n}_x \cdot \nabla_x u(x,t) u(x,t) \mathrm{d}s_q + J^{\sigma,\beta}(u,u). \end{aligned}$$

Ist $\varepsilon = 1$ so ergibt sich die behauptete Abschätzung mit der Konstante $c_1^{a_\varepsilon} = 1$. Für $\varepsilon \in \{-1, 0\}$ lässt sich die Bilinearform mit Hilfe von Lemma 2.17 nach unten abschätzen durch

$$\begin{aligned} \mathrm{a}_\varepsilon(u,u) &\geq \sum_{k=1}^{N} \|\nabla_x u\|^2_{L_2(\tau_k)} + J^{\sigma,\beta}(u,u) \\ &\quad + (\varepsilon - 1) \left[\frac{\delta}{2} \sum_{k=1}^{N} \|\nabla_x u\|^2_{L_2(\tau_k)} + \frac{n_0 c_t^2 c_E^\beta}{2\delta\sigma^*} \left[\sum_{e \in \mathcal{I}_N} \frac{\sigma}{|e|^\beta} \|[u]_e\|^2_{L_2(e)} + \sum_{\substack{e \in \mathcal{R}_N \\ e \subseteq \Sigma}} \frac{\sigma}{|e|^\beta} \|u\|^2_{L_2(e)} \right] \right] \\ &\geq \sum_{k=1}^{N} \|\nabla_x u\|^2_{L_2(\tau_k)} + J^{\sigma,\beta}(u,u) + (\varepsilon - 1) \left[\frac{\delta}{2} \sum_{k=1}^{N} \|\nabla_x u\|^2_{L_2(\tau_k)} + \frac{n_0 c_t^2 c_E^\beta}{2\delta\sigma^*} J^{\sigma,\beta}(u,u) \right] \end{aligned}$$

$$= \left(1 + (\varepsilon - 1)\frac{\delta}{2}\right) \sum_{k=1}^{N} \|\nabla_x u\|^2_{L_2(\tau_k)} + \left(1 + (\varepsilon - 1)\frac{n_0 c_t^2 c_E^\beta}{2\delta\sigma_0^*}\right) J^{\sigma,\beta}(u,u)$$

$$\geq \min\left\{1 + (\varepsilon - 1)\frac{\delta}{2}, 1 + (\varepsilon - 1)\frac{n_0 c_t^2 c_E^\beta}{2\delta\sigma_0^*}\right\} \left[\sum_{k=1}^{N} \|\nabla_x u\|^2_{L_2(\tau_k)} + J^{\sigma,\beta}(u,u)\right].$$

Wählt man nun für $\varepsilon = -1$ den noch frei wählbaren Parameter $\delta = \frac{1}{2}$, so ergibt sich für $\sigma^* \geq 4n_0 c_t^2 c_E^\beta$:

$$c_1^{a_\varepsilon} = \min\left\{1 + (\varepsilon - 1)\frac{\delta}{2}, 1 + (\varepsilon - 1)\frac{n_0 c_t^2 c_E^\beta}{2\delta\sigma_0^*}\right\} = \frac{1}{2}.$$

Ist $\varepsilon = 0$ und $\delta = 1$ so erhält man mit $\sigma^* \geq n_0 c_t^2 c_E^\beta$ ebenfalls

$$c_1^{a_\varepsilon} = \min\left\{1 + (\varepsilon - 1)\frac{\delta}{2}, 1 + (\varepsilon - 1)\frac{n_0 c_t^2 c_E^\beta}{2\delta\sigma_0^*}\right\} = \frac{1}{2}.$$

Somit ergibt sich die Behauptung des Satzes mit den jeweiligen Konstanten $c_1^{a_\varepsilon}$, das heißt

$$a_\varepsilon(u,u) \geq c_1^{a_\varepsilon} \left[\sum_{k=1}^{N} \|\nabla u\|^2_{L_2(\tau_k)} + J^{\sigma,\beta}(u,u)\right].$$

□

Mit Hilfe von Satz 2.18 lässt sich schließlich die Elliptizität der Bilinearform $\mathrm{a}_{DG}(\cdot,\cdot)$ auf $\mathcal{D}_r(\mathcal{T}_N)$ zeigen:

Satz 2.19. *Die Bilinearform $\mathrm{a}_{DG}(\cdot,\cdot)$ ist für $\varepsilon \in \{-1,0,1\}$ elliptisch auf $\mathcal{D}_r(\mathcal{T}_N)$ bezüglich der Energienorm $\|\cdot\|_{\mathrm{DG}}$, sofern für $\varepsilon \in \{-1,0\}$ der Parameter σ groß genug gewählt wird. Das heißt, es exisitiert eine Konstante $c_1^{\mathrm{a}_{DG}} > 0$, sodass*

$$\mathrm{a}_{DG}(u,u) \geq c_1^{\mathrm{a}_{DG}} \|u\|^2_{\mathrm{DG}} \quad \textit{für alle } u \in \mathcal{D}_r(\mathcal{T}_N)$$

gilt.

Beweis. Die Anwendung von Satz 2.11 und Satz 2.18 ergibt die Behauptung mit

$$\begin{aligned}
\mathrm{a}_{DG}(u,u) &= a_\varepsilon(u,v) - b(u,v) + c_T(u,v) + d_0(u,v) \\
&\geq c_1^{a_\varepsilon} \left[\sum_{k=1}^{N} \|\nabla u\|^2_{L_2(\tau_k)} + J^{\sigma,\beta}(u,u)\right] + \frac{1}{2}\left[\|u\|^2_{L_2(\Sigma_0 \cup \Sigma_T)} + \sum_{e \in \mathcal{I}_N} \left\|\sqrt{|n_t|}[u]_e\right\|^2_{L_2(e)}\right] \\
&\geq \min\left\{1, c_1^{a_\varepsilon}\right\} \|u\|^2_{\mathrm{DG}}.
\end{aligned}$$

□

Mit Hilfe der Elliptizität der Bilinearform $\mathrm{a}_{DG}(\cdot,\cdot)$ folgt schließlich die eindeutige Lösbarkeit der diskreten Variationsformulierung. Weiters lässt sich wie in [15, 23, 29] der Fehler in der Energienorm zeigen:

Satz 2.20. *Sei $u \in H^{s_x,s_t}(\mathcal{T}_N)$ die Lösung der Variationsformulierung (2.10) und $u_h \in \mathcal{D}_r(\mathcal{T}_N)$ die diskrete Lösung. Dann gilt die Fehlerabschätzung*

$$\|u - u_h\|_{\mathrm{DG}} \leq c h^{\min\{k+1,s_x,s_t\}-1} \|u\|_{H^{s_x,s_t}(\mathcal{T}_N)}.$$

2.2.3 Adjungierte Zustandsgleichung

Für eine allgemeine Form der adjungierten Differentialgleichung

$$\begin{aligned} -\frac{\partial}{\partial t}p(x,t) - \Delta p(x,t) &= f(x,t) \qquad \text{für} \quad (x,t) \in Q, \\ p(x,t) &= g(x,t) \qquad \text{für} \quad (x,t) \in \Sigma, \\ p(x,T) &= p_T(x) \qquad \text{für} \quad (x,T) \in \Sigma_T \end{aligned} \tag{2.12}$$

ergibt sich eine ähnliche Variationsformulierung wie für die Zustandsgleichung (2.2), da sich diese nur bis auf ein Vorzeichen und um die Bedingung $p(x,T) = p_T(x)$ von der Differentialgleichung (2.2) unterscheidet. Da für die adjungierte Differentialgleichung (2.12) der Endwert p_T bekannt ist, werden anstelle der Upwind–Terme Downwind-Terme verwendet:

$$u_{k\ell}^{\mathrm{down}}(x,t) := \begin{cases} u_{|\tau_\ell}(x,t) & \text{für } \underline{n}_t \geq 0, \\ u_{|\tau_k}(x,t) & \text{für } \underline{n}_t < 0 \end{cases} \quad \text{für alle } (x,t) \in \Gamma_{k\ell} \text{ mit } k < \ell.$$

Für die Variationsformulierung der adjungierten Differentialgleichung (2.12) wird deshalb anstelle der Bilinearform $d_0(\cdot,\cdot)$, die Bilinearform

$$d_T(u,v) := \sum_{e=\Gamma_{k\ell}\in\mathcal{I}_N} \int_{\Gamma_{k\ell}} n_t(x,t) u_{k\ell}^{\mathrm{down}}(x,t)[v]_e(x,t)\mathrm{d}s_q$$

eingeführt. Somit lautet die Variationsformulierung der allgemeinen adjungierten Gleichung (2.12): Gesucht ist $p \in H^{s_x,s_t}(\mathcal{T}_N)$, sodass

$$b(p,q) + a_\varepsilon(p,q) + c_0(p,q) - d_T(p,q) = \langle f, q\rangle_Q + c_T(p_T,q) + G_\varepsilon(q) \tag{2.13}$$

für alle $q \in H^{s_x,s_t}(\mathcal{T}_N)$ erfüllt ist.
Analog wie in Abschnitt 2.2.2 ergibt sich die Elliptizität der Bilinearform

$$\mathrm{a}_{DG}^*(\cdot,\cdot) := a_\varepsilon(\cdot,\cdot) + b(\cdot,\cdot) + c_0(\cdot,\cdot) - d_T(\cdot,\cdot)$$

auf $\mathcal{D}_r(\mathcal{T}_N)$ bezüglich der Energienorm $\|\cdot\|_{\mathrm{DG}}$.

2.2.4 Optimalitätssystem

Die Variationsformulierung des Optimalitätssystems (1.24)–(1.26) lässt sich nun schreiben als: Gesucht sind $u, p \in H^{s_{1,x},s_{1,t}}(\mathcal{T}_N)$ und $z \in H^{s_{2,x},s_{2,t}}(\mathcal{T}_N)$, sodass die Gleichungen

$$-b(u,v) + a_\varepsilon(u,v) + c_T(u,v) + d_0(u,v) - \langle z, v\rangle_Q = c_0(u_0,v) + G_\varepsilon(v), \tag{2.14}$$

$$b(p,q) + a_\varepsilon(p,q) + c_0(p,q) - d_T(p,q) - c_T(u,q) = -c_T(\bar{u},q), \tag{2.15}$$

$$\langle p, w\rangle_Q + \varrho\langle z, w\rangle_Q = 0 \tag{2.16}$$

für alle $v, q \in H^{s_{1,x},s_{1,t}}(\mathcal{T}_N)$ und $w \in H^{s_{2,x},s_{2,t}}(\mathcal{T}_N)$ erfüllt sind. Einschränken der Variationsformulierungen (2.14)–(2.16) auf diskrete Teilräume liefert die diskreten Variationsformulierungen des Optimalitätssystems. Diese sind äquivalent zu einem linearen Gleichungssystem, welches im nächsten Abschnitt angegeben wird. Da die diskrete Zustandsgleichung und die diskrete adjungierte Zustandsgleichung eindeutig lösbar sind, ergibt sich daraus auch die eindeutige Lösbarkeit des Optimalitätssystems.

2.3 Lineares Gleichungssystem

In diesem Abschnitt werden die Variationsformulierungen (2.14)–(2.16) auf diskrete Teilräume eingeschränkt. Dadurch erhält man diskrete Variationsformulierungen für das Optimalitätssystems (1.24)–(1.26). Diese sind äquivalent zu einem linearen Gleichungssystem. Für den Zustand $u \in H^{s_{1,x},s_{1,t}}(\mathcal{T}_N)$ und für den adjungierten Zustand $p \in H^{s_{1,x},s_{1,t}}(\mathcal{T}_N)$ wird hier ein gemeinsamer diskreter Teilraum

$$V_h(\mathcal{T}_N) := \left\{v : v_{|\tau_\ell} \in \mathbb{P}_{r_1}(\tau_\ell) \text{ für alle } \tau_\ell \in \mathcal{T}_N\right\} \subset H^{s_{1,x},s_{1,t}}(\mathcal{T}_N)$$

verwendet. Die Steuerung $z \in H^{s_{2,x},s_{2,t}}(\mathcal{T}_N)$ wird auf den diskreten Teilraum

$$W_h(\mathcal{T}_N) := \left\{v : v_{|\tau_\ell} \in \mathbb{P}_{r_2}(\tau_\ell) \text{ für alle } \tau_\ell \in \mathcal{T}_N\right\} \subset H^{s_{2,x},s_{2,t}}(\mathcal{T}_N)$$

eingeschränkt. Die Teilräume $V_h(\mathcal{T}_N)$ und $W_h(\mathcal{T}_N)$ lassen sich durch ihre Basisfunktionen aufspannen mit

$$V_h(\mathcal{T}_N) = \operatorname{span}\{\varphi_i\}_{i=1}^{N_v} \quad \text{und} \quad W_h(\mathcal{T}_N) = \operatorname{span}\{\psi_i\}_{i=1}^{N_w}.$$

Schränkt man die Lösungen $u, p \in H^{s_{1,x},s_{1,t}}(\mathcal{T}_N)$ und $z \in H^{s_{2,x},s_{2,t}}(\mathcal{T}_N)$ auf die jeweiligen diskreten Teilräume ein, also

$$u_h(x,t) = \sum_{i=1}^{N_v} u_i\varphi_i(x,t), \quad p_h(x,t) = \sum_{i=1}^{N_v} p_i\varphi_i(x,t) \quad \text{und} \quad z_h(x,t) = \sum_{k=1}^{N_w} z_k\psi_k(x,t),$$

so erhält man aus den Gleichungen (2.14)–(2.16) das lineare Gleichungssystem

$$\begin{pmatrix} C_{T,h} & -B_h - A_h - C_{0,h} + D_{T,h} & 0 \\ -B_h + A_h + C_{T,h} + D_{0,h} & 0 & -M_h^\top \\ 0 & M_h & \varrho\widetilde{M}_h \end{pmatrix} \begin{pmatrix} \underline{u} \\ \underline{p} \\ \underline{z} \end{pmatrix} = \begin{pmatrix} \underline{f}_1 \\ \underline{f}_2 \\ \underline{0} \end{pmatrix}. \tag{2.17}$$

Dabei ergeben sich die Einträge der einzelnen Matrizen durch

$$\begin{aligned} A_h[i,j] &= a_\varepsilon(\varphi_j,\varphi_i), & B_h[i,j] &= b(\varphi_j,\varphi_i),\\ C_{0,h}[i,j] &= c_0(\varphi_j,\varphi_i), & C_{T,h}[i,j] &= c_T(\varphi_j,\varphi_i),\\ D_{0,h}[i,j] &= d_0(\varphi_j,\varphi_i), & D_{T,h}[i,j] &= d_T(\varphi_j,\varphi_i), \end{aligned}$$

für $i,j = 1,\dots,N_v$ und

$$M_h[k,i] = \langle \varphi_i, \psi_k \rangle_Q, \qquad \widetilde{M}_h[k,\ell] = \langle \psi_\ell, \psi_k \rangle_Q$$

für $i = 1,\dots,N_v$ und $k,\ell = 1,\dots,N_w$. Die Vektoren auf der rechten Seite von (2.17) sind dabei gegeben durch

$$\begin{aligned} \underline{f}_1[i] &= c_T(\bar{u},\varphi_i),\\ \underline{f}_2[i] &= c_0(u_0,\varphi_i) + G_\varepsilon(\varphi_i) \end{aligned}$$

für $i = 1,\dots,N_v$. Falls die diskreten Teilräume $V_h(\mathcal{T}_N)$ und $W_h(\mathcal{T}_N)$ gleich sind, ergibt sich für die Massematrizen die Beziehung

$$\widetilde{M}_h = M_h = M_h^\top.$$

Somit lässt sich das Gleichungssystem (2.17) umschreiben zu

$$\begin{pmatrix} -B_h + A_h + C_{T,h} + D_{0,h} & -M_h \\ C_{T,h} & \varrho\,[B_h + A_h + C_{0,h} - D_{T,h}] \end{pmatrix} \begin{pmatrix} \underline{u} \\ \underline{z} \end{pmatrix} = \begin{pmatrix} \underline{f}_2 \\ \underline{f}_1 \end{pmatrix}, \tag{2.18}$$

wobei für den Koeffizientenvektor $\underline{p}$ des adjungierten Zustands die Beziehung

$$\underline{p} = -\varrho \underline{z}$$

gilt. Das lineare Gleichungssystem (2.18), welches eine nichtsymmetrische Systemmatrix besitzt, kann zum Beispiel mit dem Verfahren des verallgemeinerten minimalen Residuums (GMRES) oder mit dem stabilisierten Gradientenverfahren biorthogonaler Richtungen (BiCGStab) gelöst werden. Dazu siehe zum Beispiel [27]. Für eine mögliche Vorkonditionierung der Zustandsgleichung beziehungsweise der adjungierten Zustandsgleichung sei hier auf [12] verwiesen.
Die in Kapitel 4 angegebenen numerischen Beispiele zeigen für den Fehler in der L_2–Norm für den Zustand u und der Steuerung z eine optimale Konvergenzordnung. In dieser Arbeit wird für das Optimalitätssystem keine Fehleranalysis durchgeführt. Diese folgt jedoch durch Standardargumente, siehe [14].

3 Triangulierungen im vierdimensionalen Raum

Die im vorigen Kapitel hergeleiteten Variationsformulierungen benötigen eine Zerlegung des Raum–Zeit–Zylinders $Q \in \mathbb{R}^{d+1}$ in Finite Elemente. Für $d = 3$ wird somit eine Zerlegung des Raum–Zeit–Zylinders Q im vierdimensionalen Raum benötigt. In [5] wird die Konstruktion einer Zerlegung eines Raum–Zeit–Zylinders $Q \subset \mathbb{R}^4$ in Pentatope beschrieben. In diesem Kapitel wird ausgehend von einer zulässigen Ausgangszerlegung im vierdimensionalen Raum die weitere zulässige Verfeinerung behandelt. Dazu wird vorerst die Zerlegung eines Tetraeders untersucht. Anschließend wird dann die Zerlegung eines Pentatops behandelt. Am Ende dieses Kapitels wird kurz eine Möglichkeit vorgestellt, wie diese Triangulierungen visualisiert werden können. Dieses Kapitel basiert auf der Masterprojektarbeit [21].

3.1 Zerlegungen

In diesem Abschnitt wird zuerst die Zerlegung eines Tetraeders behandelt. Aufbauend auf die Zerlegung des Tetraeders wird in ähnlicher Weise die Zerlegung eines Pentatops untersucht. Am Ende dieses Abschnitts wird anschaulich die Zerlegung eines Hyperwürfels erlärkt.

3.1.1 Zerlegung eines Tetraeders

Durch die folgende Definition wird ein Tetraeder im vierdimensionalen Raum beschrieben:

Definition 3.1 (Tetraeder)**.** *Gegeben seien die Knoten $x_1, \ldots, x_4 \in \mathbb{R}^4$, sodass die Vektoren*

$$\{x_2 - x_1, x_3 - x_1, x_4 - x_1\}$$

linear unabhängig sind. Die konvexe Hülle dieser Punkte, also

$$\sigma := \operatorname{konv}\left(\{x_1, \ldots, x_4\}\right),$$

bildet dann einen Tetraeder.

Der Referenztetraeder wird durch die nächste Definition beschrieben:

Definition 3.2 (Referenztetraeder)**.** *Der* Referenztetraeder $\hat{\sigma}$ *ist gegeben durch*

$$\begin{aligned}\hat{\sigma} := \big\{\xi = (\xi_1, \xi_2, \xi_3)^\top \in \mathbb{R}^3 : \quad & 0 \le \xi_1 \le 1, \\ & 0 \le \xi_2 \le 1 - \xi_1, \\ & 0 \le \xi_3 \le 1 - \xi_1 - \xi_2\big\}.\end{aligned}$$

Für einen Tetraeder σ mit den Knoten $\{x_1, \ldots, x_4\}$ kann mit Hilfe der Transformationsmatrix

$$J_T := \begin{pmatrix} x_{2,1} - x_{1,1} & x_{3,1} - x_{1,1} & x_{4,1} - x_{1,1} \\ x_{2,2} - x_{1,2} & x_{3,2} - x_{1,2} & x_{4,2} - x_{1,2} \\ x_{2,3} - x_{1,3} & x_{3,3} - x_{1,3} & x_{4,3} - x_{1,3} \\ x_{2,4} - x_{1,4} & x_{3,4} - x_{1,4} & x_{4,4} - x_{1,4} \end{pmatrix}$$

jeder Punkt $x \in \sigma$ durch

$$x = x_1 + J_T \xi \quad \text{für } \xi \in \hat{\sigma}$$

dargestellt werden. Die Matrix J_T ist nicht invertierbar, da diese nicht quadratisch ist. Nach Definition 3.1 besitzt die Transformationsmatrix J_T genau drei linear unabhängige Zeilen. Durch Streichen einer linear abhängigen Zeile $J_{T,\ell}$ aus J_T ergibt sich eine invertierbare Matrix $\tilde{J}_T$. Jeder Punkt $\xi \in \hat{\sigma}$ lässt sich nun durch

$$\xi = \tilde{J}_T^{-1}(\tilde{x}_1 - \tilde{x})$$

ausdrücken. Dabei erhält man die Vektoren $\tilde{x}_1$ und $\tilde{x}$ durch Streichen der ℓ–ten Komponente von x_1 und $x \in \sigma$. Die Transformation zwischen dem Referenztetraeder $\hat{\sigma}$ und einem Tetraeder $\sigma \subset \mathbb{R}^4$ wird beim Aufstellen des linearen Gleichungssystems (2.17) benötigt. Weiters wird der Normalenvektor eines inneren Elements beziehungsweise eines Randelements benötigt. Dieser lässt sich für einen Tetraeder wie in der nächsten Definition angegeben berechnen:

Definition 3.3 (Normalenvektor)**.** *Sei $\sigma \subset \mathbb{R}^4$ ein Tetraeder mit den Knoten $\{x_1, x_2, x_3, x_4\}$. Dann kann der* Normalenvektor $\underline{n}$ *durch*

$$\underline{n} := \begin{vmatrix} \underline{e}_1 & \underline{e}_2 & \underline{e}_3 & \underline{e}_4 \\ x_{2,1} - x_{1,1} & x_{2,2} - x_{1,2} & x_{2,3} - x_{1,3} & x_{2,4} - x_{1,4} \\ x_{3,1} - x_{1,1} & x_{3,2} - x_{1,2} & x_{3,3} - x_{1,3} & x_{3,4} - x_{1,4} \\ x_{4,1} - x_{1,1} & x_{4,2} - x_{1,2} & x_{4,3} - x_{1,3} & x_{4,4} - x_{1,4} \end{vmatrix}$$

berechnet werden. Dabei sind $\underline{e}_1, \ldots, \underline{e}_4 \in \mathbb{R}^4$ die Einheitsvektoren im $\mathbb{R}^4$, welche für die Berechnung der Determinante formal als Skalar aufgefasst werden.

Nun soll die Zerlegung eines Tetraeders in acht weitere Teiltetraeder untersucht werden. Dabei werden die Knoten des Tetraeders von 1 bis 4 durchnummeriert, siehe Abbildung 3.1. Um den Tetraeder weiter zu verfeinern werden alle sechs Kanten halbiert. Dadurch erhält man sechs neue Knoten. Die Nummerierung dieser Knoten ist hier von 5 bis 10 festgelegt, siehe Abbildung 3.1. Die neuen Tetraeder in den Eckpunkten 1 bis 4 können so

gewählt werden, dass diese die gleichen Innenwinkel wie der Ausgangstetraeder aufweisen. Das Herausnehmen eines Ecktetraeders mit den Knoten $\{6, 8, 3, 10\}$ ist in Abbildung 3.2 dargestellt. Das Entfernen aller Ecktetraeder ergibt ein unregelmäßiges Oktaeder, welches noch weiter zu zerteilen ist (siehe Abbildung 3.3). Das Oktaeder kann durch Halbieren in zwei Pyramiden zerlegt werden. Es gibt drei Möglichkeiten, den Oktaeder zu halbieren, siehe Abbildungen 3.4–3.5. Die vier weiteren Tetraeder erhält man durch die Halbierung der zwei Pyramiden. Eine der drei Kanten

$$\{(x_5, x_{10}), (x_6, x_9), (x_7, x_8)\} \tag{3.1}$$

fixiert nun die Zerlegung des Oktaeders in vier weitere Tetraeder. Diese drei Kanten sind in Abbildung 3.6 dargestellt. Ist eine dieser drei Kanten (3.1) fixiert, so ist auch die Zerlegung des Teraeders eindeutig bestimmt. Dies motiviert nun die folgende Definition:

Definition 3.4 (Innere Kanten)**.** *Sei σ ein Tetraeder mit den Knoten $x_1, \ldots, x_4$. Weiters seien $x_5, \ldots, x_{10}$ die Kantenmittelpunkte des Tetraeders σ. Diese seien so nummeriert, sodass für $1 \leq i < j \leq 4$ der Knoten x_m mit*

$$m := (i-1)\left(3 - \frac{i}{2}\right) + j + 3$$

der Mittelpunkt der Kante (x_i, x_j) ist. Dann definiert die Menge der Kanten

$$\Psi := \{(x_5, x_{10}), (x_6, x_9), (x_7, x_8)\}$$

die inneren Kanten *des Tetraeders σ.*

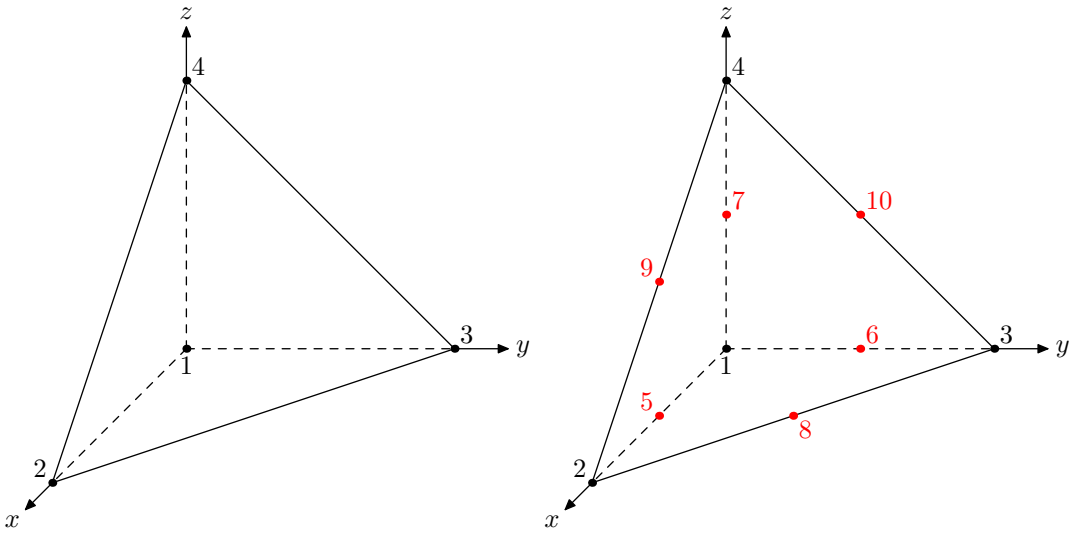


Abbildung 3.1: Tetraeder mit Eckpunkten und Kantenmittelpunkten.

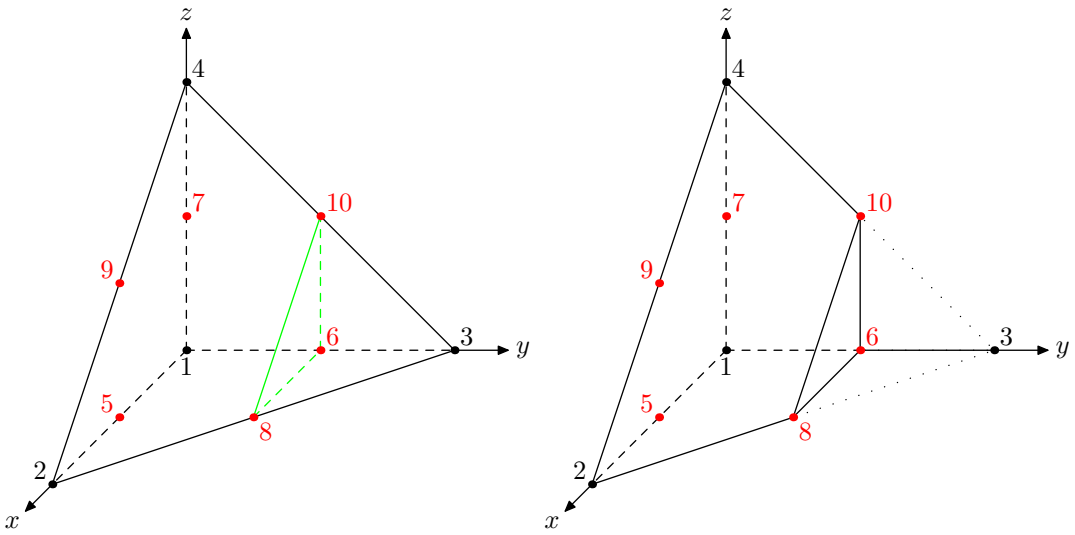


Abbildung 3.2: Entfernen eines Ecktetraeders.

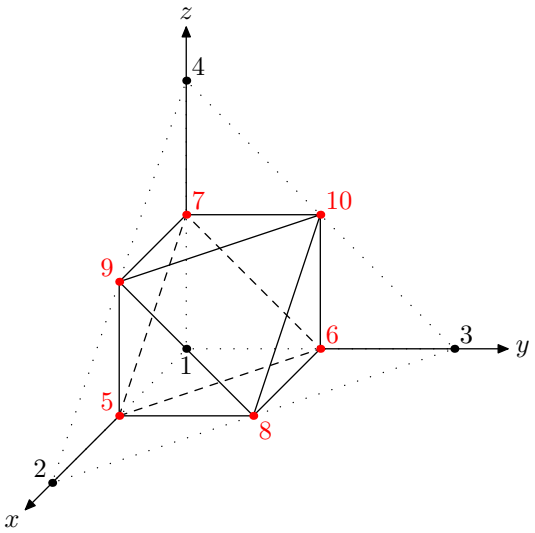


Abbildung 3.3: Tetraeder ohne Ecktetraeder.

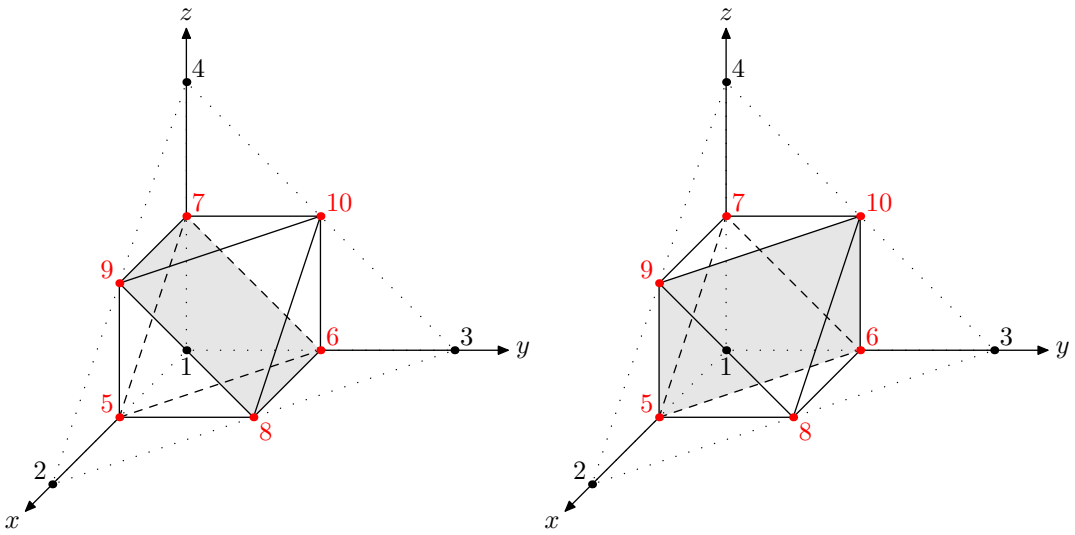


Abbildung 3.4: Zerteilung des Oktaeders I/II.

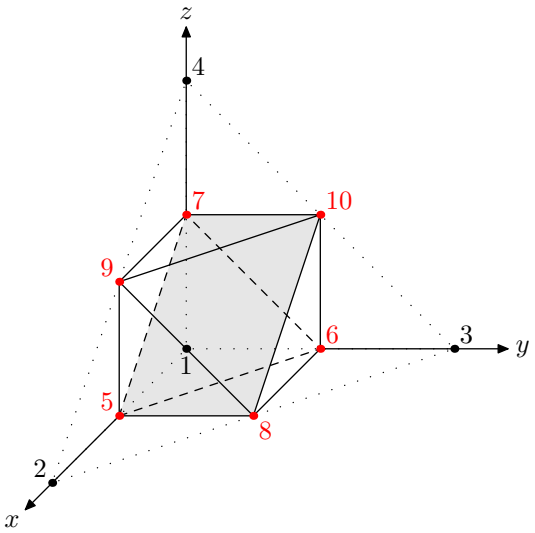


Abbildung 3.5: Zerteilung des Oktaeders III.

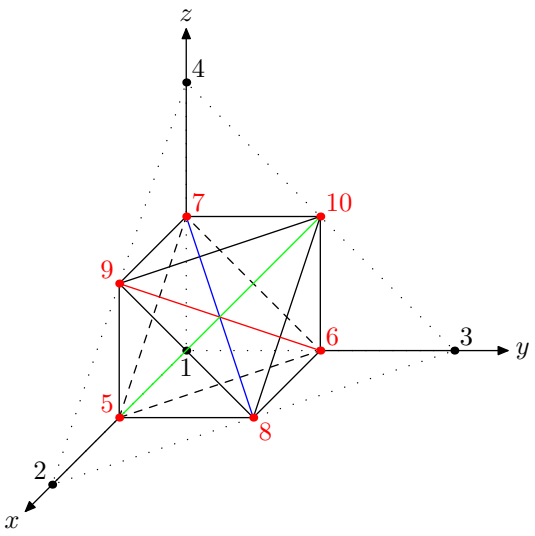


Abbildung 3.6: Inneren Kanten Ψ des Tetraeders σ.

Weiters ist zu beobachten, dass die inneren Kanten Ψ des in Abbildung 3.6 dargestellten Tetraeders σ die längsten Kanten der neuen Zerlegung sind. Im Allgemeinen müssen aber die inneren Kanten Ψ nicht immer gleich groß sein. Somit muss für eine formreguläre Zerlegung des Tetraeders σ jene Kante $e \in \Psi$ fixiert werden, die minimale Länge hat, das heißt

$$e = \arg\min_{e_n \in \Psi} |e_n| .$$

Da die Randelemente eines Pentatops Tetraeder sind, spielen die Kanten Ψ und die Wahl der zu fixierenden Kante $e \in \Psi$ eine wesentliche Rolle.

3.1.2 Zerlegung eines Pentatops

In diesem Abschnitt wird die Zerlegung eines Pentatops in 16 weitere Pentatope behandelt. Dazu wird zuerst die Oberfläche eines Pentatops, welche aus fünf Tetraeder besteht, betrachtet. Die Zerlegung eines Tetraeders ist nicht eindeutig bestimmt. Deshalb wird zuerst die Zerlegung der Oberfläche eines Pentatops eindeutig festgelegt und dann wird das Innere des Pentatops weiter zerlegt. Im folgenden wird die Definition eines Pentatops angegeben und die Transformation auf ein Referenzpentatop untersucht.

Definition 3.5 (Pentatop)**.** *Gegeben seien die Knoten $x_1, \ldots, x_5 \in \mathbb{R}^4$, sodass die Vektoren*

$$\{x_2 - x_1, x_3 - x_1, x_4 - x_1, x_5 - x_1\}$$

linear unabhängig sind. Dann bildet die konvexe Hülle dieser Punkte, also

$$\tau := \text{konv}\left(\{x_1, \ldots, x_5\}\right),$$

ein Pentatop.

Weitere Bezeichnungen für ein Pentatop sind Pentachoron, 5–Zeller, Hyperpyramide oder 4–Simplex. In Abbildung 3.7 sind zwei mögliche Projektionen eines Pentatops dargestellt.

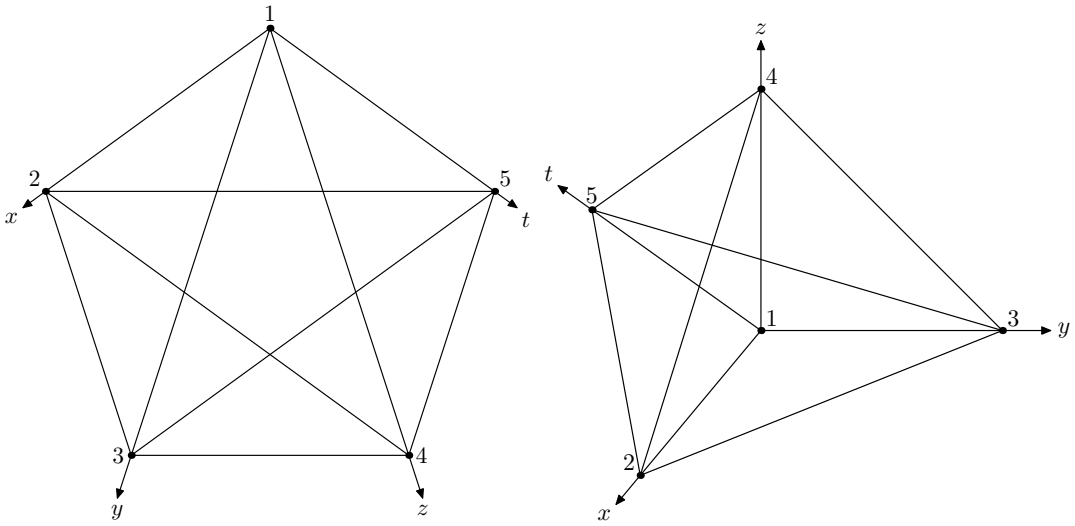


Abbildung 3.7: Zwei mögliche Projektionen eines Pentatops.

Die Anzahl n_R der Randelemente (Tetraeder), die Anzahl n_D der Dreiecke und die Anzahl der Kanten n_K eines Pentatops sind gegeben durch

$$n_R = \binom{5}{4} = \binom{5}{1} = 5,$$
$$n_D = \binom{5}{3} = \binom{5}{2} = 10,$$
$$n_K = \binom{5}{2} = 10.$$

Definition 3.6 (Referenzpentatop)**.** *Das* Referenzpentatop $\hat{\tau}$ *ist gegeben durch*

$$\begin{aligned}\hat{\tau} := \big\{\xi = (\xi_1, \xi_2, \xi_3, \xi_4)^\top \in \mathbb{R}^4 : \quad & 0 \le \xi_1 \le 1, \\ & 0 \le \xi_2 \le 1 - \xi_1, \\ & 0 \le \xi_3 \le 1 - \xi_1 - \xi_2, \\ & 0 \le \xi_4 \le 1 - \xi_1 - \xi_2 - \xi_3\big\}.\end{aligned}$$

Für ein Pentatop $\tau \subset \mathbb{R}^4$ mit den Knoten $\{x_1, \ldots, x_5\} \subset \mathbb{R}^4$ kann mit Hilfe der Transformationsmatrix

$$J_P := \begin{pmatrix} x_{2,1} - x_{1,1} & x_{3,1} - x_{1,1} & x_{4,1} - x_{1,1} & x_{5,1} - x_{1,1} \\ x_{2,2} - x_{1,2} & x_{3,2} - x_{1,2} & x_{4,2} - x_{1,2} & x_{5,2} - x_{1,2} \\ x_{2,3} - x_{1,3} & x_{3,3} - x_{1,3} & x_{4,3} - x_{1,3} & x_{5,3} - x_{1,3} \\ x_{2,4} - x_{1,4} & x_{3,4} - x_{1,4} & x_{4,4} - x_{1,4} & x_{5,4} - x_{1,4} \end{pmatrix} \tag{3.2}$$

jeder Punkt $x \in \tau$ durch

$$x = x_1 + J_P \xi \quad \text{für } \xi \in \hat{\tau} \tag{3.3}$$

dargestellt werden. Die Abbildung (3.3) ist bijektiv, da nach Definition 3.5 die Matrix J_P invertierbar ist. Das Volumen Δ eines Pentatops τ kann somit durch die Transformation auf das Referenzpentatop $\hat{\tau}$ durch

$$\begin{aligned}\Delta = \int_{\tau} \mathrm{d}x = \int_{\hat{\tau}} |\mathrm{det} J_P| \, \mathrm{d}\xi = \\ = |\mathrm{det} J_P| \int_0^1 \int_0^{1-\xi_1} \int_0^{1-\xi_1-\xi_2} \int_0^{1-\xi_1-\xi_2-\xi_3} \mathrm{d}\xi_1 \mathrm{d}\xi_2 \mathrm{d}\xi_3 \mathrm{d}\xi_4 = \frac{1}{24} |\mathrm{det} J_P|\end{aligned}$$

berechnet werden. Entsprechend lässt sich die Determinante der Matrix J_P durch

$$|\mathrm{det} J_P| = 24\Delta$$

ausdrücken.
Im weiteren wird die Zerlegung eines Pentatops behandelt. Dazu seien die Knoten des Pentatops τ von 1 bis 5 nummeriert. Die Nummerierung der Kantenmittelpunte wird dabei von 6 bis 15 festgelegt, siehe Abbildung 3.8. Wie für den Tetraeder werden auch hier zuerst die Pentatope in den Eckpunkten entfernt. Dies ist in Abbildung 3.9 für den Knoten 3 dargestellt. Das resultierende neue Eckpentatop besteht aus den Knotennummern $(7, 10, 3, 13, 14)$. Entfernt man nun alle Eckpentatope, so ergibt sich das in Abbildung 3.10 dargestellte vierdimensionale Objekt, welches noch in $16 - 5 = 11$ weitere Pentatope unterteilt werden muss. Dieses vierdimensionale Objekt wird nun über die Festlegung der Zerlegung der fünf Randtetraeder in die noch fehlenden 11 Pentatope zerteilt. Die Randtetraeder σ_i für $i = 1, \ldots, 5$ des Pentatops τ sind gegeben durch

$$\begin{aligned}\sigma_1 &= \mathrm{konv}\left(\{x_2, x_3, x_5, x_4\}\right), \\ \sigma_2 &= \mathrm{konv}\left(\{x_1, x_3, x_4, x_5\}\right), \\ \sigma_3 &= \mathrm{konv}\left(\{x_1, x_4, x_2, x_5\}\right), \\ \sigma_4 &= \mathrm{konv}\left(\{x_1, x_2, x_3, x_5\}\right), \\ \sigma_5 &= \mathrm{konv}\left(\{x_1, x_2, x_4, x_3\}\right).\end{aligned}$$

Die Zerlegung eines Randtetraeders σ_i ist von der Wahl der inneren Kanten Ψ_i abhängig. Dies motiviert nun die folgende Definition.

Definition 3.7 (Innere Randkanten)**.** *Sei τ ein Pentatop und σ_i seien die Randtetraeder für $i = 1, \ldots, 5$ des Pentatops τ. Dann definiert die Menge aller inneren Kanten Ψ_i der Randtetraeder σ_i*

$$\Phi := \{\Psi_i : \Psi_i \textit{ sind die inneren Kanten von } \sigma_i \textit{ für } i = 1, \ldots, 5\}$$

die inneren Randkanten *des Pentatops τ. Weiters werden mit $\Phi_i := \Psi_i \in \Phi$ die i–ten inneren Kanten des Randtetraeders σ_i bezeichnet.*

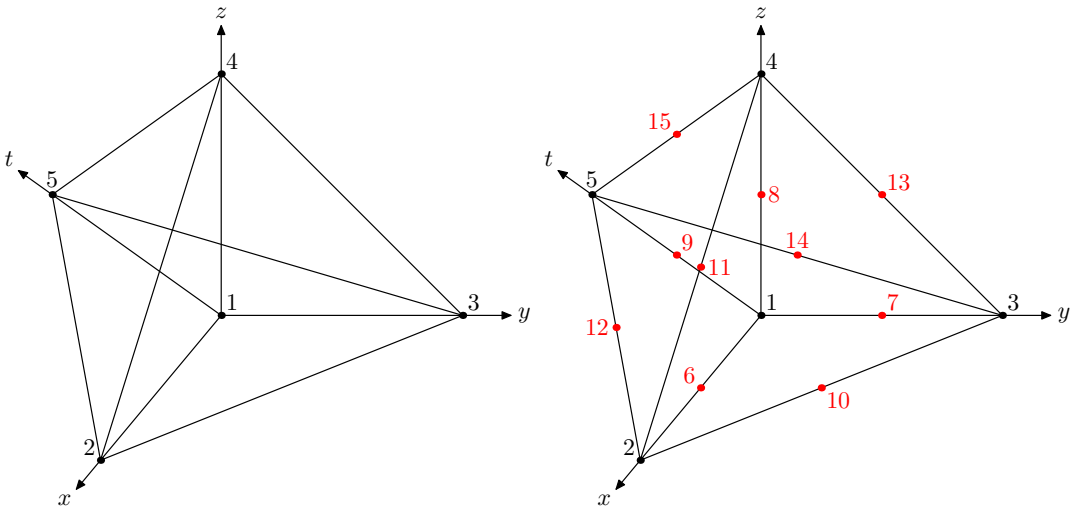


Abbildung 3.8: Pentatop mit Eckpunkten und Kantenmittelpunkten.

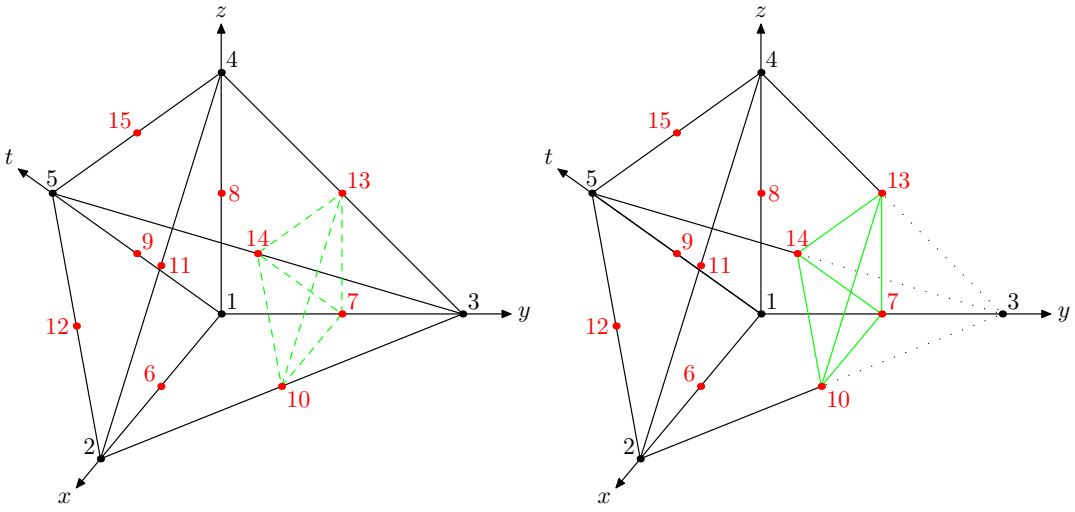


Abbildung 3.9: Entfernen eines Eckpentatops.

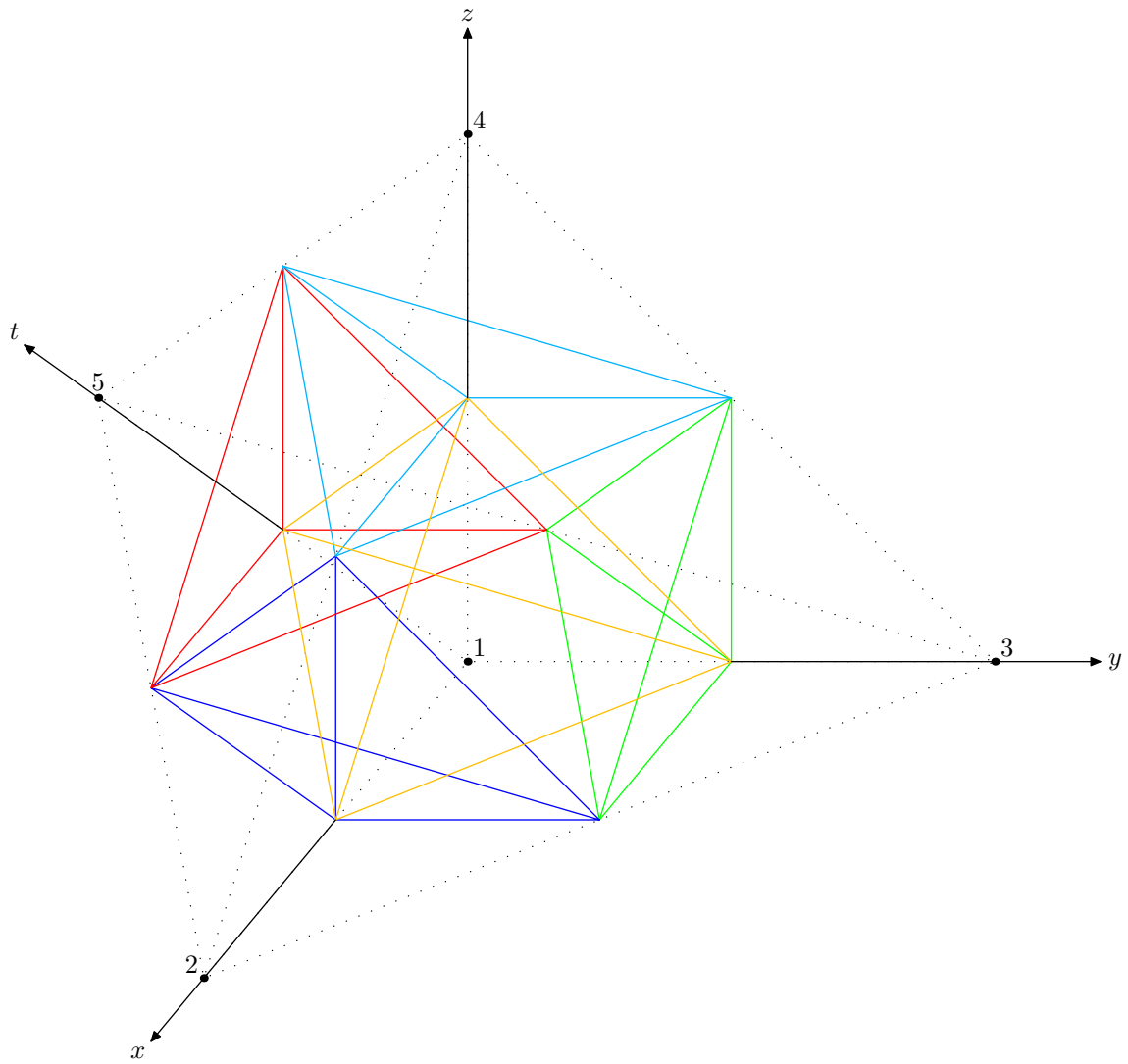


Abbildung 3.10: Pentatop ohne Eckpentatope.

Für das in den Abbildungen 3.8–3.10 dargestellte Pentatop sind die inneren Kanten Ψ_i gegeben durch:

$$\begin{aligned}
\Psi_1 &= \{(x_{10}, x_{15}), (x_{11}, x_{14}), (x_{12}, x_{13})\}, \\
\Psi_2 &= \{(x_7, x_{15}), (x_8, x_{14}), (x_9, x_{13})\}, \\
\Psi_3 &= \{(x_6, x_{15}), (x_8, x_{12}), (x_9, x_{11})\}, \\
\Psi_4 &= \{(x_6, x_{14}), (x_7, x_{12}), (x_9, x_{10})\}, \\
\Psi_5 &= \{(x_6, x_{13}), (x_7, x_{11}), (x_8, x_{10})\}.
\end{aligned}$$

Die inneren Randkanten Φ lauten somit

$$\begin{aligned}
\Phi = \big\{ &\{(x_{10}, x_{15}), (x_{11}, x_{14}), (x_{12}, x_{13})\} \\
&\{(x_7, x_{15}), (x_8, x_{14}), (x_9, x_{13})\}, \\
&\{(x_6, x_{15}), (x_8, x_{12}), (x_9, x_{11})\}, \\
&\{(x_6, x_{14}), (x_7, x_{12}), (x_9, x_{10})\}, \\
&\{(x_6, x_{13}), (x_7, x_{11}), (x_8, x_{10})\} \big\}.
\end{aligned} \tag{3.4}$$

Um die Zerlegung der Oberfläche eines Pentatops eindeutig festzulegen, muss für jeden Randtetraeder σ_i eine innere Randkante $e_i \in \Phi_i$ ausgewählt werden. Daraus ergibt sich die folgende Definition:

Definition 3.8 (Fixierte innere Randkanten)**.** *Sei τ ein Pentatop und Φ die inneren Randkanten von τ und somit Φ_i die inneren Kanten des Randtetraeders σ_i für $i = 1, \ldots, 5$. Wählt man nun eine Kante e_i aus den inneren Randkanten Φ_i aus, so bezeichnet die Menge*

$$\phi := \bigcup_{i=1}^{5} e_i \quad \textit{mit} \quad e_i \in \Phi_i \quad \textit{für alle} \quad i = 1, \ldots, 5$$

die fixierten inneren Randkanten *von τ.*

Definition 3.9. *Sei $A = \{e_1, \ldots, e_n\}$ eine Menge mit n Kanten $e_i = (x_{i,1}, x_{i,2})$. Dann bildet die Abbildung* nodes *die Menge aller Knoten der Menge A, das heißt*

$$\operatorname{nodes}(A) := \bigcup_{e_i = (x_{i,1}, x_{i,2}) \in A} \{x_{i,1}, x_{i,2}\}.$$

Definition 3.10 (Fixierte Knoten)**.** *Sei τ ein Pentatop. Weiters sei ϕ eine Fixierung der inneren Randkanten des Pentatops τ. Dann bezeichnet die Menge aller Knoten dieser Fixierung*

$$x_\phi := \operatorname{nodes}(\phi)$$

die fixierten Knoten *des Pentatops τ.*

Insgesamt gibt es $3^5 = 243$ Möglichkeiten, die inneren Randkanten Φ eines Pentatops τ zu fixieren. Nicht alle Fixierungen führen auf eine zulässige Zerlegung des Pentatops. Es gibt Fixierungen, für die das Pentatop nicht zulässig in 16 weitere Pentatope zerlegt werden kann. In dieser Arbeit werden für eine bestimmte Klasse von Fixierungen Methoden angegeben, die eine zulässige Zerlegung des fixierten Pentatops τ garantieren.
Betrachtet wird nun eine mögliche Fixierung der inneren Randkanten des in den Abbildungen 3.8–3.10 dargestellten Pentatops τ:

$$\phi_1 := \{(x_{10}, x_{15}), (x_9, x_{13}), (x_6, x_{15}), (x_9, x_{10}), (x_6, x_{13})\}.$$

Die fixierten Knoten x_{ϕ_1} der Fixierung ϕ_1 lauten somit:

$$x_{\phi_1} = \{x_6, x_9, x_{10}, x_{13}, x_{15}\}.$$

Betrachtet man die Fixierung ϕ_1 genauer, so ist zu erkennen, dass die Kanten der Fixierung ϕ_1 zusammenhängend sind.

Definition 3.11 (Zusammenhängende Fixierung)**.** *Sei τ ein Pentatop. Weiters sei ϕ eine Fixierung der inneren Randkanten des Pentatops τ mit der Eigenschaft, dass*

$$\textit{für alle Kanten } e_i \in \phi \textit{ eine Kante } e_j \in \phi \textit{ exisitiert, sodass für } i \neq j : e_i \cap e_j \neq \emptyset$$

gilt. Dann heißt die Fixierung ϕ zusammenhängende Fixierung.

Das nächste Lemma gibt eine obere Schranke für die Anzahl der fixierten Knoten einer Fixierung ϕ an.

Lemma 3.12. *Sei τ ein Pentatop und $\phi = \{e_1, \ldots, e_5\}$ eine zusammenhängende Fixierung von τ. Dann ist die Anzahl der fixierten Knoten x_ϕ nach oben beschränkt mit*

$$|x_\phi| \leq 6.$$

Beweis. Sei $B_1 := \{e_1\}$ und $B_i := B_{i-1} \cup \{e_k\}$ mit $e_k \in \phi$, sodass für $i = 2, \ldots, 5$ die Eigenschaften

$$e_k \neq e_n \quad \text{für alle} \quad e_n \in B_{i-1} \quad \text{und es existiert ein} \quad e_n \in B_{i-1} : e_k \cap e_n \neq \emptyset$$

erfüllt sind. Die Mengen B_i sind wohldefiniert, da die Fixierung ϕ zusammenhängend ist. Da weiters die neue Kante $e_k \in B_i = B_{i-1} \cup \{e_k\}$ per Definition mit einer Kante aus B_{i-1} verbunden ist und $|B_1| = 2$ ist, gilt per Induktion

$$|\text{nodes}(B_i)| = |\text{nodes}(B_{i-1} \cup \{e_k\})| \leq |\text{nodes}(B_{i-1})| + 1 \leq i + 1 \quad \text{für alle} \quad i = 1, \ldots, 5.$$

Somit ergibt sich die Behauptung mit

$$|x_\phi| = |\text{nodes}(\phi)| = |\text{nodes}(B_5)| \leq 5 + 1 = 6.$$

□

Betrachtet man weiters die inneren Randkanten (3.4), so ist zu erkennen, dass jede mögliche Fixierung ϕ mindestens fünf fixierte Knoten besitzt, das heißt

$$|x_\phi| \geq 5.$$

Mit Hilfe von Lemma 3.12 ergibt sich somit für eine zusammenhängende Fixierung ϕ die Abschätzung

$$5 \leq |x_\phi| \leq 6.$$

Dies erlaubt es nun, die zusammenhängenden Fixierungen in zwei Klassen einzuteilen:

Definition 3.13 (Zyklische und azyklische Fixierung)**.** *Sei τ ein Pentatop und ϕ eine zusammenhängende Fixierung von τ. Die Fixierung ϕ heißt* zyklische Fixierung*, falls die Anzahl der fixierten Knoten x_ϕ durch*

$$|x_\phi| = 5$$

gegeben ist. Gilt anderenfalls $|x_\phi| = 6$, so wird die Fixierung ϕ als azyklische Fixierung *bezeichnet.*

Von den 243 möglichen Fixierungen gibt es 12 zyklische Fixierungen und 75 azyklische Fixierungen, siehe Anhang. Die obige betrachtete Fixierung

$$\phi_1 = \{(x_{10}, x_{15}), (x_9, x_{13}), (x_6, x_{15}), (x_9, x_{10}), (x_6, x_{13})\}$$

ist eine zyklische Fixierung. Eine azyklische Fixierung wäre zum Beispiel

$$\phi_2 = \{(x_{10}, x_{15}), (x_9, x_{13}), (x_9, x_{11}), (x_9, x_{10}), (x_8, x_{10})\}.$$

Für zyklische Fixierungen ist die Konstruktion der Zerlegung eines Pentatops in der nächsten Definition angegeben:

Definition 3.14 (Zyklische Zerlegung)**.** *Sei τ ein Pentatop mit den Knoten $\{x_1, \ldots, x_5\}$, $\phi = \{e_1, \ldots, e_5\}$ eine zyklische Fixierung und Φ die inneren Randkanten von τ. Weiters seien die Kantenmittelpunkte von τ gegeben durch $\{x_6, \ldots, x_{15}\}$, sodass für $1 \leq i < j \leq 5$ der Knoten x_m mit*

$$m := (i-1)\left(4 - \frac{i}{2}\right) + j + 4$$

der Mittelpunkt der Kante (x_i, x_j) ist. Die Mengen χ_i sind für $i = 1, \ldots, 6$ definiert als

$$\begin{aligned}
\chi_1 &:= \{x_9, x_1, x_6, x_7, x_8\}, \\
\chi_2 &:= \{x_{11}, x_{12}, x_6, x_2, x_{10}\}, \\
\chi_3 &:= \{x_3, x_{13}, x_{14}, x_7, x_{10}\}, \\
\chi_4 &:= \{x_{11}, x_{13}, x_4, x_{15}, x_8\}, \\
\chi_5 &:= \{x_9, x_{12}, x_{14}, x_{15}, x_5\}, \\
\chi_6 &:= \text{nodes}(\phi) = x_\phi.
\end{aligned}$$

Der Index $\ell_{i,j}$ *sei für* $1 \leq i < j \leq 5$ *gegeben durch*

$$\ell_{i,j} := (i-1)\left(4 - \frac{i}{2}\right) + j + 5.$$

Somit gibt es $\binom{5}{2} = 10$ *verschiedene Indizes* $\ell_{i,j}$*. Weiters seien die Mengen* $\chi_{\ell_{i,j}}$ *definiert als*

$$\chi_{\ell_{i,j}} := \operatorname{nodes}\left(\{\mathrm{e_i}\} \cup \{\mathrm{e_j}\}\right) \cup \left[\operatorname{nodes}(\Phi_i) \cap \operatorname{nodes}(\Phi_j)\right] \quad \textit{für alle} \quad 1 \leq i < j \leq 5.$$

Dann wird die Menge der Pentatope

$$\mathcal{T}_{\text{zykl}} := \{\tau_i = \operatorname{konv}(\chi_i) : i = 1, \ldots, 16\}$$

als zyklische Zerlegung *des Pentatops* τ *bezeichnet.*

Die ersten fünf Pentatope der zyklischen Zerlegung $\mathcal{T}_{\text{zykl}}$ sind genau die fünf Eckpentatope. Das sechste Pentatop ergibt sich genau aus den Knoten x_ϕ der zyklischen Fixierung. Die restlichen zehn Pentatope der zyklischen Zerlegung $\mathcal{T}_{\text{zykl}}$ ergeben sich aus der Tatsache, dass jeweils genau eine fixierte innere Randkante in vier neuen Randtetraedern enthalten sein muss, siehe dazu Abbildung 3.6. Die zyklische Zerlegung eines Pentatops kann auch mit dem Algorithmus von Freudenthal bestimmt werden, siehe [6].
Für die zyklische Fixierung

$$\phi_1 = \{(x_{10}, x_{15}), (x_9, x_{13}), (x_6, x_{15}), (x_9, x_{10}), (x_6, x_{13})\}$$

ergibt sich nach Definition 3.14 folgende Zerlegung:

$$\begin{aligned}
\chi_1 &:= \{x_1, x_6, x_7, x_8, x_9\}, & \chi_2 &:= \{x_2, x_6, x_{10}, x_{11}, x_{12}\},\\
\chi_3 &:= \{x_3, x_7, x_{10}, x_{13}, x_{14}\}, & \chi_4 &:= \{x_4, x_8, x_{11}, x_{13}, x_{15}\},\\
\chi_5 &:= \{x_5, x_9, x_{12}, x_{14}, x_{15}\}, & \chi_6 &:= \{x_6, x_9, x_{10}, x_{13}, x_{15}\},\\
\chi_7 &:= \{x_6, x_8, x_9, x_{13}, x_{15}\}, & \chi_8 &:= \{x_7, x_9, x_{10}, x_{13}, x_{14}\},\\
\chi_9 &:= \{x_6, x_7, x_8, x_9, x_{13}\}, & \chi_{10} &:= \{x_9, x_{10}, x_{13}, x_{14}, x_{15}\},\\
\chi_{11} &:= \{x_6, x_9, x_{10}, x_{12}, x_{15}\}, & \chi_{12} &:= \{x_6, x_8, x_{11}, x_{13}, x_{15}\},\\
\chi_{13} &:= \{x_6, x_{10}, x_{11}, x_{12}, x_{15}\}, & \chi_{14} &:= \{x_6, x_7, x_9, x_{10}, x_{13}\},\\
\chi_{15} &:= \{x_9, x_{10}, x_{12}, x_{14}, x_{15}\}, & \chi_{16} &:= \{x_6, x_{10}, x_{11}, x_{13}, x_{15}\}.
\end{aligned}$$

Diese Zerlegung ist in Abbildung 3.11 als Graph dargestellt. Dabei stellen die Kreise die neuen Pentatope mit der jeweiligen Elementnummer dar. Die Kanten des Graphs verbinden jeweils zwei Pentatope miteinander und stellen daher den inneren Tetraeder zwischen diesen beiden Pentatopen dar. Die Pfeile mit den jeweiligen Nummern zeigen an, welche Knoten zwischen zwei benachbarten Pentatopen ausgetauscht werden müssen, um das jeweilige Nachbarpentatop zu erhalten. Somit lassen sich die Knotennummern des inneren

Tetraeders durch das Weglassen dieser Pfeilnummern vom jeweiligen Pentatop bestimmen. Weiters ist zu erkennen, dass das Element τ_6, welches genau aus den fixierten Knoten x_{ϕ_1} besteht, keinen Randtetraeder besitzt. Die Pentatope $\tau_1, \ldots, \tau_5$, welche die Eckpentatope sind, haben vier Randtetraeder und einen inneren Tetraeder. Die restlichen Pentatope $\tau_7, \ldots, \tau_{16}$ besitzen genau zwei Randtetraeder und drei innere Tetraeder.
Aus dem in Abbildung 3.11 dargestellten Graph lassen sich somit alle neuen Pentatope, die neuen inneren Tetraeder und die neuen Randtetraeder eindeutig bestimmen.

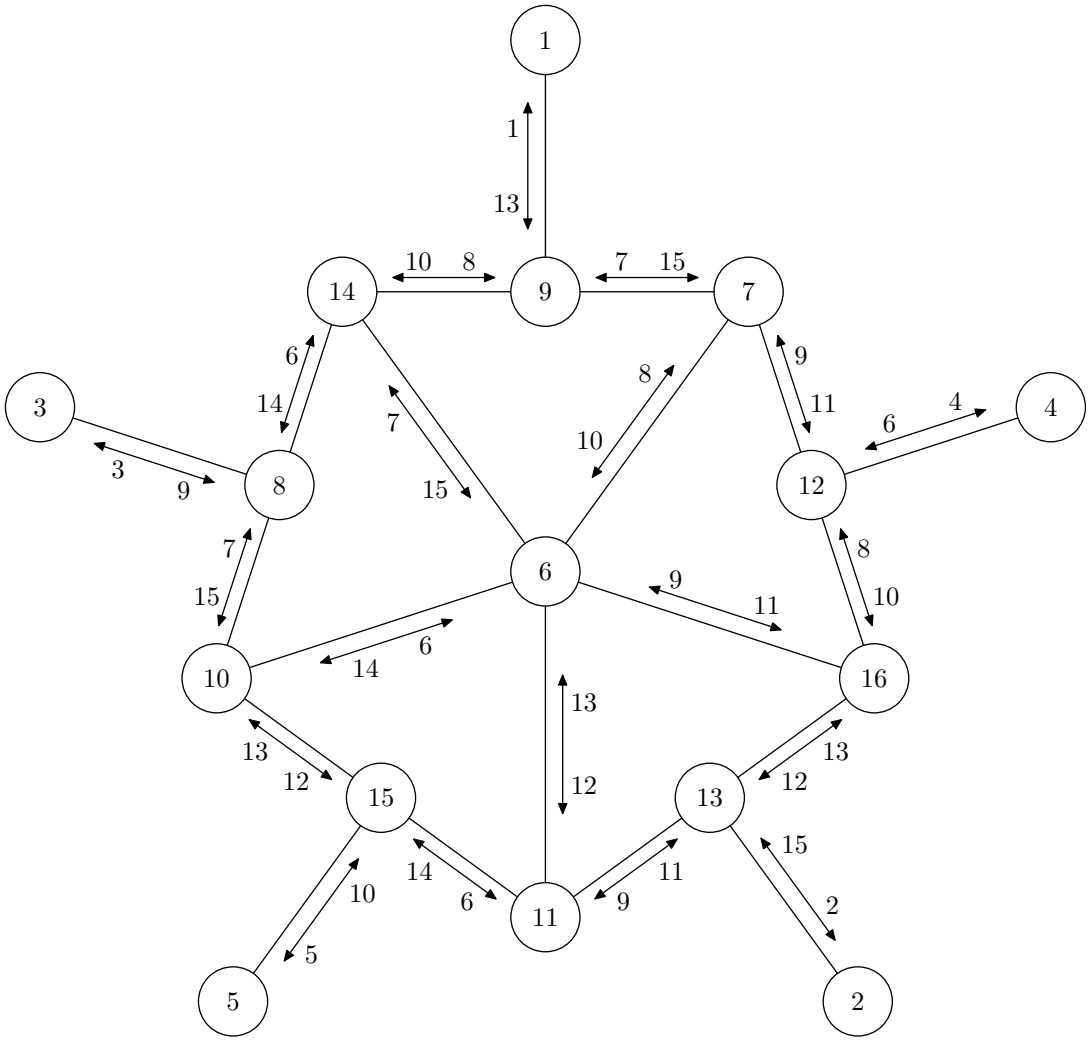


Abbildung 3.11: Zerlegung eines Pentatops mit zyklischer Fixierung.

In ähnlicher Weise lässt sich ein Pentatop τ mit einer azyklischen Fixierung ϕ zerlegen. Dies wird in der nächsten Definition behandelt:

Definition 3.15 (Azyklische Zerlegung)**.** *Sei τ ein Pentatop mit den Knoten $\{x_1, \ldots, x_5\}$, $\phi = \{e_1, \ldots, e_5\}$ eine azyklische Fixierung und Φ die inneren Randkanten von τ. Weiters*

seien die Kantenmittelpunkte $\{x_6, \ldots, x_{15}\}$ *von* τ *wie in Definition 3.11 sortiert. Im folgenden werden nun zwei bestimmte Kanten* e_{n_1} *und* e_{n_2} *aus der Fixierung* ϕ *mit bestimmten Eigenschaften ausgewählt. Dazu sei*

$$M := \{e_i \in \phi : \text{ es existiert } x \in e_i : x \notin e_j \text{ für alle } j \neq i\}.$$

- *Falls* $|M| = 3$ *ist, wähle die Kanten* $e_{n_1}, e_{n_2} \in M = \{e_{n_1}, e_{n_2}, e_{n_3}\}$ *so, dass für die Kante* e_{n_3} *die Eigenschaft*

  $$\text{nodes}(\Phi_{n_1}) \cap \text{nodes}(\Phi_{n_2}) \cap \text{nodes}(e_{n_3}) \neq \emptyset$$

  erfüllt ist.

- *Falls* $|M| = 4$ *ist, wähle die Kanten* $e_{n_1}, e_{n_2} \in M$ *wobei* $e_{n_1} \neq e_{n_2}$ *so, dass diese nichtleeren Durchschnitt haben, das heißt*

  $$e_{n_1} \cap e_{n_2} \neq \emptyset.$$

Für $i = 1, \ldots, 5$ *sind die Mengen* χ_i *definiert als*

$$\begin{aligned}
\chi_1 &:= \{x_9, x_1, x_6, x_7, x_8\}, \\
\chi_2 &:= \{x_{11}, x_{12}, x_6, x_2, x_{10}\}, \\
\chi_3 &:= \{x_3, x_{13}, x_{14}, x_7, x_{10}\}, \\
\chi_4 &:= \{x_{11}, x_{13}, x_4, x_{15}, x_8\}, \\
\chi_5 &:= \{x_9, x_{12}, x_{14}, x_{15}, x_5\}.
\end{aligned}$$

Jedem Tupel $(i, j) \neq (n_1, n_2)$ *mit* $1 \leq i < j \leq 5$ *sei ein Index* $\ell_{i,j} \in \{6, \ldots, 14\}$ *isomorph zugeordnet. Dann sind die Mengen* $\chi_{\ell_{i,j}}$ *für alle* $\ell_{i,j} \in \{6, \ldots, 14\}$ *definiert als*

$$\chi_{\ell_{i,j}} := \text{nodes}\left(\{\mathrm{e_i}\} \cup \{\mathrm{e_j}\}\right) \cup \left[\text{nodes}(\Phi_i) \cap \text{nodes}(\Phi_j)\right].$$

Weiters sind

$$\begin{aligned}
\chi_{15} &:= \text{nodes}(\phi \setminus \{e_{n_1}\}), \\
\chi_{16} &:= \text{nodes}(\phi \setminus \{e_{n_2}\}).
\end{aligned}$$

Dann wird die Menge der Pentatope

$$\mathcal{T}_{\text{azykl}} := \{\tau_i = \text{konv}(\chi_i) : i = 1, \ldots, 16\}$$

als azyklische Zerlegung *des Pentatops* τ *bezeichnet.*

Für alle möglichen 75 azyklischen Fixierungen liefert die in Definition 3.15 angegebene Vorschrift zulässige Zerlegungen in 16 gleich große Pentatope. Für die azyklische Fixierung

$$\phi_2 = \{(x_{10}, x_{15}), (x_9, x_{13}), (x_9, x_{11}), (x_9, x_{10}), (x_8, x_{10})\}$$

ergibt sich die azyklische Zerlegung

$$\begin{aligned}
\chi_1 &:= \{x_1, x_6, x_7, x_8, x_9\}, & \chi_2 &:= \{x_2, x_6, x_{10}, x_{11}, x_{12}\}, \\
\chi_3 &:= \{x_3, x_7, x_{10}, x_{13}, x_{14}\}, & \chi_4 &:= \{x_4, x_8, x_{11}, x_{13}, x_{15}\}, \\
\chi_5 &:= \{x_5, x_9, x_{12}, x_{14}, x_{15}\}, & \chi_6 &:= \{x_8, x_9, x_{11}, x_{13}, x_{15}\}, \\
\chi_7 &:= \{x_7, x_9, x_{10}, x_{13}, x_{14}\}, & \chi_8 &:= \{x_7, x_8, x_9, x_{10}, x_{13}\}, \\
\chi_9 &:= \{x_9, x_{10}, x_{13}, x_{14}, x_{15}\}, & \chi_{10} &:= \{x_6, x_9, x_{10}, x_{11}, x_{12}\}, \\
\chi_{11} &:= \{x_6, x_8, x_9, x_{10}, x_{11}\}, & \chi_{12} &:= \{x_9, x_{10}, x_{11}, x_{12}, x_{15}\}, \\
\chi_{13} &:= \{x_6, x_7, x_8, x_9, x_{10}\}, & \chi_{14} &:= \{x_9, x_{10}, x_{12}, x_{14}, x_{15}\}, \\
\chi_{15} &:= \{x_9, x_{10}, x_{11}, x_{13}, x_{15}\}, & \chi_{16} &:= \{x_8, x_9, x_{10}, x_{11}, x_{13}\}.
\end{aligned}$$

Der in Abbildung 3.12 dargestellte Graph spiegelt genau diese Zerlegung wider. Dabei ist zu erkennen, dass die Pentatope τ_{15} und τ_{16} vier innere Tetraeder und einen Randtetraeder besitzen. Die Pentatope $\tau_6, \ldots, \tau_{14}$ haben genau zwei Randtetraeder und drei innere Tetraeder. Auch hier kann mithilfe des in Abbildung 3.12 dargestellten Graphs die Zerlegung eindeutig bestimmt werden.

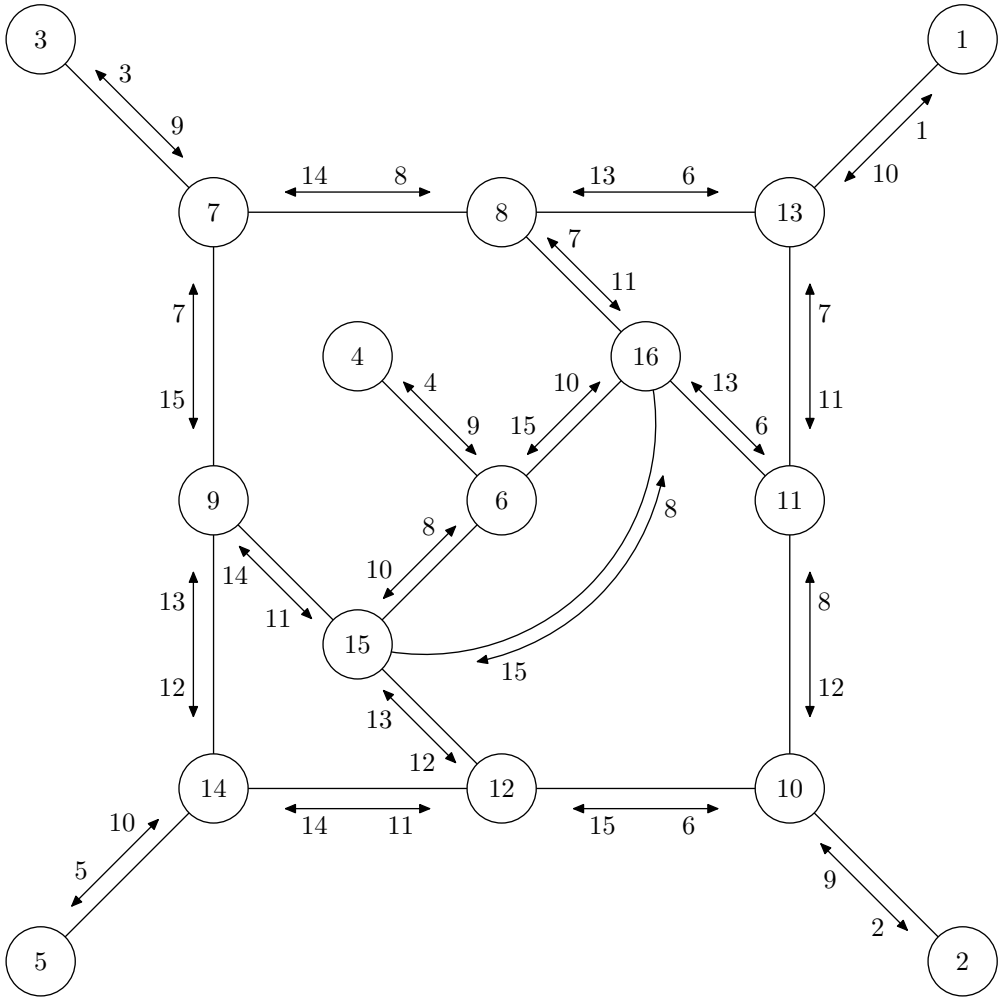


Abbildung 3.12: Zerlegung eines Pentatops mit azyklischer Fixierung.

3.1.3 Zerlegung eines Hyperwürfels

In diesem Abschnitt wird die Zerlegung eines vierdimensionalen Würfels beziehungsweise eines Hyperwürfels in 96 gleich große Pentatope angegeben. In [13, 22] werden Zerlegungen des Hyperwürfels in $N \leq 4! = 24$ gleich große Pentatope behandelt.

Definition 3.16 (n–dimensionaler Würfel)**.** *Sei $a \in \mathbb{R}^+$ und $n \in \mathbb{N}$, dann ist*

$$W_n := [0, a]^n$$

*ein n–*dimensionalen Würfel *mit der Kantenlänge a. Der Index n ist dabei die Dimension des Würfls W_n.*

In [4] wird die Zerlegung des n–dimensionalen Würfels behandelt. Hier wird anschaulich eine Zerlegung des Hyperwürfels $W_4 \subset \mathbb{R}^4$ vorgestellt.

Lemma 3.17. *Sei $n \in \mathbb{N}$. Der n–dimensionale Würfel W_n besitzt genau*

$$N_{W_{n-1}} = 2n$$

$n-1$–dimensionale Würfel als Oberfläche.

Beweis. Siehe zum Beispiel [10]. □

Ein Hyperwürfel W_4 besitzt somit $N_{W_3} = 8$ Würfel W_3^i mit $i = 1, \ldots, 8$ als Oberfläche. Eine mögliche Projektion eines Hyperwürfels ist in Abbildung 3.13 angegeben, in der auch die acht Würfel W_3^i zu sehen sind. Weiters besteht laut Lemma 3.17 ein Würfel W_3^i aus $N_{W_2} = 6$ Quadraten $W_2^{i,j}$ für $i = 1, \ldots, 8$ und $j = 1, \ldots, 6$. Ausgehend vom Mittelpunkt $M = \left(\frac{a}{2}, \frac{a}{2}, \frac{a}{2}, \frac{a}{2}\right)^\top$ des Hyperwürfels W_4 lässt sich dieser nun in 96 Pentatope τ_k zerlegen: Dazu werden die Mittelpunkte M_i für jeden Würfel W_3^i für $i = 1, \ldots, 8$ betrachtet. Jedes Quadrat $W_2^{i,j}$ lässt sich nun in zwei Dreiecke zerteilen. Jedes Dreieck bildet mit dem Mittelpunkt M_i des Würfels W_3^i und mit dem Mittelpunkt M des Hyperwürfels W_4 ein Pentatop τ_k. In Abbildung 3.13 ist dies für ein Pentatop veranschaulicht. Die resultierende Zerlegung besteht somit aus

$$N = 2N_{W_2}N_{W_3} = 96$$

gleich großen Pentatopen τ_k.

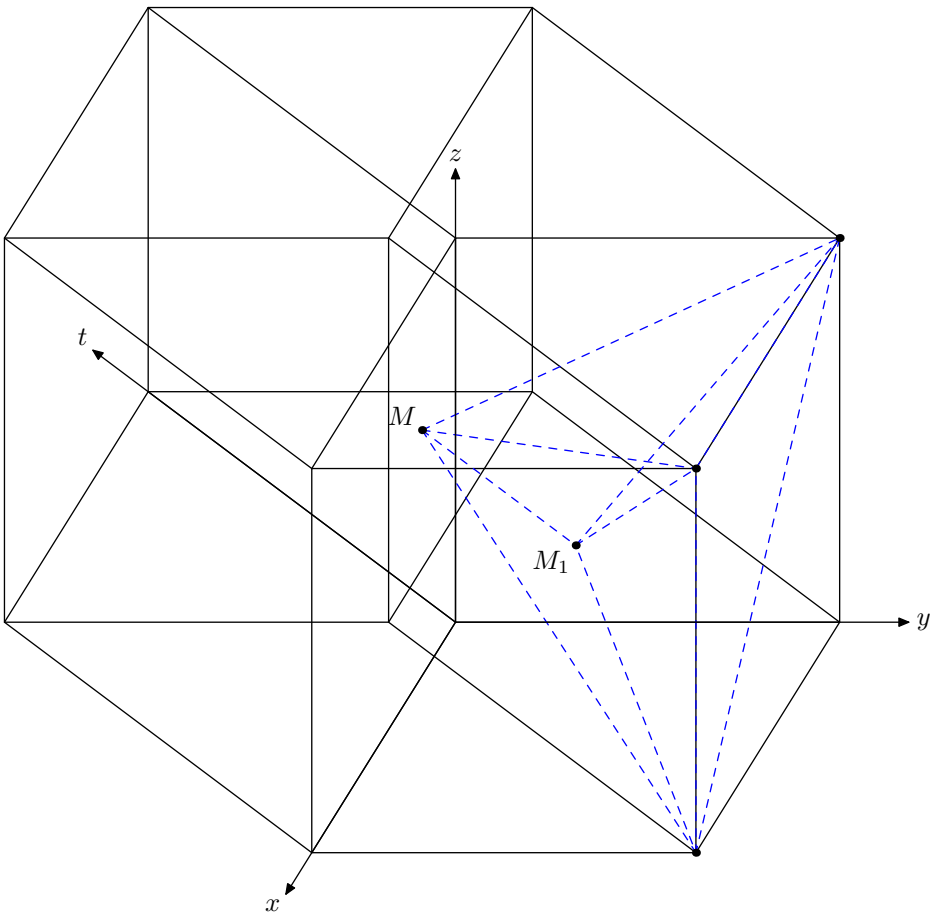


Abbildung 3.13: Das Zerlegen eines Hyperwürfels.

3.2 Uniforme Verfeinerung

In diesem Abschnitt wird die uniforme Verfeinerung einer gegebenen Triangulierung $\mathcal{T}_N$ in Pentatope $\tau_k \in \mathcal{T}_N$ behandelt. In Abschnitt 3.1.2 wurde die Zerlegung eines Pentatops untersucht. Dazu musste für ein Pentatop τ_k eine zyklische oder eine azyklische Fixierung ϕ_k gewählt werden. Aus der Information über die Fixierung ϕ_k kann dann das Pentatop τ_k in 16 weitere Pentatope zerlegt werden. Damit für zwei benachbarte Pentatope $\tau_k, \tau_\ell \in \mathcal{T}_N$ die weitere Zerlegung zulässig bleibt, muss die gemeinsame fixierte Kante für beide Pentatope übereinstimmen. Denn dann stimmt die Zerlegung des gemeinsamen Randtetraeders der Pentatope τ_k, τ_ℓ für beide Pentatope überein. Dies motiviert die folgende Definition:

Definition 3.18 (Zulässig fixiert)**.** *Sei $\mathcal{T}_N$ eine zulässige Triangulierung. Zwei Pentatope $\tau_k, \tau_\ell \in \mathcal{T}_N$ heißen* zulässig fixiert*, falls für die fixierte Kante $e_{k,\ell}$ des gemeinsamen Tetraeders $\sigma_{k,\ell}$ also $\sigma_{k,\ell} = \tau_k \cap \tau_\ell$ die Bedingung*

$$e_{k,\ell} \in \phi_k \quad \textit{und} \quad e_{k,\ell} \in \phi_\ell$$

gilt. Dabei sind ϕ_k und ϕ_ℓ zyklische beziehungsweise azyklische Fixierungen der Pentatope τ_k und τ_ℓ.

Wenn alle benachbarten Pentatope $\tau_k, \tau_\ell \in \mathcal{T}_N$ zulässig fixiert sind, führen die in den Definitionen 3.14 und 3.15 angegebenen Zerlegungen für die Pentatope $\tau_k \in \mathcal{T}_N$ auf eine zulässige feinere Zerlegung $\mathcal{T}_{16N}$. Ein möglicher Algorithmus, um alle benachbarten Pentatope zulässig zu fixieren, ist in Algorithmus 3.14 angegeben. Dabei werden für die fixierten Kanten immer die mit minimaler Länge ausgewählt.
In den Tabellen 3.1–3.3 sind die wichtigsten Merkmale einer uniformen Zerlegung eines einzelnen Pentatops und eines Hyperwürfels angegeben. Dabei wurde für den Hyperwürfel eine Ausgangszerlegung mit $N = 96$ Pentatopen (siehe Abschnitt 3.1.3) und $N = 24$ Pentatopen verwendet.
Falls es möglich ist, die Ausgangszerlegung nur mit zyklischen Fixierungen festzulegen, so kann man durch rekursives Anwenden der Zerlegungen auf die Söhne wieder zulässige Zerlegungen erhalten. Dazu müssen in der Ausgangszerlegung die Knoten der Pentatope wie in Definition 3.19 angeordnet sein. Dies entspricht dann der globalen Anwendung des Algorithmus von Freudenthal, siehe [6].

Definition 3.19 (Zulässige Nummerierung)**.** *Eine Zerlegung $\mathcal{T}_N$ heißt* zulässig nummeriert*, falls für zwei benachbarte Pentatope*

$$\tau_k = \operatorname{konv}(\{x_1, \ldots, x_5\}), \quad \tau_\ell = \operatorname{konv}(\{y_1, \ldots, y_5\}) \in \mathcal{T}_N$$

Indizes $i_1 < i_2 < i_3 < i_4 < i_5$ und $j_1 < j_2 < j_3 < j_4 < j_5$ existieren, sodass

$$\{x_1, \ldots, x_5\} \cap \{y_1, \ldots, y_5\} \cong \{x_{i_1}, x_{i_2}, x_{i_3}, x_{i_4}, x_{i_5}\} \equiv \{x_{j_1}, x_{j_2}, x_{j_3}, x_{j_4}, x_{j_5}\}$$

gilt. Dabei bedeutet $\cong$, dass die zwei Mengen isomorph sind und $\equiv$ bedeutet, dass die Mengen gleich sind und zusätzlich die Anordnung der Elemente identisch ist.

Für ein einzelnes Pentatop und für die Zerlegung des Hyperwürfels in $N = 24$ Elementen ist es möglich, die Anfangszerlegung geeignet zu nummerieren. Die resultierenden feineren Zerlegungen mit rein zyklischen Fixierungen sind in den Tabellen 3.4–3.5 angegeben. Dabei ist zu erkennen, dass sich der maximale Durchmesser $\max_{k=1,\dots,N} d_k$ der Pentatope nach dem zweitem Level immer halbiert. Dies ist in den Tabllen 3.1 und 3.3, welche die Triangulierungen mit zyklischen und azyklischen Fixierungen beinhalten, nicht immer der Fall. Jedoch ist der maximale Durchmesser der Pentatope mit einer rein zyklschen Fixierung für den Hyperwürfel größer als mit zyklischen und azyklischen Fixierungen, siehe Tabelle 3.3 und 3.5.

For $k = 1, 2, \dots, N$:
  Betrachte alle Tetraeder $\sigma_{k,i}$ der Elemente $\tau_k \in \mathcal{T}_N$.
  For $i = 1, \dots, 5$:
    Bestimme die inneren Kanten $\Psi_{k,i}$.
    Berechne die minimale innere Kante von $\sigma_{k,i}$

$$e_{k,i} := \arg \min_{e_n \in \Psi_{k,i}} |e_n| .$$

    Bestimme weiters die Menge

$$A_{k,i} := \{e_n \in \Psi_{k,i} : |e_n| = |e_{k,i}|\}.$$

    Falls $|A_{k,i}| = 1$ ist, fixiere die Kante $e_{k,i}$ des Tetraeders $\sigma_{k,i}$.
  EndFor
EndFor
For $k = 1, 2, \dots, N$:

  For $i = 1, \dots, 5$:
    Falls der Tetraeder $\sigma_{k,i}$ des Elements $\tau_k \in \mathcal{T}_N$ noch nicht fixiert ist, fixiere eine geeignete Kante $e \in A_{k,i}$.
  EndFor
EndFor
For $k = 1, 2, \dots, N$:

  Falls die Fixierung ϕ_k des Tetraeders τ_k nicht zusammenhängend ist, tausche geeignete Kanten der Fixierung ϕ_k aus.
EndFor

Algorithmus 3.14: Fixierung aller inneren Randkanten Φ_k einer Triangulierung $\mathcal{T}_N$.

Level	Elemente	Randelemente	innere Elemente	Knoten	$h_{\max}$	$\max_{k=1,\dots,N} d_k$
1	1	5	0	5	0.4518	1.4142
2	16	40	20	15	0.2259	1.0000
3	256	320	480	70	0.1130	0.5590
4	4096	2560	8960	495	0.05648	0.2795
5	65536	20480	153600	4845	0.02824	0.1531
6	1048576	163840	2539520	58905	0.01412	0.07655

Tabelle 3.1: Triangulierungen eines Pentatops.

Level	Elemente	Randelemente	innere Elemente	Knoten	$h_{\max}$	$\max_{k=1,\dots,N} d_k$
1	96	96	192	25	0.3195	1.4142
2	1536	768	3456	169	0.1597	0.7071
3	24576	6144	58368	1681	0.07987	0.3750
4	393216	49152	958464	21025	0.03994	0.1875
5	6291456	393216	15532032	297025	0.01997	0.1083

Tabelle 3.2: Triangulierungen eines Hyperwürfels.

Level	Elemente	Randelemente	innere Elemente	Knoten	$h_{\max}$	$\max_{k=1,\dots,N} d_k$
1	24	48	36	16	0.4518	2.0000
2	384	384	768	81	0.2259	1.1180
3	6144	3072	13824	625	0.1130	0.5590
4	98304	24576	233472	6561	0.05648	0.3062
5	1572864	196608	3833856	83521	0.02824	0.1531

Tabelle 3.3: Triangulierungen eines Hyperwürfels.

Level	Elemente	Randelemente	innere Elemente	Knoten	$h_{\max}$	$\max_{k=1,\dots,N} d_k$
1	1	5	0	5	0.4518	1.4142
2	16	40	20	15	0.2259	1.0000
3	256	320	480	70	0.1130	0.5000
4	4096	2560	8960	495	0.05648	0.2500
5	65536	20480	153600	4845	0.02824	0.1250
6	1048576	163840	2539520	58905	0.01412	0.0625

Tabelle 3.4: Zyklische Triangulierungen eines Pentatops.

Level	Elemente	Randelemente	innere Elemente	Knoten	$h_{\max}$	$\max\limits_{k=1,\dots,N} d_k$
1	24	48	36	16	0.4518	2.0000
2	384	384	768	81	0.2259	1.5811
3	6144	3072	13824	625	0.1130	0.7906
4	98304	24576	233472	6561	0.05648	0.3953
5	1572864	196608	3833856	83521	0.02824	0.1976

Tabelle 3.5: Zyklische Triangulierungen eines Hyperwürfels.

3.3 Visualisierung

Hier wird eine Möglichkeit behandelt, eine Triangulierung $\mathcal{T}_N$ zu visualisieren. Dazu wird diese in endlich viele dreidimensionale Mannigfaltigkeiten zerschnitten. Um dies zu erreichen, wird eine Hyperebene benötigt:

Definition 3.20 (Hyperebene)**.** *Sei $P_0 \in \mathbb{R}^4$ ein beliebiger Vektor und $P_1, P_2, P_3 \in \mathbb{R}^4$ seien linear unabhängige Vektoren. Dann beschreibt die Menge*

$$E_4 := \{x : x = P_0 + \mu_1 P_1 + \mu_2 P_2 + \mu_3 P_3 \quad \textit{für} \quad \mu_1, \mu_2, \mu_3 \in \mathbb{R}\}$$

eine Hyperebene.

Die Triangulierung $\mathcal{T}_N$ wird mit einer Hyperebene E_4 geschnitten, indem jedes einzelne Pentatop $\tau_k \in \mathcal{T}_N$ mit dieser Hyperebene geschnitten wird. Dazu wird für jede Kante $e_i = (x_{i,1}, x_{i,2})$ eines Pentatops τ_k der Schnittpunkt mit dieser Ebene berechnet. Jeder Punkt x auf der Kante e_i lässt sich darstellen als

$$x = x_{i,1} + \lambda(x_{i,2} - x_{i,1}) \quad \text{für } \lambda \in [0,1].$$

Ein Schnittpunkt x_i der Kante e_i mit der Hyperebene E_4 muss somit die Gleichung

$$x_{i,1} + \lambda(x_{i,2} - x_{i,1}) = P_0 + \mu_1 P_1 + \mu_2 P_2 + \mu_3 P_3$$

erfüllen. Dies ist äquivalent zum linearen Gleichungssystem

$$\underbrace{\begin{pmatrix} & & & \\ P_1 & P_2 & P_3 & x_{i,1} - x_{i,2} \\ & & & \end{pmatrix}}_{=:A_i} \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \lambda \end{pmatrix} = \left(x_{i,1} - P_0\right). \tag{3.5}$$

Die Matrix A_i ist genau dann invertierbar wenn der Vektor $x_{i,1} - x_{i,2}$ zu den Vektoren P_1, P_2 und P_3 linear unabhängig ist. Das heißt, die Matrix A_i ist nicht invertierbar, wenn die Kante e_i parallel zur Hyperebene E_4 liegt. In diesem Fall gibt es entweder keinen oder

unendlich viele Schnittpunkte. Ist die Matrix A_i invertierbar, so können die Koeffizienten $\mu_1, \mu_2, \mu_3, \lambda \in \mathbb{R}$ eindeutig bestimmt werden. Gilt zusätzlich $0 \leq \lambda \leq 1$, so ist

$$x_i = x_{i,1} + \lambda(x_{i,2} - x_{i,1}) = P_0 + \mu_1 P_1 + \mu_2 P_2 + \mu_3 P_3$$

ein Schnittpunkt der Kante e_i mit der Hyperebene E_4. Für ein Pentatop τ_k beinhaltet die Menge

$$D_k := \{x_i : x_i \text{ ist Schnittpunkt der Kante } e_i \text{ mit der Hyperebene } E_4\}$$

alle Schnittpunkte, die benötigt werden, um ein dreidimensionales Objekt darzustellen. Je nach Anzahl der Schnittpunkte gibt es folgende Fälle:

- Ist $|D_k| = 4$, so bilden die Schnittpunkte D_k einen Tetraeder.
- Für $|D_k| = 6$ ergibt sich im Allgemeinen ein schiefes Prisma.
- Ist $|D_k| \leq 3$, so sind die Schnittpunkte D_k für die Visualisierung nicht relevant.

Für $t \in [0, T]$ lässt sich mit den Vektoren

$$P_0 = (0,0,0,t)^\top, P_1 = (1,0,0,0)^\top, P_2 = (0,1,0,0)^\top \text{ und } P_3 = (0,0,1,0)^\top$$

zu jedem Zeitunkt t ein dreidimensionales Objekt berechnen, welches mit den bekannten Methoden visiualisiert werden kann [16].

4 Numerische Beispiele

In diesem Kapitel werden numerische Beispiele für das in Kapitel 1 vorgestellte Kontrollproblem (1.1)–(1.2) für $d = 1$ und $d = 3$ betrachtet. Dazu wird das in Kapitel 2 diskretisierte Optimalitätssystem (2.14)–(2.16) für verschiedene Paramter ε gelöst und die jeweiligen Fehler in der L_2–Norm angegeben. Dabei ergeben sich für die L_2–Fehler des Zustandes u und der Steuerung z die erwarteten Konvergenzordnungen.

4.1 Modellprobleme für $d = 1$

4.1.1 Beispiel zur Konvergenzuntersuchung

Gegeben sei das Gebiet $\Omega = (0, 1)$ und die Zeit $T = 1$. Der Raum Zeit–Zylinder Q ist somit durch $Q = (0, 1)^2$ gegeben. Weiters sei der Kostenkoeffizient $\varrho = 1$ und die gewünschte Temperaturverteilung sei

$$\bar{u}(x) = \left(1 + 2\pi^2\right) 10^{-3} e^{\pi^2} \sin(\pi x) \quad \text{für } x \in \Omega.$$

Am Anfang ist die Temperatur mit $u_0(x) = 10^{-3} \sin(\pi x)$ vorgegeben. Am Rand Σ sind homogene Dirichletdaten gegeben. Die exakten Lösungen des Optimalitätssystems (1.24)–(1.26) lauten somit

$$\begin{aligned} u(x,t) &= 10^{-3} e^{\pi^2 t} \sin(\pi x), \\ z(x,t) &= 2\pi^2 10^{-3} e^{\pi^2 t} \sin(\pi x), \\ p(x,t) &= -2\pi^2 10^{-3} e^{\pi^2 t} \sin(\pi x). \end{aligned}$$

Um näherungsweise Lösungen für das Optimalitätsystem (1.24)–(1.26) zu bestimmen, wurde das diskrete Optimalitätssystem (2.14)–(2.16) gelöst. Dabei wurde für den Raum–Zeit–Zylinder $Q = (0, 1)^2$ eine Ausgangszerlegung mit $N = 4$ Dreiecken verwendet. Für weitere Level wurden diese immer gleichmäßig verfeinert. Weiters wurden für die Steuerung z und für den Zustand u die gleichen Ansatzräume verwendet. Somit kann zur Bestimmung der Näherungslösungen u_h und z_h das Gleichungssystem (2.18) verwendet werden. Dieses wurde mit dem Löser PARDISO [24, 25] gelöst. Für die Parameter $\varepsilon \in \{-1, 0, 1\}$, $\sigma = 1$, $\beta = 100$ und für die Polynomgrade $r = 1, 2, 3, 4$ sind in den Tabellen 4.1–4.12 die jeweiligen Fehler in der L_2–Norm angegeben. Dabei ergibt sich die Anzahl der Freiheitsgrade des Gleichungssystems (2.18) durch die Beziehung

$$N_{\text{dof}} = (r + 1)(r + 2)N,$$

wobei N die Anzahl der Elemente bezeichnet. Weiters wird mit

$$\text{eoc} := \log_2 \left(\frac{\left\| u - u_{h_{l-1}} \right\|_{L_2(Q)}}{\left\| u - u_{h_l} \right\|_{L_2(Q)}} \right)$$

die numerisch erhaltene Konvergenzrate zum Level l bezeichnet.
Für $\varepsilon = -1$ ist für die L_2–Fehler des Zustandes u und der Steuerung z immer eine optimale Konvergenzordnung von $\text{eoc} = r + 1$ bezüglich der Polynomgrade $r = 1, 2, 3, 4$ zu beobachten, siehe dazu die Tabellen 4.1–4.4. Für $\varepsilon \in \{0, 1\}$ ist jedoch nur für ungerade Polynomgrade, also $r = 1, 3$ eine optimale Konvergenzordnung von $\text{eoc} = r + 1$ gegeben. Für $r = 2, 4$ ergibt sich nur eine Konvergenzordnung von $\text{eoc} = r$. Dieses Verhalten für den Fehler in der L_2–Norm ist auch für den rein elliptischen Fall zu beobachten, siehe [20, 23].

Level	Elemente	Freiheitsgrade	$\|u - u_h\|_{L_2(Q)}$	eoc	$\|z - z_h\|_{L_2(Q)}$	eoc
0	4	24	$3.0356 + 0$	–	$5.6442 + 1$	–
1	16	96	$1.1869 + 0$	1.35	$2.2435 + 1$	1.33
2	64	384	$4.6311 - 1$	1.36	$9.0583 + 0$	1.31
3	256	1536	$1.4842 - 1$	1.64	$3.0984 + 0$	1.55
4	1024	6144	$4.0841 - 2$	1.86	$9.1944 - 1$	1.75
5	4096	24576	$1.0556 - 2$	1.95	$2.5091 - 1$	1.87
6	16384	98304	$2.6752 - 3$	1.98	$6.5568 - 2$	1.94
7	65536	393216	$6.7318 - 4$	1.99	$1.6766 - 2$	1.97
8	262144	1572864	$1.6885 - 4$	2.00	$4.2405 - 3$	1.98

Tabelle 4.1: Ergebnisse für $\varepsilon = -1$, $\beta = 1$, $\sigma = 100$ und $r = 1$.

Level	Elemente	Freiheitsgrade	$\|u - u_h\|_{L_2(Q)}$	eoc	$\|z - z_h\|_{L_2(Q)}$	eoc
0	4	48	$1.4287 + 0$	–	$2.0095 + 1$	–
1	16	192	$2.9668 - 1$	2.27	$4.6402 + 0$	2.11
2	64	768	$5.3210 - 2$	2.48	$9.5004 - 1$	2.29
3	256	3072	$7.1474 - 3$	2.90	$1.5109 - 1$	2.65
4	1024	12288	$8.8310 - 4$	3.02	$2.1051 - 2$	2.84
5	4096	49152	$1.0794 - 4$	3.03	$2.7679 - 3$	2.93
6	16384	196608	$1.3296 - 5$	3.02	$3.5461 - 4$	2.96
7	65536	786432	$1.6479 - 6$	3.01	$4.4869 - 5$	2.98
8	262144	3145728	$2.0499 - 7$	3.01	$5.6423 - 6$	2.99

Tabelle 4.2: Ergebnisse für $\varepsilon = -1$, $\beta = 1$, $\sigma = 100$ und $r = 2$.

Level	Elemente	Freiheitsgrade	$\|u - u_h\|_{L_2(Q)}$	eoc	$\|z - z_h\|_{L_2(Q)}$	eoc
0	4	80	$4.3473 - 1$	–	$5.7248 + 0$	–
1	16	320	$5.5692 - 2$	2.96	$8.6034 - 1$	2.73
2	64	1280	$4.8595 - 3$	3.52	$9.0357 - 2$	3.25
3	256	5120	$3.4060 - 4$	3.83	$7.0670 - 3$	3.68
4	1024	20480	$2.1719 - 5$	3.97	$4.7727 - 4$	3.89
5	4096	81920	$1.3352 - 6$	4.02	$3.0461 - 5$	3.97
6	16384	327680	$8.1300 - 8$	4.04	$1.9038 - 6$	4.00
7	65536	1310720	$4.9535 - 9$	4.04	$1.1807 - 7$	4.01

Tabelle 4.3: Ergebnisse für $\varepsilon = -1$, $\beta = 1$, $\sigma = 100$ und $r = 3$.

Level	Elemente	Freiheitsgrade	$\|u - u_h\|_{L_2(Q)}$	eoc	$\|z - z_h\|_{L_2(Q)}$	eoc
0	4	120	$1.2464 - 1$	–	$1.6641 + 0$	–
1	16	480	$1.1446 - 2$	3.44	$1.7667 - 1$	3.24
2	64	1920	$5.3339 - 4$	4.42	$9.7153 - 3$	4.18
3	256	7680	$1.9862 - 5$	4.75	$3.8069 - 4$	4.67
4	1024	30720	$6.4761 - 7$	4.94	$1.3989 - 5$	4.77
5	4096	122880	$1.9714 - 8$	5.04	$3.9565 - 7$	5.14
6	16384	491520	$5.9418 - 10$	5.05	$1.2032 - 8$	5.04

Tabelle 4.4: Ergebnisse für $\varepsilon = -1$, $\beta = 1$, $\sigma = 100$ und $r = 4$.

Level	Elemente	Freiheitsgrade	$\|u - u_h\|_{L_2(Q)}$	eoc	$\|z - z_h\|_{L_2(Q)}$	eoc
0	4	24	$3.0158 + 0$	–	$5.6245 + 1$	–
1	16	96	$1.2122 + 0$	1.31	$2.2528 + 1$	1.32
2	64	384	$4.7327 - 1$	1.36	$9.0747 + 0$	1.31
3	256	1536	$1.5049 - 1$	1.65	$3.0952 + 0$	1.55
4	1024	6144	$4.1228 - 2$	1.87	$9.1725 - 1$	1.75
5	4096	24576	$1.0639 - 2$	1.95	$2.5020 - 1$	1.87
6	16384	98304	$2.6946 - 3$	1.98	$6.5378 - 2$	1.94
7	65536	393216	$6.7751 - 4$	1.99	$1.6716 - 2$	1.97
8	262144	1572864	$1.7002 - 4$	1.99	$4.2282 - 3$	1.98

Tabelle 4.5: Ergebnisse für $\varepsilon = 0$, $\beta = 1$, $\sigma = 100$ und $r = 1$.

Level	Elemente	Freiheitsgrade	$\|u-u_h\|_{L_2(Q)}$	eoc	$\|z-z_h\|_{L_2(Q)}$	eoc
0	4	48	$1.5296+0$	–	$2.0657+1$	–
1	16	192	$3.2101-1$	2.25	$4.7307+0$	2.13
2	64	768	$5.6576-2$	2.50	$9.5897-1$	2.30
3	256	3072	$7.7587-3$	2.87	$1.5294-1$	2.65
4	1024	12288	$1.0194-3$	2.93	$2.2118-2$	2.79
5	4096	49152	$1.5369-4$	2.73	$2.9173-3$	2.92
6	16384	196608	$2.9307-5$	2.39	$4.1464-4$	2.81
7	65536	786432	$6.7093-6$	2.13	$6.9486-5$	2.58
8	262144	3145728	$1.6458-6$	2.03	$1.4446-5$	2.27

Tabelle 4.6: Ergebnisse für $\varepsilon = 0$, $\beta = 1$, $\sigma = 100$ und $r = 2$.

Level	Elemente	Freiheitsgrade	$\|u-u_h\|_{L_2(Q)}$	eoc	$\|z-z_h\|_{L_2(Q)}$	eoc
0	4	80	$5.1042-1$	–	$6.2151+0$	–
1	16	320	$6.1105-2$	3.06	$8.8904-1$	2.81
2	64	1280	$5.0745-3$	3.59	$9.1191-2$	3.29
3	256	5120	$3.4957-4$	3.86	$7.1196-3$	3.68
4	1024	20480	$2.2138-5$	3.98	$4.8191-4$	3.88
5	4096	81920	$1.3577-6$	4.03	$3.0818-5$	3.97
6	16384	327680	$8.2650-8$	4.04	$1.9289-6$	4.00
7	65536	1310720	$5.0380-9$	4.04	$1.1978-7$	4.01

Tabelle 4.7: Ergebnisse für $\varepsilon = 0$, $\beta = 1$, $\sigma = 100$ und $r = 3$.

Level	Elemente	Freiheitsgrade	$\|u-u_h\|_{L_2(Q)}$	eoc	$\|z-z_h\|_{L_2(Q)}$	eoc
0	4	120	$1.5732-1$	–	$1.8823+0$	–
1	16	480	$1.2211-2$	3.69	$1.8366-1$	3.36
2	64	1920	$5.5402-4$	4.46	$9.8942-3$	4.21
3	256	7680	$2.2027-5$	4.65	$3.9243-4$	4.66
4	1024	30720	$9.1798-7$	4.58	$1.3864-5$	4.82
5	4096	122880	$4.7604-8$	4.27	$5.2686-7$	4.72
6	16384	491520	$2.9434-9$	4.02	$2.5818-8$	4.35

Tabelle 4.8: Ergebnisse für $\varepsilon = 0$, $\beta = 1$, $\sigma = 100$ und $r = 4$.

Level	Elemente	Freiheitsgrade	$\|u - u_h\|_{L_2(Q)}$	eoc	$\|z - z_h\|_{L_2(Q)}$	eoc
0	4	24	$3.0027 + 0$	–	$5.6101 + 1$	–
1	16	96	$1.2383 + 0$	1.28	$2.2631 + 1$	1.31
2	64	384	$4.8352 - 1$	1.36	$9.0944 + 0$	1.32
3	256	1536	$1.5262 - 1$	1.66	$3.0934 + 0$	1.56
4	1024	6144	$4.1635 - 2$	1.87	$9.1550 - 1$	1.76
5	4096	24576	$1.0727 - 2$	1.96	$2.4961 - 1$	1.87
6	16384	98304	$2.7153 - 3$	1.98	$6.5216 - 2$	1.94
7	65536	393216	$6.8290 - 4$	1.99	$1.6675 - 2$	1.97
8	262144	1572864	$1.7126 - 4$	2.00	$4.2175 - 3$	1.98

Tabelle 4.9: Ergebnisse für $\varepsilon = 1$, $\beta = 1$, $\sigma = 100$ und $r = 1$.

Level	Elemente	Freiheitsgrade	$\|u - u_h\|_{L_2(Q)}$	eoc	$\|z - z_h\|_{L_2(Q)}$	eoc
0	4	48	$1.6093 + 0$	–	$2.1149 + 1$	–
1	16	192	$3.4373 - 1$	2.23	$4.8244 + 0$	2.13
2	64	768	$6.0113 - 2$	2.52	$9.7040 - 1$	2.31
3	256	3072	$8.5241 - 3$	2.82	$1.5560 - 1$	2.64
4	1024	12288	$1.2544 - 3$	2.76	$2.2349 - 2$	2.80
5	4096	49152	$2.3008 - 4$	2.45	$3.2399 - 3$	2.79
6	16384	196608	$5.1915 - 5$	2.15	$5.4231 - 4$	2.58
7	65536	786432	$1.2722 - 5$	2.03	$1.1213 - 4$	2.27
8	262144	3145728	$3.1800 - 6$	2.00	$2.6451 - 5$	2.08

Tabelle 4.10: Ergebnisse für $\varepsilon = 1$, $\beta = 1$, $\sigma = 100$ und $r = 2$.

Level	Elemente	Freiheitsgrade	$\|u - u_h\|_{L_2(Q)}$	eoc	$\|z - z_h\|_{L_2(Q)}$	eoc
0	4	80	$5.6573 - 1$	–	$6.6047 + 0$	–
1	16	320	$6.6370 - 2$	3.09	$9.2143 - 1$	2.84
2	64	1280	$5.3104 - 3$	3.64	$9.2555 - 2$	3.32
3	256	5120	$3.6176 - 4$	3.88	$7.2136 - 3$	3.68
4	1024	20480	$2.2824 - 5$	3.99	$4.8933 - 4$	3.88
5	4096	81920	$1.3984 - 6$	4.03	$3.1350 - 5$	3.96
6	16384	327680	$8.5170 - 8$	4.04	$1.9648 - 6$	4.00
7	65536	1310720	$5.1977 - 9$	4.03	$1.2211 - 7$	4.01

Tabelle 4.11: Ergebnisse für $\varepsilon = 1$, $\beta = 1$, $\sigma = 100$ und $r = 3$.

Level	Elemente	Freiheitsgrade	$\|u - u_h\|_{L_2(Q)}$	eoc	$\|z - z_h\|_{L_2(Q)}$	eoc
0	4	120	$1.8144 - 1$	–	$2.0706 + 0$	–
1	16	480	$1.3032 - 2$	3.80	$1.9185 - 1$	3.43
2	64	1920	$5.9341 - 4$	4.46	$1.0179 - 2$	4.24
3	256	7680	$2.7246 - 5$	4.44	$4.3461 - 4$	4.55
4	1024	30720	$1.3999 - 6$	4.28	$1.6287 - 5$	4.74
5	4096	122880	$8.4580 - 8$	4.05	$7.6411 - 7$	4.41
6	16384	491520	$5.4128 - 9$	3.97	$4.4192 - 8$	4.11

Tabelle 4.12: Ergebnisse für $\varepsilon = 1$, $\beta = 1$, $\sigma = 100$ und $r = 4$.

4.1.2 Anwendungsbeispiel

Gegeben sei das Gebiet $\Omega = (0, 1)$, welches einen Draht darstellen soll. Am Rand Γ wird dabei die Temperatur zu jeder Zeit durch homogene Dirichletrandbedingungen modelliert. Am Anfang weist der Draht eine Temperatur $u_0 = x(1 - x)$ auf. Wegen der homogenen Dirichletrandbedingungen würde der Draht auf die Temperatur $u = 0$ abkühlen. Ziel ist es nun, den Draht so zu erwärmen, dass dieser zum Zeitpunkt $T = 1$ wieder seine Anfangstemperatur aufweist, also $\bar{u} = x(1 - x)$. Dabei ist der Kostenkoeffizient mit $\varrho = 10^{-3}$ gegeben. Diese Aufgabenstellung führt auf das in dieser Arbeit vorgestellte Kontrollproblem. Für die näherungsweisen Lösungen des Optimalitätssystems (1.24)–(1.26) wurde der Raum–Zeit–Zylinder $Q = (0, 1)^2$ gleichmäßig in 16384 Elemente unterteilt. Weiters wurden für die Steuerung z und für die Zustände u und p die gleichen Ansatzräume mit einem Polynomgrad $r = 3$ verwendet. Die näherungsweisen Lösungen des Zustands u, des adjungierten Zustands p und der Steuerung z sind in den Abbildungen 4.1–4.3 dargestellt. Dabei ist zu erkennen, dass die Temperatur u relativ schnell auf die Temperatur $u = 0$ abkühlt. Durch die Wirkung der Steuerung z, welche erst kurz vor dem Zeitpunkt $T = 1$ einsetzt, wird die Temperatur u wieder auf die vorgegebene Temperatur $\bar{u}$ gebracht.
Da der Zustand u und die Steuerung z ein unterschiedliches Verhalten aufweisen, ist es von Vorteil für beide unterschiedliche Netze zu verwenden. Für stationäre Probleme siehe dafür zum Beispiel [34].

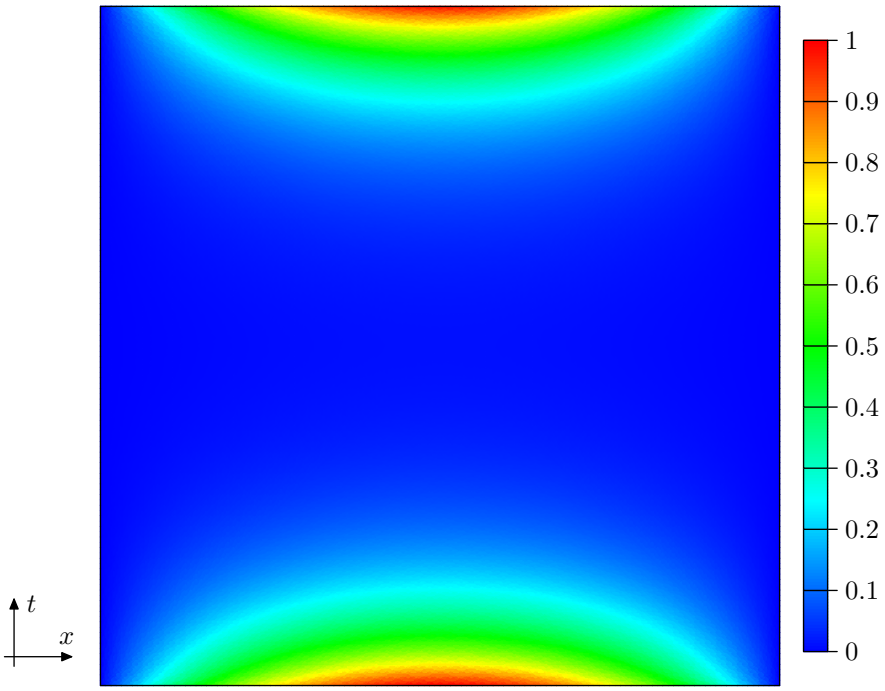


Abbildung 4.1: Zustand u

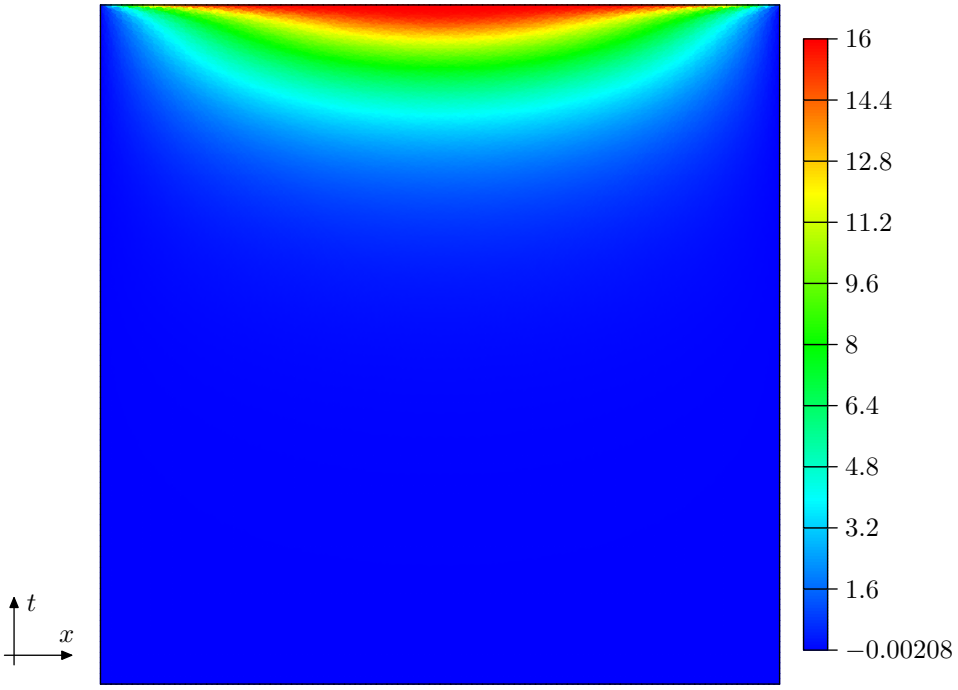


Abbildung 4.2: Steuerung z

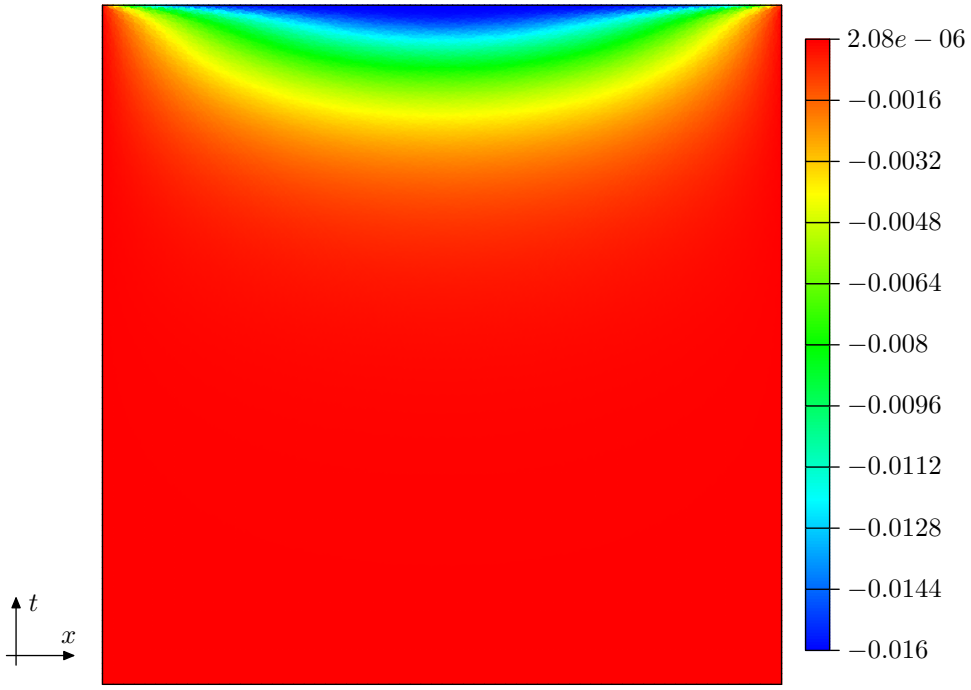


Abbildung 4.3: Adjungierter Zustand p

4.2 Modellprobleme für $d = 3$

Analog wie für den Fall $d = 1$ wird hier ein numerisches Beispiel für $d = 3$ betrachtet. Dazu sei das Gebiet $\Omega = (0, 9)^3$, die Zeit $T = 9$, der Kostenkoeffizient $\varrho = 1$ und eine gewünschte Temperaturverteilung

$$\bar{u}(x, y, z) = 10^{-3} \left(1 + \frac{2\pi^2}{27}\right) e^{\frac{\pi^2}{3}} \sin\left(\frac{\pi x}{9}\right) \sin\left(\frac{\pi y}{9}\right) \sin\left(\frac{\pi z}{9}\right)$$

gegeben. Am Anfang sei eine Temperatur

$$u_0(x, y, z) = 10^{-3} \sin\left(\frac{\pi x}{9}\right) \sin\left(\frac{\pi y}{9}\right) \sin\left(\frac{\pi z}{9}\right)$$

vorgegeben. Weiters sei am Rand Γ die Temperatur u gleich Null. Die exakten Lösungen des Optimalitätssystems (1.24)–(1.26) lauten somit

$$\begin{aligned}
u(x, y, z, t) &= 10^{-3} e^{\frac{\pi^2}{27} t} \sin\left(\frac{\pi x}{9}\right) \sin\left(\frac{\pi y}{9}\right) \sin\left(\frac{\pi z}{9}\right), \\
z(x, y, z, t) &= 10^{-3} \frac{2\pi^2}{27} e^{\frac{\pi^2}{27} t} \sin\left(\frac{\pi x}{9}\right) \sin\left(\frac{\pi y}{9}\right) \sin\left(\frac{\pi z}{9}\right), \\
p(x, y, z, t) &= -10^{-3} \frac{2\pi^2}{27} e^{\frac{\pi^2}{27} t} \sin\left(\frac{\pi x}{9}\right) \sin\left(\frac{\pi y}{9}\right) \sin\left(\frac{\pi z}{9}\right).
\end{aligned}$$

Das Lösen des diskreten Optimalitätssystems (2.14)–(2.16) für $\varepsilon = -1$ beziehungsweise $\varepsilon = 0$ liefert die in den Tabellen 4.13–4.18 angegebenen L_2–Fehler. Dabei wurden für den Zustand u und für die Steuerung z wie in Abschnitt 4.1.1 die gleichen Ansatzräume verwendet. Weiters wurde der Parameter $\beta = 1$ gewählt und für den Penalty σ wurde der in den Tabellen 4.13–4.18 angegebene Wert verwendet. Für den Raum–Zeit–Zylinder $Q = (0, 9)^4$ wurden die in Tabelle 3.3 angegebenen Zerlegungen benützt. Für die Anzahl der Freiheitsgrade des Gleichungssystems (2.18) ergibt sich die Formel

$$N_{\text{dof}} = \frac{1}{12}(r+1)(r+2)(r+3)(r+4)N.$$

Für verschiedene Polynomgrade $r = 1, 2, 3$ ist in den Tabellen 4.13–4.18 für die L_2–Fehler eine optimale Konvergenzordnung von eoc $= r + 1$ zu beobachten. Für den Fall $\varepsilon = 1$ und einem Polynomgrad von $r = 2$ sollte eigentlich nur eine Konvergenzordnung von eoc $= r$ zu beobachten sein. Dies ist jedoch in den ersten vier Level noch nicht zu sehen, siehe dazu auch Tabelle 4.6.

Level	Elemente	Freiheitsgrade	$\|u - u_h\|_{L_2(Q)}$	eoc	$\|z - z_h\|_{L_2(Q)}$	eoc
0	24	240	$1.7950 - 1$	–	$1.2312 - 1$	–
1	384	3840	$1.1780 - 1$	0.61	$6.2051 - 2$	0.99
2	6144	61440	$4.9120 - 2$	1.26	$2.9581 - 2$	1.07
3	98304	983040	$1.4085 - 2$	1.80	$9.4513 - 3$	1.65
4	1572864	15728640	$3.7658 - 3$	1.90	$2.7401 - 3$	1.79

Tabelle 4.13: Ergebnisse für $\varepsilon = -1$, $\beta = 1$, $\sigma = 1000$ und $r = 1$.

Level	Elemente	Freiheitsgrade	$\|u - u_h\|_{L_2(Q)}$	eoc	$\|z - z_h\|_{L_2(Q)}$	eoc
0	24	720	$9.4717 - 2$	–	$6.0743 - 2$	–
1	384	11520	$3.5922 - 2$	1.4	$2.2498 - 2$	1.43
2	6144	184320	$6.0095 - 3$	2.58	$4.0194 - 3$	2.48
3	98304	2949120	$7.9776 - 4$	2.91	$5.6269 - 4$	2.84

Tabelle 4.14: Ergebnisse für $\varepsilon = -1$, $\beta = 1$, $\sigma = 5000$ und $r = 2$.

Level	Elemente	Freiheitsgrade	$\|u-u_h\|_{L_2(Q)}$	eoc	$\|z-z_h\|_{L_2(Q)}$	eoc
0	24	1680	$5.4525-2$	–	$3.6752-2$	–
1	384	26880	$7.9829-3$	2.77	$5.1685-3$	2.83
2	6144	430080	$5.9965-4$	3.73	$4.1652-4$	3.63
3	98304	6881280	$3.9719-5$	3.92	$2.8469-5$	3.87

Tabelle 4.15: Ergebnisse für $\varepsilon = -1$, $\beta = 1$, $\sigma = 5000$ und $r = 3$.

Level	Elemente	Freiheitsgrade	$\|u-u_h\|_{L_2(Q)}$	eoc	$\|z-z_h\|_{L_2(Q)}$	eoc
0	24	240	$1.7457-1$	–	$1.2199-1$	–
1	384	3840	$1.1752-1$	0.05	$6.0605-2$	1.01
2	6144	61440	$4.9033-2$	1.31	$2.9534-2$	1.04
3	98304	983040	$1.4075-2$	1.80	$9.4532-3$	1.64
4	1572864	15728640	$3.7665-3$	1.90	$2.7565-3$	1.78

Tabelle 4.16: Ergebnisse für $\varepsilon = 0$, $\beta = 1$, $\sigma = 1000$ und $r = 1$.

Level	Elemente	Freiheitsgrade	$\|u-u_h\|_{L_2(Q)}$	eoc	$\|z-z_h\|_{L_2(Q)}$	eoc
0	24	720	$9.3427-2$	–	$5.9502-2$	–
1	384	11520	$3.5944-2$	1.38	$2.2491-2$	1.4
2	6144	184320	$6.0079-3$	2.58	$4.0199-3$	2.48
3	98304	2949120	$7.9775-4$	2.91	$5.6274-4$	2.84

Tabelle 4.17: Ergebnisse für $\varepsilon = 0$, $\beta = 1$, $\sigma = 5000$ und $r = 2$.

Level	Elemente	Freiheitsgrade	$\|u-u_h\|_{L_2(Q)}$	eoc	$\|z-z_h\|_{L_2(Q)}$	eoc
0	24	1680	$5.5751-2$	–	$3.6616-2$	–
1	384	26880	$8.0054-3$	2.8	$5.1695-3$	2.82
2	6144	430080	$6.0023-4$	3.74	$4.1679-4$	3.63
3	98304	6881280	$3.9710-5$	3.92	$2.8472-5$	3.87

Tabelle 4.18: Ergebnisse für $\varepsilon = 0$, $\beta = 1$, $\sigma = 5000$ und $r = 3$.

5 Ausblick

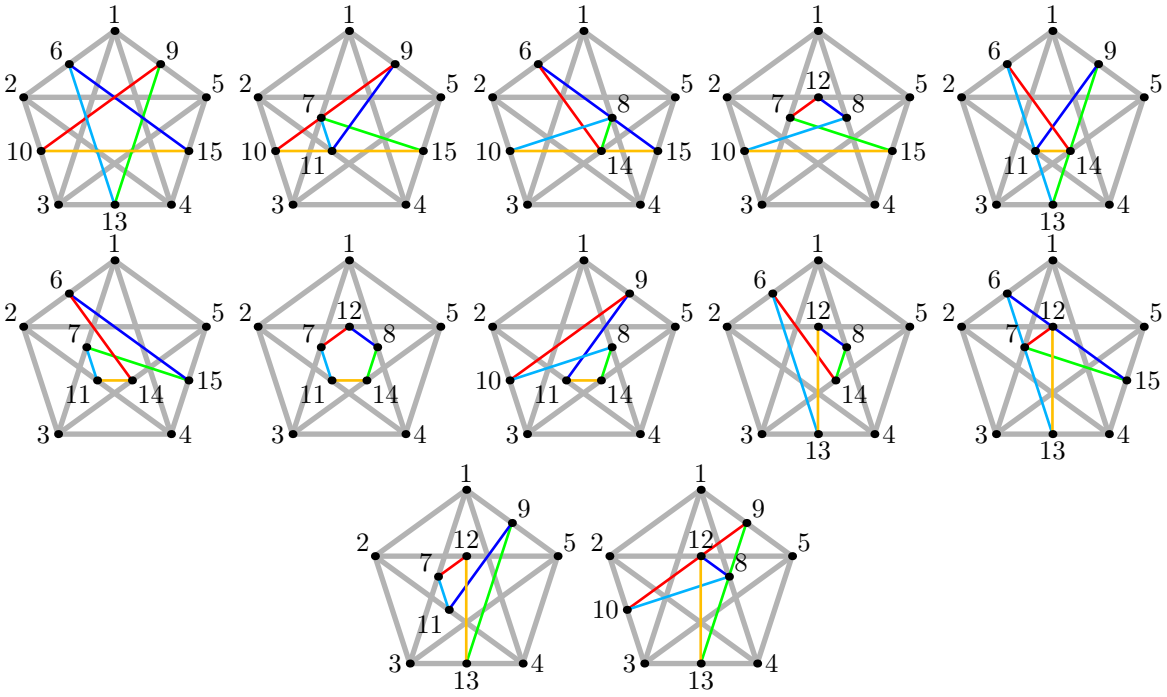
In dieser Arbeit wurde ein optimales Kontrollproblem mit verteilter Steuerung betrachtet. Für dieses wurde ein äquivalentes Optimalitätssystem hergeleitet, welches mithilfe der Discontinuous Galerkin Methode diskretisiert wurde. Im vierten Kapitel wurden numerische Beispiele angegeben, die die erwartete Konvergenzordnung zeigen. Das diskrete Optimalitätssystem wurde dabei ohne Vorkonditionierung gelöst. Für die Triangulierung eines Raum–Zeit–Zylinder $Q \subset \mathbb{R}^4$ steigt die Anzahl der Elemente bei einer gleichmäßigen Verfeinerung um den Faktor 16 an, wodurch sich auch die Anzahl der Freiheitsgrade des diskreten Optimalitätssystems um den Faktor 16 multipliziert. Deshalb ist die Untersuchung geeigneter Vorkonditionierungsstrategien notwendig, um das Optimalitätssystem effizient lösen zu können. Eine Multigrid Vorkonditionierung für Advektions–Diffusions–Probleme wird zum Beispiel in [12] beschrieben.
Da die Anzahl der Freiheitsgrade des diskreten Optimalitätssystems für einen Raum–Zeit–Zylinder $Q \subset \mathbb{R}^4$ sehr groß werden kann, ist ein weiterer interessanter Aspekt die Anwendung von Gebietszerlegungsmethoden auf diese Problemstellung, siehe hier zum Beispiel [9]. Weiters ist die numerische Behandlung von anderen Zustandsgleichungen, welche andere physikalische Prozesse modellieren, ein weiterer interessanter Gesichtspunkt, siehe hier zum Beispiel [31].

Anhang

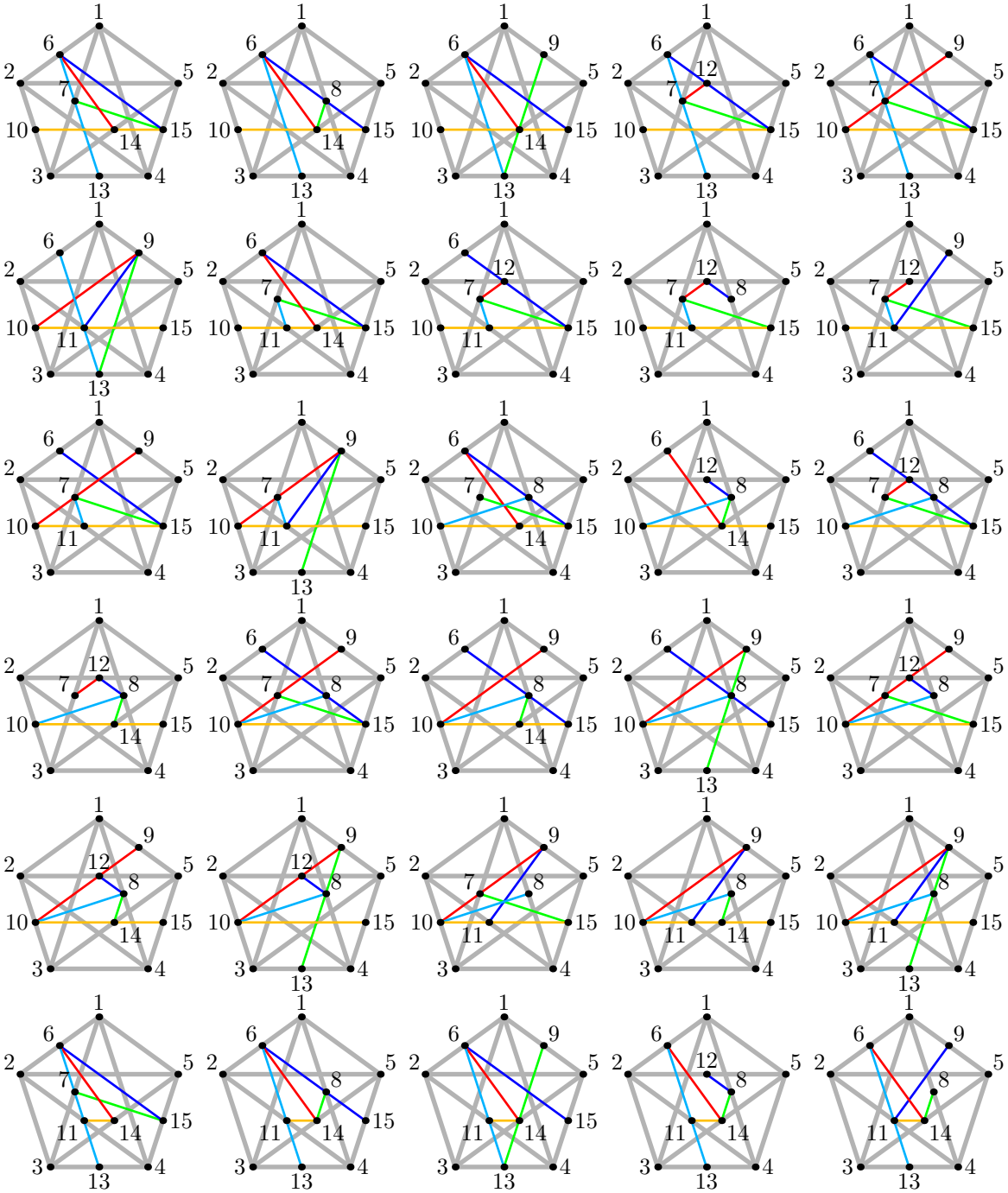
Zusammenhängende Fixierungen

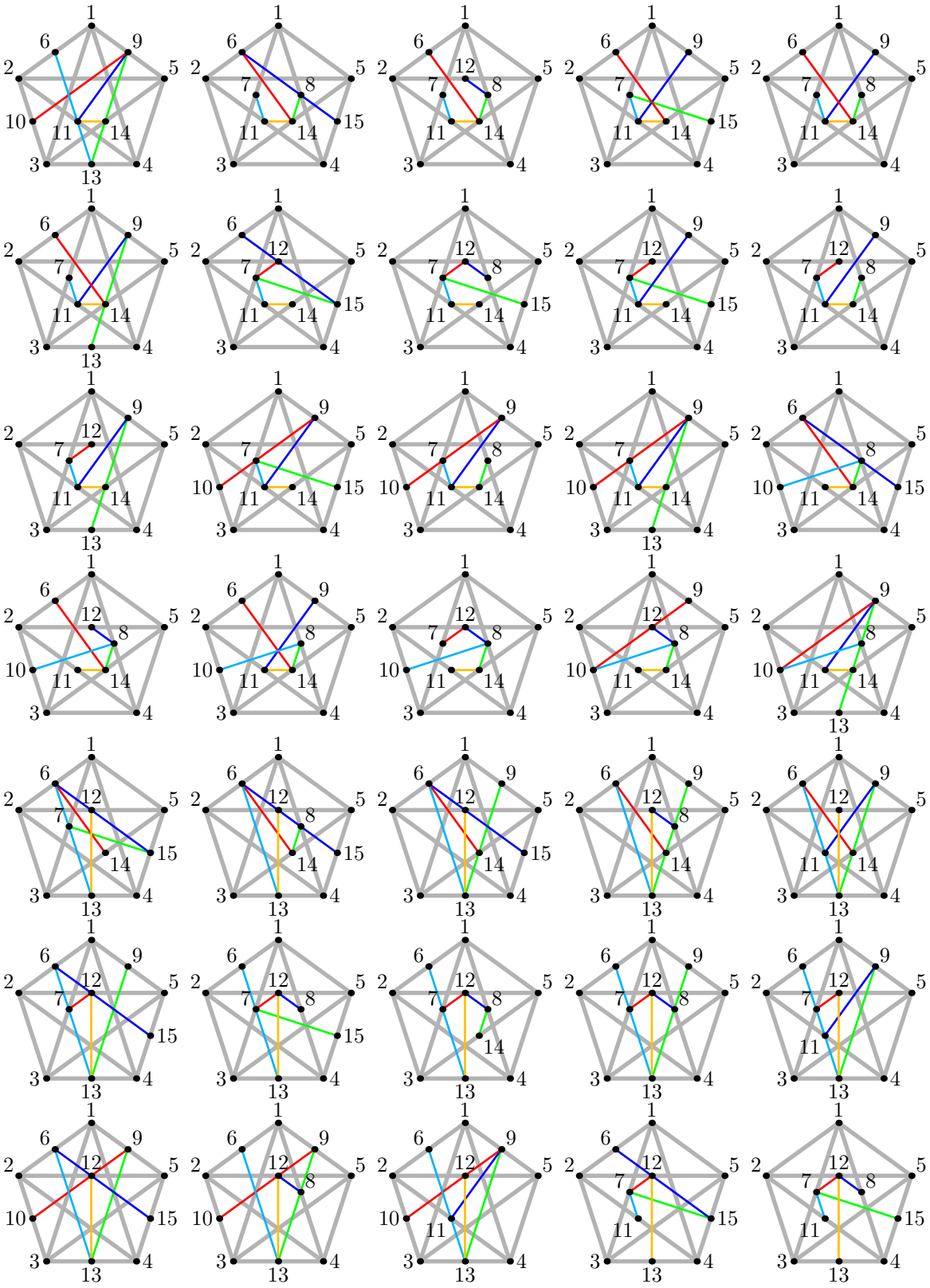
Hier sind alle möglichen zusammenhängenden Fixierungen grafisch dargestellt. Von den 243 möglichen Fixierungen gibt es insgesamt 87 zusammenhängende Fixierungen. Davon sind 12 zyklische und 75 sind azyklische Fixierungen.

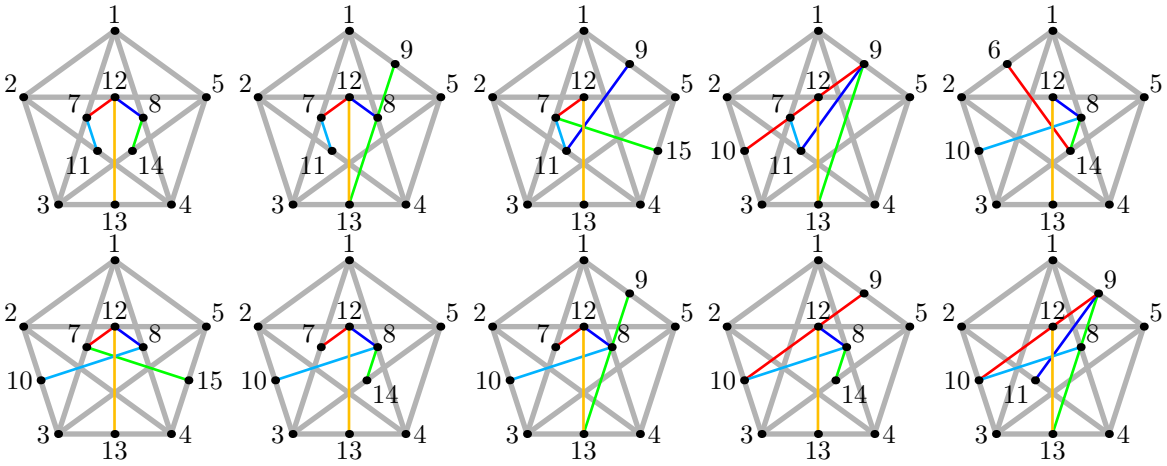
Zyklische Fixierungen



Azyklische Fixierungen






1 2 5 12 7 8 11 14 3 13 4
1 9 2 5 12 7 8 11 3 13 4
1 9 2 5 12 7 15 11 3 13 4
1 9 2 5 12 7 10 11 3 13 4
1 6 2 5 12 8 10 14 3 13 4
1 2 5 12 7 8 10 15 3 13 4
1 2 5 12 7 8 10 14 3 13 4
1 9 2 5 12 7 8 10 3 13 4
1 9 2 5 12 8 10 14 3 13 4
1 9 2 5 12 8 10 11 3 13 4

Literaturverzeichnis

[1] H. Amann and J. Escher. *Analysis I.* Birkhäuser, Basel-Bosten-Berlin, 2002.

[2] D. N. Arnold, F. Brezzi, B. Cockburn, and D. Marini. Unified analysis of discontinuous Galerkin methods for elliptic problems. *SIAM J. Numer. Anal.*, 39:1749–1779, 2002.

[3] C. E. Baumann and J. T. Oden. A discontinuous hp finite element method for convection-diffusion problems. *Comput. Methods Appl. Mech. Engrg*, 175:311–341, 1999.

[4] L. Baumgartner. Zerlegung des n-dimensionalen Raumes in kongruente Simplexe. *Math. Nachr.*, 48:213–224, 1971.

[5] M. Behr. Simplex space–time meshes in finite element simulations. *Internat. J. Numer. Methods Fluids*, 57:1421–1434, 2008.

[6] J. Bey. Simplicial grid refinement on Freudenthal's algorithm and the optimal number of congruence classes. *Numer. Math.*, 85:1–29, 2000.

[7] L.T. Biegler, O. Ghattas, M. Heinkenschloss, D. Keyes, and B. van Bloemen Waanders, editors. *Real–Time PDE–Constrained Optimization.* SIAM, Philadelphia, 2007.

[8] B. Cockburn. Discontinuous Galerkin methods for convection-dominated problems. High-order methods for computational physics. *Lect. Notes Comput. Sci. Eng.*, 9:69–224, 1999.

[9] B. Cockburn, J. Gopalakrishnan, and R. Lazarov. Unified hybridization of discontinuous Galerkin, mixed, and continuous Galerkin methods for second order elliptic problems. *SIAM J. Numer. Anal.*, 47:1319–1365, 2009.

[10] H. S. M. Coxeter. *Regular Polytopes.* Dover Publications, New York, 1973.

[11] H. Gajewski. *Nichtlineare Operatorgleichungen und Operatordifferentialgleichungen.* Akademie-Verlag.

[12] J. Gopalakrishnan and G. Kanschat. A multilevel discontinuous Galerkin method. *Numer. Math.*, 95:527–550, 2003.

[13] M. Haiman. A Simple and Relatively Efficient Triangulation of the n-Cube. *Discrete and Computational Geometry*, 6:287–289, 1991.

[14] M. Hinze, R. Pinnau, M. Ulbrich, and S. Ulbrich. *Optimization with PDE constraints.* Springer, Dordrecht, 2009.

[15] P. Houston, C. Schwab, and E. Süli. Discontinuous hp-Finite Element Methods for Advection-Diffusion-Reaction Problems. *SIAM J. Numer. Anal.*, 39:2133–2163, 2002.

[16] C. R. Johnson and C. Hansen. *The Visualization Handbook.* Elsevier-Butterworth Heinemann, Oxford, 2005.

[17] B.Q. Li. *Discontinuous Finite Elements in Fluid Dynamics and Heat Transfer.* Springer, London, 2006.

[18] J. L. Lions. *Optimal Control Of Systems Governed By Partial Differential Equations.* Springer, New York, 1971.

[19] M. Neumüller. Zur Lösbarkeit optimaler Kontrollprobleme mit elliptischen Randwertaufgaben. Seminar Bakkalaureat TM, Institut für Numerische Mathematik, Technische Universität Graz, 2008.

[20] M. Neumüller. Discontinuous Galerkin Methoden für elliptische Randwertprobleme. Seminar TM, Institut für Numerische Mathematik, Technische Universität Graz, 2009.

[21] M. Neumüller. Triangulierungen im vierdimensionalen Raum. Projekt TM, Institut für Numerische Mathematik, Technische Universität Graz, 2009.

[22] D. Orden. Asymptotically efficient triangulations of the d–cube. *Discrete Comput. Geom.*, 30:509–528, 2003.

[23] B. Rivière. *Discontinuous Galerkin Methods for Solving Elliptic and Parabolic Equations.* SIAM, Philadelphia, 2008.

[24] O. Schenk, M. Bollhöfer, and R. A. Römer. On Large Scale Diagonalization Techniques For The Anderson Model Of Localization. *SIAM Review*, 50(1):91–112, 2008. SIGEST Paper.

[25] O. Schenk, A. Wächter, and M. Hagemann. Matching-based Preprocessing Algorithms to the Solution of Saddle-Point Problems in Large-Scale Nonconvex Interior-Point Optimization. *Comput. Optim. Appl.*, 36(2-3):321–341, April 2007.

[26] C. Schwab. *p- and hp- Finite Element Methods.* Oxford University Press, Oxford-New York, 2004.

[27] O. Steinbach. *Lösungsverfahren für lineare Gleichungssysteme.* Teubner, Stuttgart-Leipzig-Wiesbaden, 2005.

[28] O. Steinbach. *Numerical Approximation Methods for Elliptic Boundary Value Problems.* Springer, New York, 2007.

[29] J. Sudirham. Space-Time Discontinuous Galerkin Methods for Convection-Diffusion Problems. Dissertation, University of Twente, 2005.

[30] V. Thomée. *Galerkin Finite Element Methods for Parabolic Problems.* Springer, Berlin-Heidelberg, 1997.

[31] F. Tröltzsch. *Optimale Steuerung partieller Differentialgleichungen.* Vieweg-Teubner, Wiesbaden, 2009.

[32] T. Warburton and J.S. Hesthaven. On the constants in hp-finite element trace inverse inequalities. *Comput. Methods Appl. Mech. Engrg*, 192:2765–2773, 2003.

[33] J. Wloka. *Partial differential equations.* Cambridge University Press, Cambridge, 1986.

[34] L. Yuan and D. Yang. A posteriori error estimate of optimal control problem of PDE with integral constraint for state. *J. Comput. Math.*, 27:525–542, 2009.

[35] E. Zeidler. *Nonlinear Functional Analysis and its Applications II/A.* Springer, New York, 1988.

Erklärung

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den Quellen wörtlich und inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Ort, Datum	Unterschrift