Karlsruher Institut für Technologie

Bachelorarbeit

Die Multilevel Monte Carlo Methode und deren Anwendung am Beispiel der linearen Transportgleichung

Tim Buchholz

30.02.20

Betreuung: Prof.Dr. Christian Wieners und M.Sc. Niklas Baumgarten

Fakultät für Mathematik

Karlsruher Institut für Technologie

Inhaltsverzeichnis

1 Einleitung

TODO(Einleitung wird zu einem späterem Zeitpunkt noch ausgebaut und nachgebessert mehr cites mehr forschung mehr inhalt) Monte Carlo Methoden sind weit verbreitet und finden in verschiedenen Bereichen der Mathematik ihre Anwendung. Sie dienen dabei als statistische Schätzer für Erwartungswerte. Eine der bekanntesten Anwendungen ist wohl die Monte Carlo Quadratur, welche zur numerischen Integration genutzt werden kann.

Nachdem Giles (cite ...) ... gewöhnliche DGL ... kam ... für SPDE's zu nutzen ...cite .

So entstehende Problemstellungen fallen in das Gebiet der Uncertainty Quantification, einem 'Zusammentreffen der Wahrscheinlichkeitstheorie, Numerik, Statistik und der echten Welt' [28]. Allerdings besitzt die Monte Carlo Methode einen entscheidenden Nachteil, will man sie im Zusammenhang unsicherer Ausgangsdaten für die Lösung von partiellen Differentialgleichungen nutzen, sie konvergiert im Normalfall relativ langsam und das numerische Lösen von PDE's ist oft sehr aufwendig. Es werden also unter Umständen sehr viele, sehr teure Zufallssamples benötigt, um ein vernünftiges Ergebnis zu erhalten.

Diese Thesis soll sich daher mit der Multilevel Monte Carlo Methode (im Folgenden MLMC Methode genannt) beschäftigen, welche an die Monte Carlo Methode angelehnt ist, aber durch die geschickte Auswertung der (Zufalls-Samples) deutliche Effizienzvorteile gegenüber der Standard Monte Carlo Methode besitzt. Die MLMC Methode soll nach einer ausführlichen theoretischen Analyse auch praktisch auf das Transportproblem angewandt werden. Genauer soll für

- ein beschränktes Gebiet $\mathbb{D} \subseteq \mathbb{R}^d$
- ein Zeitintervall $\mathbb{T} = [0, T]$
- ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$
- ein zufälliges Flussvektorfeld $q : \Omega \times \overline{\mathbb{D}} \to \mathbb{R}^d$
- eine Anfangskonzentration eines (zu transportierenden) Stoffes $\rho_0 : \overline{\mathbb{D}} \to \mathbb{R}^d$
- einen Einfluss $\rho_{\text{in}} : \Gamma_{\text{in}} \times \mathbb{T} \to \mathbb{R}$ über den Einflussrand $\Gamma_{\text{in}} := \{z \in \partial\mathbb{D} : q(z) \cdot n(z) \leq 0\} \subset \partial\mathbb{D}$ mit $n(z)$ als äußeren Normalenvektor im (Rand-)Punkt z

der Erwartungswert eines Funktionals der Konzentration des Stoffes $\rho : \overline{\mathbb{D}} \times \mathbb{T} \to \mathbb{R}_{\geq 0}$ bestimmt werden. Dabei erhält man ρ als Lösung der folgenden partiellen Differentialgleichung:

$$\text{Bestimme } \rho : \overline{\mathbb{D}} \times \mathbb{T} \to \mathbb{R}_{\geq 0}, \text{ sodass}$$

$$\text{(TP)} \begin{cases} \partial_t \rho + \operatorname{div}(\rho q) = 0 & \text{in } \mathbb{D} \times (0, T) \\ \rho(x, t) = \rho_{\text{in}}(x, t) & \text{auf } \Gamma_{\text{in}} \times (0, T) \\ \rho(x, 0) = \rho_0(x) & \text{auf } \mathbb{D}. \end{cases}$$

Außerdem muss zunächst ein zwar zufälliges, aber dennoch sinnvolles Vektorfeld q erzeugt werden. Wir nutzen hierbei das Darcy-Gesetz, welches als Modellierung von Fluiden in porösen Bodenschichten bereits oft genutzt wurde (vgl. z.B. [11]). Dabei soll später, bevor wir das eigentliche Transportproblem lösen, stets zunächst für einen zufälligen Permeabilitätstensor, welcher die unbekannte Bodenbeschaffenheit modellieren soll, ein entsprechendes Flussvektorfeld q über das sogenannte Potentialströmungsproblem, welches sich aus dem Darcy-Gesetz ableitet, berechnet werden. Die genauere Modellierung des so entstehenden Gesamtproblems soll aber an späterer Stelle erfolgen.
Die Thesis ist dazu folgendermaßen unterteilt:
Abschnitt 2 sammelt verschiedene Grundlagen aus den Bereichen der Stochastik, der Analysis und Numerik partieller Differentialgleichungen. Besonders werden wir hierbei auf einige zentrale Aussagen der Wahrscheinlichkeitstheorie eingehen, welche für die Konvergenzanalyse von Monte Carlo Methoden im Allgemeinen eine wichtige Rolle spielen. In Abschnitt 3 betrachten wir einige Aspekte der (standard) Monte Carlo Methode, welche auch der MLMC Methode als theoretischer Unterbau dienen sollen. Dabei erklären wir die Monte Carlo Methoden zunächst anhand des Beispiels der numerischen Integration, gehen dann aber auch abstrakter auf Konvergenz und Genauigkeit der Methode ein.
Anschließend werden wir in Abschnitt 4 die Multilevel Monte Carlo Methode an sich erklären. Dazu greifen wir das Beispiel der numerischen Integration aus Abschnitt 3 in einer etwas abgewandelten Form wieder auf. Auch hier wollen wir dann aber auch etwas abstrakter Eigenschaften der Methode betrachten, welche uns auch später bei der Anwendung auf das Transportproblem wieder beschäftigen werden.
In Abschnitt 5 werden dann das Transportproblem und das Potentialströmungsproblem beschrieben, welches wir lösen müssen, um an die entsprechenden Ausgangsdaten zu kommen. Anschließend wird die numerische Lösung der beiden Probleme mit Finite Elemente Methoden behandelt, bevor schließlich in Abschnitt 6 auf die Anwendung der Multilevel Monte Carlo Methode auf das Transportproblem mit unsicheren Ausgangsdaten am Beispiel der Permeabilität κ eingegangen wird.
Der siebte und letzte Abschnitt befasst sich mitder konkreten Durchführung und Implementierung des zuvor theoretisch beleuchteten Problem innerhalb der parallelen Finite Elemente Softwarebibliothek "M++"[1], welche am Institut für Angewandte und Numerische Mathematik 3 (KIT) von Herrn Prof. Dr. C. Wieners entwickelt wurde.
Am Schluss der Thesis steht eine kleine Zusammenfassung der bis dahin erarbeiteten Resultate und der Ausblick auf Möglichkeiten verschiedener Art an, diese weiter zu entwickeln.

2 Grundlagen

2.1 Analytische/numerische Grundlagen

Sei $\mathcal{D} \subseteq \mathbb{R}^d$ offen für $d \in \mathbb{N}$ und $\|\cdot\|$ eine Norm auf $\mathbb{R}^d$ Die folgenden Definitionen und Sätze sollen als Grundlagen für die weiteren Betrachtungen dieser Thesis dienen. Insbesondere wollen wir hierbei meist auf konkrete Beweise verzichten und verweisen dahingehend auf die Literatur. Die analytischen Grundlagen bauen zum Teil auf der Vorlesung Rand- und Eigenwertprobleme aus dem Sommersemester 2019 von Herrn Prof. Dr. Reichel auf, sind aber auch z.B. in [12] oder [14] zu finden.

Definition 2.1. (Einige Operatoren)

(a) Für $F : \mathbb{R}^d \to \mathbb{R}^d$ ist die Divergenz von F definiert durch

$$\operatorname{div} F = \nabla \cdot F \coloneqq \sum_{i=1}^{d} \frac{\partial F_i}{\partial x_i}$$

(b) Für $f : \mathbb{R}^d \to \mathbb{R}$ und $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbb{N}^d$ ist die partielle Ableitung von f nach dem sogenannten Multiindex α definiert durch

$$\partial^\alpha f \coloneqq \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}} = \frac{\partial^{\alpha_1 + \cdots + \alpha_d} f}{\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}}$$

Satz 2.2. (Gaußscher Integralsatz für Lipschitz-Gebiete)
Sei $\mathcal{D} \subset \mathbb{R}^d$ ein beschränktes Lipschitz-Gebiet und sei n der äußere Einheitsnormalenvektor an $\partial\mathcal{D}$. Dann gilt:

$$\int_{\mathcal{D}} \frac{\partial f}{\partial x_i} \,\mathrm{d}x = \int_{\partial\mathcal{D}} f n_i \,\mathrm{d}a$$

für jede Funktion $f \in C^1(\overline{\mathcal{D}})$.
Oft erscheint der Gaußsche Integralsatz auch in folgender Form:

$$\int_{\mathcal{D}} \operatorname{div} F \,\mathrm{d}x = \int_{\partial\mathcal{D}} F \cdot n \,\mathrm{d}a$$

wobei $F : \mathcal{D} \to \mathbb{R}^d$ ein Vektorfeld ist. Die Komponentenfunktionen von $F = (F_1, \ldots, F_d)$ sollen dann $F_i \in C^1(\overline{\mathcal{D}})$ für $i = 1, \ldots, n$ erfüllen.

Folgerung 2.3. (mehrdimensionale partielle Integration)
Sei $u \in C^1(\mathbb{R}^d, \mathbb{R})$ und $\vec{v} : \mathcal{D} \to \mathbb{R}^d$ ein stetig partiell differenzierbares Vektorfeld. Dann gilt:

$$\int_{\mathcal{D}} u \operatorname{div}(\vec{v}) \,\mathrm{d}x = \int_{\partial\mathcal{D}} u\vec{v} \cdot n \,\mathrm{d}a - \int_{\mathcal{D}} \vec{v} \cdot \nabla\phi \,\mathrm{d}x$$

Definition 2.4. (schwache Ableitung)
Sei $u \in L^1_{\text{loc}}(\mathcal{D})$. Dabei bezeichne $L^1_{\text{loc}}(\mathcal{D})$ den Raum der lokal integrierbaren Funktionen auf $\mathcal{D}$. Wir sagen v besitzt eine schwache Ableitung zum Multiindex α, falls eine Funktion $v \in L^1_{\text{loc}}$ existiert, mit

$$\int_{\mathcal{D}} u \partial^\alpha \Phi \, \mathrm{d}x = (-1)^{|\alpha|} \int_{\mathcal{D}} v \Phi \, \mathrm{d}x \qquad \forall \Phi \in C_0^\infty(\mathcal{D})$$

In diesem Zusammenhang nennen wir Φ auch Testfunktion und wir definieren $D^\alpha u \coloneqq v$ als die schwache Ableitung von u zum Multiindex α.

Bemerkung. Per Konvention ist für $\alpha = (0, \ldots, 0) \quad \partial^\alpha u = u$

Definition 2.5. (Sobolevräume)
Sei $\mathcal{D} \subseteq \mathbb{R}^d$ offen, $k \in \mathbb{N}$ und $1 \leq p \leq \infty$. Weiter sei $L : C^1(\mathcal{D}, \mathbb{R}^m) \to L^\infty(\mathcal{D}, \mathbb{R}^k)$ ein linearer Differentialoperator erster Ordnung und $L^\star : C^1(\mathcal{D}, \mathbb{R}^k) \to L^\infty(\mathcal{D}, \mathbb{R}^m)$ der zugehörige adjungierte Operator. Es gelte also

$$\int_{\mathcal{D}} Lu \cdot \phi \, \mathrm{d}x = \int_{\mathcal{D}} u \cdot L^\star \phi \, \mathrm{d}x \text{ für } u \in C_c^1(\mathcal{D}, \mathbb{R}^m) \ , \ \phi \in C_c^1(\mathcal{D}, \mathbb{R}^k)$$

Dann sind:

(a) $W^{k,p}(\mathcal{D}) \coloneqq \{u \in L^p(\mathcal{D})$ und die schwachen Ableitungen $\partial^\alpha u$ existieren, mit $\partial^\alpha u \in L^p(\mathcal{D})$ für alle $\alpha \in \mathbb{N}_0^d, |\alpha| \leq k\}$

(b) $\|u\|_{k,p} = \|u\|_{W^{k,p}(\mathcal{D})} \coloneqq \begin{cases} (\sum\limits_{|\alpha| \leq k} \int_{\mathcal{D}} |\partial^\alpha u|^p \, \mathrm{d}x)^{\frac{1}{p}} & , 1 \leq p < \infty \\ \sum\limits_{|\alpha| \leq k} \|\partial^\alpha u\|_\infty & , p = \infty \end{cases}$

(c) $W_0^{k,p}(\mathcal{D}) \coloneqq \overline{C_c^\infty(\mathcal{D})}^{\|\cdot\|_{k,p}}$. Über den sogenannten Spursatz erhält man eine äquivalente Charakterisierung durch: $W_0^{k,p}(\mathcal{D}) = \{v \in W^{k,p}(\mathcal{D}) : v|_{\partial\mathcal{D}} = 0\}$

(d) Im Falle $p = 2$ schreibt man aufgrund der Tatsache, dass es sich dann bei $W^{k,p}(\mathcal{D})$ um einen Hilbertraum handelt, oft auch $H^k(\mathcal{D}) \coloneqq W^{k,p}(\mathcal{D})$

(e) $H(L, \mathcal{D}) \coloneqq \{u \in L^2(\mathcal{D}, \mathbb{R}^m) : \exists v \in L^2(\mathbb{D}, R^k)$ mit $\int_{\mathcal{D}} v \cdot \phi \, \mathrm{d}x = \int_{\mathcal{D}} u \cdot L^\star \phi \, \mathrm{d}x \ \forall \phi \in C_c^1(\mathcal{D}, \mathbb{R}^k)\}$

Satz 2.6. (Multiplikation mit Testfunktionen und Integration)
Sei $\mathcal{D} \subset \mathbb{R}^d$ offen und $u \in L^1_{\text{loc}}(\mathcal{D})$ und $\int_{\mathcal{D}} u\psi \, \mathrm{d}x = 0 \ \forall \psi \in C_c^\infty(\mathcal{D})$,
dann gilt $u \equiv 0$.

2.2 Stochastische Grundlagen

An dieser Stelle wollen wir an einige grundlegende Resultate der Wahrscheinlichkeitstheorie erinnern. Außerdem führen wir dabei auch Teile der Notation ein, die wir an späterer Stelle noch brauchen werden. Als Referenzen sind vor allem [5], die Vorlesung Wahrscheinlichkeitstheorie von Herrn Prof. Dr. Henze (SS18) sowie [19] zu nennen.
Sei $\Omega \neq \emptyset$ eine beliebige nichtleere Teilmenge. Einige grundlegende Begriffe der Maßtheorie wollen wir an dieser Stelle voraussetzen, sie sind aber ebenfalls in [19] zu finden. Dazu zählen:

- eine σ-Algebra $\mathcal{A} \subset \mathcal{P}(\Omega)$
- die von einem Mengensystem $\mathcal{M} \subset \mathcal{P}(\Omega)$ erzeugte σ-Algebra $\sigma(\mathcal{M})$
- ein Maß μ auf einer σ-Algebra $\mathcal{A}$
- das Maß-Integral einer messbaren Funktion $f : \Omega \to \overline{\mathbb{R}}$ über einem Maßraum $(\Omega, \mathcal{A}, \mu)$
- die Borel'sche σ-Algebra $\mathcal{B}$, sowie die Begriffe 'Borelmenge' und 'Borel-messbar'.

Definition 2.7. (Wahrscheinlichkeitsraum)
Ein Wahrscheinlichkeitsraum ist ein Tripel $(\Omega, \mathcal{A}, \mathbb{P})$. Dabei seien:

(a) Ω eine beliebige nichtleere Teilmenge

(b) $\mathcal{A}$ eine σ-Algebra über Ω

(c) $\mathbb{P} : \mathcal{A} \to \mathbb{R}$ eine Funktion mit:
   (i) $\mathbb{P}(A) \geq 0$ für jedes $A \in \mathcal{A}$
   (ii) $\mathbb{P}(\Omega) = 1$
   (iii) Sind $A_1, A_2, \ldots \in \mathcal{A}$ paarweise disjunkt , dann gilt $\mathbb{P}(\sum_{j=1}^{\infty} A_j) = \sum_{j=1}^{\infty} \mathbb{P}(A_j)$

Insbesondere erfüllt $\mathbb{P}$ auch die Bedingungen eines Maßes. Somit sind Wahrscheinlichkeitsräume Spezialfälle eines Maßraumes. Jede Menge $A \in \mathcal{A}$ heißt dann auch Ereignis, zu $\mathbb{P}$ sagen wir Wahrscheinlichkeitsmaß und wir nennen $\mathbb{P}(A)$ die Wahrscheinlichkeit des Ereignisses A. Ein Tupel $(\Omega, \mathcal{A})$ heißt Messraum oder auch messbarer Raum.

Definition 2.8. (Zufallsvariable und deren Verteilung.)

(a) Seien $(\Omega, \mathcal{A})$ und $(\Omega', \mathcal{A}')$ Messräume. Eine (Ω'-wertige) Zufallsvariable ist ein eine $(\mathcal{A}, \mathcal{A}')$-messbare Funktion $X : \Omega \to \Omega'$, d.h. es gilt: $X^{-1}(A') \in \mathcal{A} \quad \forall A' \in \mathcal{A}'$.
Der Wert $X(\omega)$ heißt auch Realisierung der Zufallsvariablen X zum Ausgang $\omega \in \Omega$

(b) Sei in obiger Situation zusätzlich $(\Omega, \mathcal{A})$ ausgerüstet mit Wahrscheinlichkeitsmaß $\mathbb{P}$, also ein Wahrscheinlichkeitsraum. Dann ist durch

$$\begin{aligned} \mathbb{P}^X : \mathcal{A}' &\to [0,1] \\ A' &\mapsto \mathbb{P}(X^{-1}(A')), \quad A' \in \mathcal{A}' \end{aligned}$$

ein Maß auf $\mathcal{A}'$ definiert. $\mathbb{P}^X$ heißt dann die Verteilung von X.

Sei ab nun $(\Omega, \mathcal{A}, \mathbb{P})$ stets ein Wahrscheinlichkeitsraum und $\mathcal{B}^n$ die Borelsche σ-Algebra über $\mathbb{R}^n$.

Definition 2.9. (stetig verteilte Zufallsvariablen und Zufallsvektoren)
Ein Zufallsvektor $X = (X_1, \ldots, X_n)$ heißt stetig verteilt, wenn eine nichtnegative (Borel-)messbare Funktion $f : \mathbb{R}^n \to \mathbb{R}$ mit $\int_{\mathbb{R}^n} f(x)\,\mathrm{d}x = 1$ existiert, sodass

$$\mathbb{P}^X(B) = \mathbb{P}(X \in B) = \int_B f(x)\,\mathrm{d}x, \quad B \in \mathcal{B}^n$$

In diesem Fall heißt f Dichte von X (bzw. von $\mathbb{P}^X$).
Ist $n = 1$ spricht man einfach von einer stetig verteilten Zufallsvariable mit Dichte f.

Definition 2.10. (Unabhängigkeit)
Sei $\mathcal{J}$ eine Menge mit mindestens zwei Elementen.

(a) Es seien $A_j \in \mathcal{A}$ für $j \in \mathcal{J}$ Ereignisse. Die Familie $(A_j)_{j\in\mathcal{J}}$ heißt unabhängig, falls gilt:

$$\mathbb{P}\left(\bigcap_{j\in J} A_j\right) = \prod_{j\in J} \mathbb{P}(A_j) \;\forall J \subset \mathcal{J} \text{ mit } 2 \le |J| \le \infty \tag{2.1}$$

(b) Seien $\mathcal{M}_j \subset \mathcal{A}$ für $j \in \mathcal{J}$ Mengensysteme. Die Familie $(\mathcal{M}_j)_{j\in\mathcal{J}}$ von Mengensystemen heißt unabhängig, falls Bedingung (2.1) für jede endliche mindestens zweielementige Teilmenge $J \subset \mathcal{J}$ und jede Wahl $A_j \in \mathcal{M}_j,\ j \in J$ erfüllt ist.

(c) Seien $(\Omega_j, \mathcal{A}_j)_{j\in\mathcal{J}}$ messbare Räume und $X_j : \Omega \to \Omega_j$ für $j \in \mathcal{J}$ Zufallsvariablen. Die Familie $(X_j)_{j\in\mathcal{J}}$ heißt unabhängig, falls die Familie der erzeugten σ-Algebren

$$(\sigma(X_j))_{j\in\mathcal{J}} \coloneqq \sigma\left(\bigcup_{j\in\mathcal{J}} X_j^{-1}(\mathcal{A}_j)\right)$$

unabhängig ist.

Definition 2.11. (Erwartungswert)
$X : \Omega \to \overline{\mathbb{R}}$ sei eine Zufallsvariable. Der Erwartungswert von X existiert genau dann,

wenn $\int_\Omega |X| \, \mathrm{d}\mathbb{P} < \infty$. In diesem Fall heißt

$$\mathbb{E}[X] \coloneqq \int_\Omega X \, \mathrm{d}\mathbb{P}$$

der Erwartungswert von X. Ist X eine stetig verteilte Zufallsvariable mit Dichte f, so gilt:

$$\int_\Omega X \, \mathrm{d}\mathbb{P} = \int_{\mathbb{R}} x f(x) \, \mathrm{d}x$$

Definition 2.12. (Normalverteilung und multivariate Normalverteilung)

(a) Eine Zufallsvariable X heißt normalverteilt mit Parametern μ und σ^2, falls X die Dichte

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad x \in \mathbb{R}$$

besitzt. In diesem Fall schreiben wir oft auch $X \sim N(\mu, \sigma^2)$. Ist spezieller $\mu = 0$ und $\sigma^2 = 1$ so heißt X standardnormalverteilt.

(b) Sei nun $X = (X_1, \ldots, X_n)$ ein Zufallsvektor, $\mu = (\mu_1, \ldots, \mu_n) \in \mathbb{R}^n$ und $C = (\sigma_{ij})_{1 \le i,j \le n} \in \mathbb{R}^{n \times n}$ eine symmetrische positiv-definite Matrix. X bestitzt eine (nicht ausgeartete) multivariate Normalverteilung mit Parametern μ und C, falls die Dichte von X durch

$$f(x) = \frac{1}{(2\pi)^{\frac{k}{2}}\sqrt{\det C}} \exp\left(-\frac{1}{2}(x-\mu)^\top C^{-1}(x-\mu)\right), \quad x \in \mathbb{R}^n$$

gegeben ist. Wir schreiben auch $X \sim N_n(\mu, C)$. Insbesondere kann man mit recht elementaren Methoden einsehen, dass dann auch für jedes $j \in \{1, \ldots, n\}$ $X_j \sim N(\mu_j, \sigma_{jj})$ und außerdem die Einträge σ_{ij} der Matrix C gerade die Kovarianzen $\mathrm{Cov}(X_i, X_j)$ darstellen. Ein Beweis hierfür findet sich im Appendix.

Satz 2.13. (schwaches Gesetz großer Zahlen - wird eventuell gekürzt)
Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge unabhängiger reellwertiger Zufallsvariablen auf $(\Omega, \mathcal{A}, \mathbb{P})$ mit identischer Verteilung $\mathbb{P}^{X_1} = \mathbb{P}^{X_n}$ $\forall n \in \mathbb{N}$. Wir nennen (X_n) dann auch eine u.i.v-Folge, dabei steht u.i.v. für 'unabhängig identisch verteilt'. Ist zudem $\mathbb{E}[X_1^2] < \infty$ so gilt:

$$\frac{1}{n}\sum_{j=1}^{n} X_j \xrightarrow{\mathbb{P}} \mathbb{E}[X_1]$$

Mit $\xrightarrow{\mathbb{P}}$ bezeichnen wir dabei die Konvergenz bezüglich des Wahrscheinlichkeitsmaßes $\mathbb{P}$. Es gilt also $\lim_{n\to\infty} \mathbb{P}(|\frac{1}{n}\sum_{j=1}^{n} X_j - \mathbb{E}[X_1]| > \epsilon) = 0$. Wir sagen auch $\frac{1}{n}\sum_{j=1}^{n} X_j$ konvergiert stochastisch gegen $\mathbb{E}[X_1]$. Im Falle eines Zufallsvektors (einer $\mathbb{R}^d$-wertigen Zufallsvariable) kann der Betrag gegen eine beliebige Norm auf $\mathbb{R}^d$ ersetzt werden.

Satz 2.14. (starkes Gesetz großer Zahlen)
Es sei $(X_n)_{n\in\mathbb{N}}$ eine u.i.v.-Folge und es gelte $\mathbb{E}[|X_1|] < \infty$. Dann gilt für fast alle $\omega \in \Omega$

$$\mathbb{E}[X_1] = \lim_{n\to\infty} \sum_{i=1}^{n} X_i(\omega)$$

das heißt es existiert eine Menge $N \subset \Omega$ mit $\mathbb{P}(N) = 0$ und obige Aussage gilt für alle $\omega \notin N$. Eine solche Menge N heißt auch Nullmenge. In der Literatur findet man diese Art der Konvergenz oft auch unter dem Namen der ($\mathbb{P}$)-fast sicheren Konvergenz.

Bemerkung. (wird gekürzt wenn schwaches Gesetz großer Zahlen gekürzt)
Aus Konvergenz für fast alle $\omega \in \Omega$ folgt insbesondere Konvergenz bezüglich des Wahrscheinlichkeitsmaßes $\mathbb{P}$.
Denn falls $\mathbb{E}[X] = \lim_{n\to\infty} \sum_{i=1}^{n} X_i(\omega)$ für fast alle $\omega \in \Omega$, dann ist das gleichbedeutend mit $\mathbb{P}(\{\omega \in \Omega : \lim_{n\to\infty} \sum_{i=1}^{n} X_i(\omega) \neq \mathbb{E}[X]\}) = 0$ und somit insbesondere $\frac{1}{n}\sum_{j=1}^{n} X_j \xrightarrow{\mathbb{P}} \mathbb{E}[X_1]$.

Satz 2.15. (Chebyshev Ungleichung)
Sei X eine Zufallsvariable mit $\mathbb{E}[X^2] < \infty$, dann gilt für jedes $\epsilon > 0$:

$$\mathbb{P}\left(|X - \mathbb{E}[X]| \geq \epsilon\right) \leq \frac{\mathbb{V}[X]}{\epsilon^2}$$

Satz 2.16. (Zentraler Grenzwertsatz)
Sei $(X_n)_{n\in\mathbb{N}}$ eine u.i.v.-Folge und es gelte $\mathbb{E}[X_1^2] < \infty$. Mit $\mathbb{V}[X_1]$ bezeichnen wir die Varianz der Zufallsvariable X_1:

$$\mathbb{V}[X_1] = \mathbb{E}[X_1^2] - \mathbb{E}[X_1]^2 = \mathbb{E}[(X_1 - \mathbb{E}[X_1])^2]$$

Dann gilt für eine standardnormalverteilte Zufallsvariable N:

$$\hat{S}_n := \frac{\sum_{j=1}^{n} X_j - n\mathbb{E}[X_1]}{\sqrt{n\mathbb{V}[X_1]}} \xrightarrow{\mathcal{D}} N$$

Dabei bezeichnet $\xrightarrow{\mathcal{D}}$ die Konvergenz in Verteilung und ist genau dann erfüllt, wenn

$$\lim_{n\to\infty} \mathbb{P}^{\hat{S}_n}((-\infty, x]) = \mathbb{P}^N((-\infty, x])$$

für alle Stetigkeitsstellen der Verteilungsfunktion $\mathbb{P}^N((-\infty, \cdot])$ von N erfüllt ist.

Definition 2.17. (Zufallsfelder)

(a) Sei $\mathcal{D} \subset \mathbb{R}^d$ eine nichtleere Menge und $d \geq 1$. Wir nennen $X : \Omega \times \mathcal{D} \to \mathbb{R}$ ein Zufallsfeld, wenn für jedes feste $x \in \mathcal{D}$ die Funktion $X(x) := X(\cdot, x)$ eine Zufallsvariable ist.
Ist spezieller $d = 1$ so wird die Parametermenge auch oft mit T bezeichnet und man

spricht von einem stochastischen Prozess. In der Literatur findet man Zufallsfelder vor allem unter der englischen Bezeichnung 'random fields'.
Außerdem definieren wir die Erwartung eines Zufallsfeldes durch

$$\mu(x) \coloneqq \mathbb{E}[X(x)] = \int_\Omega X(w,x)\,\mathrm{d}\mathbb{P}(\omega) \quad \text{für } x \in \mathcal{D}$$

und die zu einem Zufallsfeld gehörige Kovarianzfunktion $C : \mathcal{D} \times \mathcal{D} \to \mathbb{R}$ durch

$$C(x,y) \coloneqq \mathrm{Cov}[X(x), X(y)] = \mathbb{E}\left[(X(x) - \mathbb{E}[X(x)])\,(X(y) - \mathbb{E}[X(y)])\right]$$

(b) Ein Zufallsfeld X heißt weiter Gauß'sches Zufallsfeld falls für jede Wahl $n \in \mathbb{N}$ und $x = (x_1, \ldots, x_n) \in \mathcal{D}$ der Zufallsvektor $\hat{X} = (\hat{X}_1, \ldots, \hat{X}_n)^\top = (X(x_1), \ldots, X(x_n))$ eine (nicht ausgeartete) multivariate Normalverteilung mit Parametern $\mu(x)$ und $C(x,y)$ besitzt. Wir haben in 2.12 bereits gesehen, dass wir so gerade den Erwartungswertvektor μ und die Kovarianzmatrix C festlegen.

(c) Ein Zufallsfeld $Y : \Omega \times \mathcal{D} \to \mathbb{R}$ heißt hingegen log-normal verteilt oder kurz lognormal-Feld falls das durch

$$\begin{aligned} X : \Omega \times \mathcal{D} &\to \mathbb{R} \\ (\omega, x) &\mapsto \log(Y(\omega, x)) \end{aligned}$$

definierte Feld ein Gauß'sches Zufallsfeld ist. Umgekehrt lässt sich jedes lognormal-Feld also durch ein Gauß'sches Zufallsfeld X erzeugen. So ist dann $\tilde{Y} : \Omega \times \mathcal{D} \to \mathbb{R}$, $(\omega, x) \mapsto \exp(X(\omega, x))$ auch ein lognormal-Feld.

3 Die Monte Carlo Methode

3.1 Herleitung und Beispiel

Wie in [22] wollen wir uns, um die Monte Carlo Methode von Grund auf einzuführen, zunächst mit der numerischen Integration beschäftigen. Grundsätzlich handelt es sich bei der Monte Carlo Methode um einen sogenannten Erwartungswertschätzer. Bevor wir also ein Problem mithilfe der Monte Carlo Methode lösen können, müssen wir die Größe, welche wir berechnen wollen, zunächst in der Form eines Erwartungswertes ausdrücken. Wir suchen dann also einen Erwartungswert $\mathbb{E}[X]$ wobei X eine Zufallsvariable, einen Zufallsvektor oder gar ein Zufallsfeld beschreiben kann. Mithilfe der Monte-Carlo-Methode können wir dann versuchen eben diesen Erwartungswert zu schätzen. Dazu müssen wir X simulieren können. Damit ist gemeint, dass wir in der Lage sein müssen eine Realisierung $(x_1, \ldots, x_n)$ von $(X_1, \ldots, X_n)$ zu generieren (oft sagt man auch in Anlehnung an das Bernoulli'sche Urnenmodell 'zu ziehen'). Dabei sollen die Zufallsgrößen $X_1, \ldots, X_n$ unabhängig sein und die gleiche Verteilung besitzen wie die Zufallsgröße X. Außerdem sei vorausgesetzt dass der Erwartungswert $\mathbb{E}[X] < \infty$ existiert. Anschließend wird der gesuchte Erwartungswert durch

$$\mathbb{E}[X] \approx \frac{1}{n}(x_1 + \ldots x_n)$$

approximiert.

Beispiel 3.1. (Integral über $[0,1]^d$ - aus [22])
Angenommen wir wollen für $d \geq 1$ folgendes Integral berechnen:

$$I = \int_{[0,1]^d} f(u_1, \ldots, u_d)\, \mathrm{d}u_1 \ldots\, \mathrm{d}u_d$$

Wir können das Integral dann wie folgt als Erwartungswert ausdrücken: Sei $X = f(U_1, \ldots, U_d)$ ein Zufallsvektor, wobei $U_1, \ldots, U_d$ unabhängig und auf $[0,1]$ gleichverteilt sind, d.h. jedes U_i besitzt als Dichte $f_i(x) = \mathbb{1}_{[0,1]}(x)$. Dann ergibt sich so

$$I = \int_{[0,1]^d} f(u_1, \ldots, u_d)\, \mathrm{d}u_1 \ldots\, \mathrm{d}u_d = \mathbb{E}[f(U_1, \ldots, U_d)] = \mathbb{E}[X]$$

Wir haben also das Integral, welches wir berechnen wollen als Erwartungswert ausgedrückt. Nun müssen wir die Zufallsvariable $X = f(U_1, \ldots, U_d)$ simulieren. Dazu nehmen wir an, gleichverteilte Zufallsvariablen simulieren zu können. Die Simulation solcher Zufallsvariablen spielt in der numerischen Stochastik eine ganz besondere Rolle, denn oft werden andere Verteilungen durch Transformationen auf den Fall einer Gleichverteilung auf $[0,1]$ reduziert. Sei also $(U_i)_{i\geq 1}$ eine Folge unabhängiger Zufallsvariablen mit Gleichverteilung auf $[0,1]$. Wir können dann mithilfe der simulierten Realisierungen $(u_i)_{i\geq 1}$ von

$(U_i)_{i\geq 1}$ die Zufallsvariable X wie folgt definieren: Wir setzen

$$\begin{aligned} X_1 &= f(U_1, \ldots, U_d), & x_1 &= f(u_1, \ldots, u_d) \\ X_2 &= f(U_{d+1}, \ldots, U_{2d}), & x_2 &= f(u_{d+1}, \ldots, u_{2d}) \\ X_i &= f(U_{(i-1)d+1}, \ldots, U_{id}), & x_2 &= f(u_{(i-1)d+1}, \ldots, u_{id}) \end{aligned}$$

Da $(U_i)_{i\geq 1}$ eine Folge unabhängiger Zufallsvariablen ist, erhalten wir so unter der einzigen echten Voraussetzung, dass f messbar ist, nach Blockungslemma ebenfalls eine Folge unabhängiger Zufallsvariablen $(X_i)_{i\geq 1}$. Außerdem erhalten wir so für ein großes $n \in \mathbb{N}$ eine gute Approximation von I durch:

$$I = \mathbb{E}[X] \approx \frac{1}{n}(x_1 + \cdots + x_n) = \frac{1}{n}(f(u_1, \ldots, u_d) + \cdots + f(u_{(n-1)d+1}, \ldots, u_{nd}))$$

Inbesondere haben wir keinerlei Regularität an f vorausgesetzt, es genügt bereits die bloße Messbarkeit von f

Oft wollen wir über eine andere Grundmenge als $[0,1]^d$ integrieren. Bei endlichen Mengen, etwa einer beschränkten Borelmenge $B \subset \mathbb{R}^d$ mit $0 < |B| \coloneqq \lambda^d(B)$ (hierbei ist $\lambda^d(\cdot)$ das Borel-Lebesgue-Maß) lässt sich $I = \int_B f(x)\,\mathrm{d}x$ ähnlich wie in obigem Beispiel berechnen. Für einen Zufallsvektor U mit Gleichverteilung $U(B)$ auf B existiert nämlich der Erwartungswert $f(U)$ und es gilt:

$$\mathbb{E}[f(U)] = \int_B f(x)\frac{1}{|B|}\,\mathrm{d}x = \frac{I}{|B|}$$

Wieder simulieren wir $(U_i)_{i\geq 1}$ als Folge unabhängiger Zufallsvariablen mit identischer Verteilung zu U. Dann erhalten wir:

$$I = |B| \cdot \mathbb{E}[f(U)] \approx \frac{|B|}{n}\sum_{j=1}^{n} f(u_j)$$

Wollen wir hingegen ein Integral über $\mathbb{R}^d$ auswerten, muss es uns in der Form

$$I = \int_{\mathbb{R}^d} g(x)f(x)\,\mathrm{d}x = \int_{\mathbb{R}^d} g(x_1, \ldots, x_d)f(x_1, \ldots, x_d)$$

vorliegen. Dabei sei $f(x)$ nichtnegativ und $\int_{\mathbb{R}^d} f(x)\,\mathrm{d}x = 1$. Dann lässt sich I schreiben als $I = \mathbb{E}[g(X)]$ für eine Zufallsvariable X mit Werten in $\mathbb{R}^d$ und Verteilung $f(x)\,\mathrm{d}x$. Wir können also I approximieren durch

$$I \approx \frac{1}{n}\sum_{i=1}^{n} g(x_i)$$

wobei $(x_i)_{i\geq 1}$ Realisierungen der Zufallsvariablen $(X_i)_{i\geq 1}$ sind, welche unabhängig und identisch zu X verteilt seien.

Betrachten wir nun wieder die Monte Carlo Methode in einem etwas abstrakteren Sinne ganz allgmein. An der Stelle, an der wir letztlich die Realisierungen einer Zufallsvariable eingesetzt haben, also einen Erwartungswert durch $\mathbb{E}[X] \approx \frac{1}{n}(x_1 + \dots x_n)$ approximiert haben, haben wir stet gefordert, dass n groß ist. Es stellt sich nun die Frage, wann n groß genug ist. Wir wollen uns deshalb noch abschließend damit beschäftigen, wann und wie die Methode konvergiert und was wir über die Genauigkeit der Approximation aussagen können.

3.2 Konvergenz und Genauigkeit

Damit die Methode überhaupt in irgendeiner Weise als nützlich zu erachten ist, bedarf es Möglichkeiten den Fehler

$$\epsilon_n = \frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}[X]$$

abzuschätzen. Um diesem Problem beizukommen, bedienen wir uns zweier zentraler Aussagen der Wahrscheinlichkeitstheorie. Zum einen sagt uns das starke Gesetzt großer Zahlen 2.14, dass unter der Voraussetzung $\mathbb{E}[|X|] < \infty$ der Fehler ϵ_n für $n \to \infty$ für fast alle $\omega \in \Omega$ gegen 0 konvergiert. Wir erhalten also zunächst Konvergenz der Methode in einem sehr grundlegendem Sinn. Aus dem zentralen Grenzwertsatz 2.16 lassen sich zum anderen Aussagen über die Genauigkeit der Methode und letztlich somit auch der Art der Konvergenz ableiten. Nach 2.16 erhalten wir nämlich für eine u.i.v.-Folge $(X_i)_{i\in\mathbb{N}}$ mit gleicher Verteilung wie X und $\mathbb{E}[X^2] < \infty$, dass

$$\frac{\sqrt{n}}{\sqrt{\mathbb{V}[X]}}\epsilon_n = \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i - \sqrt{n}\mathbb{E}[X]}{\sqrt{\mathbb{V}[X]}} = \frac{\sum_{i=1}^{n} X_i - n\mathbb{E}[X]}{\sqrt{n\mathbb{V}[X]}} =: \hat{S}_n \xrightarrow{\mathcal{D}} N \text{ für } n \to \infty$$

wobei N eine standardnormalverteilte Zufallsvariable ist. Die Wurzel der Varianz wird im Folgenden noch des Öfteren auftauchen, weswegen wir an dieser Stelle die sogenannte Standardabweichung $\sigma := \sqrt{\mathbb{V}[X]}$ einführen. Da also

$$\lim_{n\to\infty} \mathbb{P}^{\hat{S}_n}((-\infty, x]) = \mathbb{P}^N((-\infty, x])$$

gilt, ist insbesondere für $a \leq b$

$$\lim_{n\to\infty} \mathbb{P}(\frac{\sigma}{\sqrt{n}}a \leq \epsilon_n \leq \frac{\sigma}{\sqrt{n}}) = \lim_{n\to\infty} \mathbb{P}(a \leq \hat{S}_n \leq b) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\,\mathrm{d}x\ .$$

An dieser Stellen wollen wir kurz innehalten und uns überlegen was obiges Resultat für den Fehler der Monte Carlo Methode denn praktisch gesehen bedeutet.

- Der zentrale Grenzwertsatz liefert uns kein zu der Folgerung aus dem starken Gesetz großer Zahlen vergleichbares Resultat, denn es ist $\lim_{n\to\infty} \mathbb{P}(\epsilon_n = 0) = 0$ nach obiger Überlegung.

- Der zentrale Grenzwertsatz erlaubt uns ebenso nicht eine für andere Verfahren typische Fehlerschranke der Form $\epsilon_n \leq M_n$ für eine von n und möglicherweise anderen Faktoren, wie z.B. Ausgangsdaten, abhängigen Schranke M aufzustellen. Grund dafür ist, dass der Träger von $\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$ ganz $\mathbb{R}$ ist.

- Was der zentrale Grenzwertsatz uns jedoch erlaubt, ist, ein sogenanntes 95% Konfidenzintervall für ϵ_n zu bestimmen. Das bedeutet, dass das tatsächliche Ergebnis mit einer Wahrscheinlichkeit von mindestens 95% im gegebenen Intervall enthalten ist. Denn, da
$$\mathbb{P}(|N| \leq 1.96) \approx 0.95$$
können wir wegen
$$\lim_{n\to\infty} \mathbb{P}(-1.96\frac{\sigma}{n} \leq \epsilon_n \leq 1.96\frac{\sigma}{n}) \approx 0.95 \quad (\star)$$
ein Konfidenzintervall für $\mathbb{E}[X]$ der Form
$$[m - 1.96\frac{\sigma}{\sqrt{n}}, m + 1.96\frac{\sigma}{\sqrt{n}}]$$
angeben. In der Praxis nehmen wir näherungsweise an, dass $(\star)$ auch für ein festes $n \in \mathbb{N}$ erfüllt ist und entledigen uns so des Grenzwertes. Somit wird dann insbesondere die Wahl $m = \frac{1}{n}\sum_{i=1}^{n}(x_1, \ldots, x_n)$ gerechtfertigt.

Wir erhalten also (unter den eben erklärten Annahmen) eine Konvergenzrate des (wahrscheinlichen) Fehlers von $\frac{\sigma}{\sqrt{n}}$. Dieses Resultat mag auf den ersten Blick relativ ernüchternd wirken, allerdings existieren Fälle, in denen solch eine langsame Methode die bestmögliche ist. [22] nennt hierzu zum Beispiel Integrale in mehr als 100 Dimensionen oder besonders schwere parabolische Differentialgleichungen. Außerdem lohnt es sich zu erwähnen, dass wir im Falle der numerischen Integration - bis auf Integrierbarkeit und Messbarkeit - keine Voraussetzungen an die Regularität der zu Funktion f gestellt haben.
Obiges Resultat legt außerdem nahe, dass es entscheidend für eine Aussage über die Konvergenz und Güte der Methode ist, die Standardabweichung σ zu kennen, oder zumindest über einen guten Schätzer für σ zu verfügen. Falls uns σ bzw. $\mathbb{V}$ nämlich sogar exakt bekannt ist, können wir die sogenannte Chebyshev Ungleichung 2.15 ausnutzen: Da $(X_i)_{i\in\mathbb{N}}$ eine u.i.v.-Folge mit Verteilung wie X ist, gilt nämlich mit den üblichen Rechenregeln für die Varianz (zu finden z.B. in [5] auf den Seiten 778 und 779)
$$\mathbb{V}[\frac{1}{n}\sum_{i=1}^{n} X_i] = \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}[X] = \frac{\mathbb{V}[X]}{n}$$

Dann besagt die Chebychev Ungleichung für alle $t \geq 0$:

$$\mathbb{P}\left(|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}[X]| \geq t\right) \leq \frac{\mathbb{V}[X]}{nt^2}$$

Für uns bedeutet das insbesondere, dass für jedes $\epsilon \in (0, 1]$ die berechnete Monte-Carlo Approximation $\frac{1}{n}\sum_{i=1}^{N}$ mit einer Wahrscheinlichkeit von $1 - \epsilon$ weniger als $(\frac{\mathbb{V}[X]}{n\epsilon})^{\frac{1}{2}}$ von dem tatsächlichen Erwartungswert $\mathbb{E}[X]$ entfernt ist. In der Literatur (z.B. in [28]) finden sich einige Weiterentwicklungen der Monte Carlo Methode. Abgesehen von der Multilevel Monte Carlo Methode, welche wir in Abschnitt 4 behandeln werden, wollen wir uns hier auf die oben erklärte Standard-Variante beschränken.

4 Die Multilevel Monte Carlo Methode

4.1 Motivation und Beispiel

Nachdem wir im dritten Abschnitt die Monte Carlo Methode betrachtet haben, wollen wir uns nun einer Weiterentwicklung der Monte Carlo Methode, der sogenannten Multi Level Monte Carlo Methode zuwenden. Grundsätzlich liegt dieselbe Situation vor wie bei der Monte Carlo Methode: Wir wollen wieder eine Größe bestimmen, welche sich nach geeigneter Modellierung in der Form eines Erwartungswertes $\mathbb{E}[X]$ einer Zufallsvariablen X schreiben lässt. Besonders wenn diese Größe mit der Lösung von gewöhnlichen oder partiellen Differentialgleichungen, wie wir sie später betrachten wollen, hat man nun jedoch die Wahl, wie genau die numerische Lösung des zugrunde liegenden Problems, z.B. der Differentialgleichung, erfolgen soll. Beispielsweise können wir im Falle der numerischen Lösung von Differentialgleichung Zeitschrittweiten und/oder Gitterweiten der Ortsdiskretisierung festlegen. Wir werden dann in diesem Zusammenhang auch von verschiedenen (Genauigkeits-)Leveln sprechen. An dieser Stelle tritt stets ein typischer Zwiespalt auf:

- Zu Einen wollen wir möglichst genau rechnen. Dies legt die Wahl von besonders kleinen Zeitschrittweiten bzw. feinen Gittern zur Ortsdiskretisierung nahe.
- Zum Anderen wollen wir die Anzahl der Rechenschritte bzw. die Rechenzeit möglichst gering halten. Dies spricht hingegen für große Zeitschritte bzw. grobe Gitter.

Zusätzlich zu der oft bereits alleine anspruchsvollen Aufgabe solche Probleme numerisch zu lösen, müssen wir also stets einen für unsere Bedürfnisse passenden Kompromiss aus möglichst genauer numerischer Approximation und geringem (oder zumindest machbarem) Rechenaufwand eingehen. Obwohl dies zunächst wie eine zusätzliche Hürde erscheint und Mehraufwand vermuten lässt, stellt sich heraus, dass solche eine Wahl der Genauigkeit im Kontext von Monte Carlo Methode sich durchaus als Nützlich erweisen kann. Die Multi Level Monte Carlo Methode, wir werden im folgenden auch oft vom sogenannten Multi Level Monte Carlo Schätzer sprechen, ist der Prototyp einer Familie sogenannter Varianz-reduzierender Methoden, welche das Ziel haben die naive Monte Carlo Methode in Sachen Konvergenzrate und Effizienz zu schlagen. Bevor wir erklären wie genau die Multilevel Monte Carlo Methode im Allgemeinen dabei vorgeht, möchten wir die Funktionsweise wieder anhand eines Beispiels erklären, welches in [18] ausführlich erklärt wird.

Beispiel 4.1. (Wieder ein Integral über $[0,1]^d$)
Wie bereits im letzten Abschnitt setzen wir uns die Aufgabe das Integral einer Funktion f zunächst über $[0,1]^d$ zu bestimmen. Damit wir aber überhaupt in oben erklärte Situation kommen und von verschiedenen 'Leveln' sprechen können, sei f nun zusätzlich abhängig von einem Parameter $\lambda \in \Lambda \subseteq \mathbb{R}^{d_2}$, also $f : \Lambda \times [0,1]^d \to \mathbb{R}$. Um bei den folgenden Überlegungen die Notation so schlank wie möglich zu halten, betrachten wir an dieser Stelle nur einen konkreten Spezialfall:

Sei $d = d_2 = 1$ und $f \in C([0,1]^2, \mathbb{R})$, d.h. wir wollen das Integral

$$I(\lambda) = \int_0^1 f(\lambda, u)\,\mathrm{d}u$$

für alle $\lambda \in \Lambda = [0,1]$ bestimmen, wir suchen also nach einer Funktion in Abhängigkeit von λ.

Monte Carlo Schätzer für $I(\lambda)$
Wollen wir an dieser Stelle einen normalen Monte Carlo Schätzer nutzen, stellt sich die Frage, wie wir mit dem zusätzlichen Parameter umsetzen sollen. Die wohl naheliegendste und einfachste Idee ist, zunächst für ein festes $h \in \mathbb{N}$ ein Gitter $\{\lambda_i = \frac{i}{h}, i = 0, \dots, h\}$ festzulegen und für jedes λ_i wie im letzten Abschnitt vorzugehen und für ein $n \in \mathbb{N}$

$$I(\lambda_i) \approx \hat{I}(\lambda_i) \coloneqq \frac{1}{n}\sum_{k=1}^{n} f(\lambda_i, x_k)$$

zu schätzen. Dabei seien wieder $(x_k)_{k=1,\dots,n}$ Realisierungen von unabhängigen auf $[0,1]$ gleichverteilten Zufallsvariablen $(X_k)_{k=1,\dots,n}$. Anschließend lässt sich aus den so ermittelten Werten durch Interpolation einen Schätzer für die gesamte Funktion $I(\lambda)$ bestimmen. Grundsätzlich sind verschiedene Interpolationsansätze möglich. Für dieses grundlegende Beispiel wählen wir stückweise lineare Interpolation. Wir erhalten so für alle $\lambda \in \Lambda$:

$$I(\lambda) \approx (PI)(\lambda) = \sum_{i=0}^{h} \hat{I}(\lambda_i)\varphi_i(\lambda)$$

mit $\varphi_i \coloneqq \mathbb{1}_{\{|\lambda-\lambda_i|\le h\}}(1 - h|\lambda - \lambda_i|)$. Ein solcher Interpolationsansatz lässt sich insbesondere auf mehrdimensionale Gitter übertragen. Somit erhalten wir für $I(\lambda)$:

$$I(\lambda) \approx \mathcal{I}_{MC}(\lambda) \coloneqq \sum_{i=0}^{h}\left(\frac{1}{n}\sum_{k=1}^{n} f(\lambda_i, x_k)\right)\varphi_i(\lambda) = \frac{1}{n}\sum_{k=1}^{n}(Pf(\cdot, x_k))(\lambda)$$

Als Fehler dieser Methode können wir den sogenannten 'root mean square error' verbunden mit einer beliebigen Norm betrachten, wir wählen hierbei die L^2-Norm. Wir erhalten so

$$\epsilon(\mathcal{I}_{MC}) = \left(\mathbb{E}[\|I - \mathcal{I}_{MC}\|^2_{L^2([0,1])}]\right)^{\frac{1}{2}} = \left(\mathbb{E}\left[\int_0^1 |I(\lambda) - \mathcal{I}_{MC}(\lambda)|^2\,\mathrm{d}\lambda\right]\right)^{\frac{1}{2}}$$

Ist f zusätzlich stetig differenzierbar im Parameter λ, kann gezeigt werden, dass

$$\epsilon(\mathcal{I}_{MC}) = \mathcal{O}(n^{-\frac{1}{2}} + h^{-1})\ .$$

. Gleichzeitig ist die Anzahl der arithmetischen Operationen, Funktionsaufrufe und generierter Zufallszahlen in $\mathcal{O}(hn)$. Wir sehen also, dass wir an dieser Stelle genau diesen

Zwiespalt antreffen, welchen wir zuvor abstrakt beschrieben haben. Aus diesem Grund wollen wir nun einen Multilevel Monte Carlo Schätzer für $I(\lambda)$ einführen.

Multilevel Monte Carlo Schätzer für $I(\lambda)$
Wir betrachten nun eine Familie von Gittern $\{\lambda_{li} = \frac{i}{h_l} : h_l = 2^l \ , i = 0, 1, \dots h_l\}$ für $l = 0, \dots, m$. Analog zu oben führen wir zugehörige Interpolationsoperatoren

$$(P_l I)(\lambda) = \sum_{i=0}^{h_l} \hat{I}(\lambda_{li})\varphi_{li} \quad (l = 0, \dots, m)$$

ein. Wir können nun also insbesondere $P \coloneqq P_m$ also Teleskopsumme darstellen. Es gilt nämlich:

$$P = P_m = P_0 + \sum_{l=1}^{m}(P_l - P_{l-1}) \ .$$

Der Monte Carlo Schätzer von oben lässt sich (mit $P_{-1} \coloneqq 0$)dann durch

$$\mathcal{I}_{MC} = \sum_{l=0}^{m} \frac{1}{n} \sum_{k=1}^{n}(P_l - P_{l-1})f(\cdot, x_k)$$

umschreiben. Um nun tatsächlich einen Nutzen aus der Aufteilung in verschiedene Level zu ziehen und einen guten Kompromiss zwischen Kosten und Fehler herzustellen erlauben wir nun zusätzlich die Anzahl der Zufallsauswertungen n von Level zu Level zu variieren. Wir wählen also $(n_l)_{l=0,\dots,m} \in \mathbb{N}^{m+1}$. Außerdem seien $\{X_{lj}, l = 0, \dots, m \ , j = 0, \dots, n_l\}$ unabhängige auf $[0, 1]$ gleichverteilte Zufallsvariablen und $(x_{lj})_{l=0,\dots,m \ ,j=0,\dots,n_l}$ zugehörige Realisierungen. Dann erhalten wir den Multilevel Monte Carlo Schätzer

$$I(\lambda) \approx \mathcal{I}_{MLMC}(\lambda) = \sum_{l=0}^{m} \frac{1}{n_l} \sum_{k=1}^{n_l}((P_l - P_{l-1})f(\cdot, x_{lj}))(\lambda) \ .$$

Der bedeutendste Schritt ist an dieser Stelle eine passende Wahl der n_l. Bei diesem Beispiel wollen wir uns darauf beschränken eine passende Wahl anzugeben und den Nutzen hervorzuheben, welchen wir durch diese Wahl erlangen. So zeigt sich, dass eine passende Wahl beispielsweise durch $n_l = \Theta(2^{-\frac{3l}{2}}n)$ für ein $n \in \mathbb{N}$ groß genug gegeben ist. Dann kann für den analog wie für den MC-Schätzer definierten (RMSE-)Fehler gezeigt werden, dass

$$\epsilon(\mathcal{I}_{MLMC}) = \mathcal{O}(n^{-\frac{1}{2}} + n^{-\frac{1}{2}}) = \mathcal{O}(n^{-\frac{1}{2}})$$

. Zugleich ist die Anzahl der benötigten Rechenoperationen inklusive Funktions- und Zufallszahlauswertungen diesmal in $\mathcal{O}(n)$. Verglichen mit der Standard (Ein-Level) Monte Carlo Methode können wir nun also eine Approximation für die gesamte Familie von Integralen $I(\lambda)$ mit einem Fehler von $\mathcal{O}(n^{-\frac{1}{2}})$, aber den Kosten von $\mathcal{O}(n)$ berechnen. Das ist durchaus erstaunlich, denn bereits die Kosten der Auswertung eines einzigen Integrals $I(\lambda)$ für ein festes $\lambda \in \Lambda$ liegen in $\mathcal{O}(n)$.

Wir sehen also, dass die Multilevel Monte Carlo Methode in Situationen, in denen wir bei der Wahl von Zeitschrittweiten und/oder feinen Gittern zur Ortsdiskretisierung zwischen Anzahl an Rechenoperationen und Genauigkeit einen Kompromiss finden müssen, einen Ein-Level Ansatz, wie die Standard Monte Carlo Methode, durchaus übertreffen kann. Der Kern dieser Methode bildet dabei eine geschickte Wahl der Anzahl n_l der Zufallssamples, welche wir auf je einem Level auswerten. Wie wir in unserem Fall diese Wahl durchführen soll an anderer Stelle in Abschnitt 6 ausführlich erläutert werden, in welchem wir die bisher zunächst beispielhaft anhand der Integration eingeführte Multilevel Monte Carlo Methode auf das probabilistische Transportproblem, welches wir in Abschnitt 5 bereits näher beleuchtet haben, übertragen werden. Mehr zu Monte Carlo und Multilevel Monte Carlo Methoden für Parameterintegrale findet sich neben [18] auch in [17].

4.2 Konvergenz und Genauigkeit

Da wir in Abschnitt 6 noch einmal ausführlich auf die Eigenschaften des Verfahrens für unsere konkrete Anwendung eingehen werden, soll dieser Unterabschnitt eher noch einmal etwas allgemeiner auf die stochastischen Hintergründe eingehen. Als Referenz ist hierbei das Kapitel 9.5 über Monte-Carlo Funktionen in [28] zu nennen. Betrachten wir also wieder etwas allgemeiner eine Folge unabhängig Zufallsvariablen $Y_1, Y_2 \dots$ mit zugehörigen Realisierungen $y_1, y_2, \dots$. Diese sollen dabei alle die identische Verteilung wie eine weitere Zufallsvariable Y mit zugehöriger Dichte g_Y besitzen. Wir wollen diesmal den Erwartungswert einer Zufallsvariablen X berechnen, wobei wir X mithilfe einer messbaren Funktion f, die alle (unten aufgeführten) Voraussetzungen des Transformationssatzes erfülle, folgendermaßen ausgedrückt werden kann:

$$X = f(Y)$$

Wir fordern nun wieder, dass der Erwartungswert $\mathbb{E}[|X|] < \infty$ existiert, und diese Forderung ist für die Konvergenz der Methode ebenso wichtig wie scharf. Wollen wir uns überlegen, was das maßtheoretisch bedeutet, erhalten wir: Ist für $\{g_Y > 0\} \subseteq O$ die Menge O offen und $f : \mathbb{R} \to \mathbb{R}$ eine Borel-messbare Abbildung, deren Restriktion auf O stetig differenzierbar ist, eine nirgends verschwindende Funktionaldeterminante besitze und O bijektiv auf eine Menge $V \subset \mathbb{R}$ abbilde. Dann muss die Dichte g_X der Zufallsvariable X auf V integrierbar sein. Dabei gilt für die Dichte von X nach dem Transformationssatz:

$$g_X(t) := \begin{cases} \frac{g_y(f^{-1}(t))}{|\det f'(f^{-1}(t))|} & , \text{falls } t \in V \\ 0 & , \text{sonst} \end{cases}$$

Wir nehmen nun außerdem an, dass wir über eine Hierarchie $(f_l)_{l \in \{0,\dots,L\}}$ verfügen. Dabei sei $f = f_L$ und wir bezeichnen l als Level-Parameter. Aufgrund der Linearität des Erwartungswertes kann der Erwartungswert von X dann folgendermaßen ausgedrückt

werden:

$$\mathbb{E}[X] = \mathbb{E}[f(Y)] = \mathbb{E}[f_0[Y]] + \sum_{l=1}^{L} \mathbb{E}[f_l(Y) - f_{l-1}(Y)]$$

Wir können nun jeden Summanden einzeln durch einen Monte Carlo Ansatz schätzen. Dazu seien $(Y_{l,i})_{l\in\{0,\dots,L\},i\in\{1,\dots,n_l\}}$ unabhängige Zufallsvariablen aus der Folge $(Y_i)_{i\in\mathbb{N}}$. Dann gilt:

$$\mathbb{E}[X] \approx \frac{1}{n_0}\sum_{i=1}^{n_0} f_0(Y_{0,i}) + \sum_{l=1}^{L}\frac{1}{n_l}\sum_{i=1}^{n_l}(f_l(Y_{l,i}) - f_{l-1}(Y_{l,i}))$$

An dieser Stelle scheint obige Darstellung keinen wirklichen Vorteil gegenüber dem Standard Monte Carlo Schätzer zu besitzen, wir müssen aber nun beachten, dass zum Einen die Anzahl der benötigten Rechenoperation zur Berechnung von $f_l(Y_{l,i})$ unter Umständen für niedrige Werte l deutlich geringer ausfällt, als dies für größere l der Fall ist. Zum Anderen gilt für den Fehler der Monte Carlo Schätzung (vgl. Abschnitt 3), dass der Fehler des l-ten Summanden wie $\sqrt{\frac{\mathbb{V}[f_l(Y)-f_{l-1}(Y)]}{n_l}}$ konvergiert. Das heißt insbesondere, dass falls $\mathbb{V}[f_l(Y)-f_{l-1}(Y)]$ klein ausfällt, auch kleinere n_l gewählt werden können, als bei der Standard Monte Carlo Methode, für welche bei einer großen Varianz $\mathbb{V}[f(Y)] = \mathbb{V}[f_L(Y)]$ ein sehr großes $n = n_L$ für eine gute Approximation benötigt werden. Andererseits müssen wir aber auch beachten, dass wir uns damit zusätzlich zum Schätzfehler, durch die Approximation von f, auch einen Approximationsfehler einhandeln. In der Praxis, wie z.B. beim partiellen Differentialgleichungen, verfügen wir aber auch oft nicht über f selbst, sondern nur verschieden gute Approximationen. Wir müssen also in der späteren Anwendung ganz genau prüfen, wie und ob wir aus der Multilevel Monte Carlo Methode tatsächlich einen Nutzen ziehen können. Diese Grundidee, die Varianz klein zu halten, damit die benötigte Anzahl der auszuwertenden Zufallssamples gering gehalten werden kann, ist namensgebend für die sogenannten 'Varianz reduzierenden Methoden' zur Verbesserung der Monte Carlo Methode. Weitere Details der Konvergenzanalyse finden sich in für unser Problem angepasster Form in Abschnitt 6. Bevor wir dazu kommen können, müssen wir aber im Folgenden Abschnitt zunächst das zu betrachtende Problem sowie zugehörige Löser einführen, welche später die Rolle der Approximationen f_l übernehmen werden.

5 Das lineare Transportproblem

5.1 Problemstellung

5.1.1 Deterministisches Problem

Sei $\mathbb{T} = [0, T]$ ein Zeitintervall für $T > 0$ und $\mathbb{D} \subset \mathbb{R}^d, d \in \mathbb{N}$ ein beschränktes, offenes und konvexes Lipschitz-Gebiet mit Rand $\partial\mathbb{D} = \Gamma_D \dot\cup \Gamma_N$. Wie bereits in der Einleitung beschrieben, wollen wir den Transport eines Stoffes in einer porösen Bodenschicht auf Grundlage eines vorhandenen Flusses beschreiben. Als modellhaftes Problem soll uns hierfür die Regenwasserversickerung dienen: In einer porösen Bodenschicht befindet sich zum Zeitpunkt $t = 0$ ein Stoff (beispielsweise Öl) in einer gegebenen Anfangskonzentration und -verteilung. Nun sickert Regenwasser in diese poröse Bodenschicht ein. Zusätzlich wollen wir weitere Zuflüsse des Fremdstoffes über den Einflussrand $\Gamma_{\text{in}} \subset \partial\mathbb{D}$ zulassen. Wir sind letztendlich an der Konzentration dieser Substanz an einer Stelle $x \in \overline{\mathbb{D}}$ zu einem Zeitpunkt $t \in \mathbb{T}$ interessiert.
Bevor allerdings die Konzentration als Lösung des Transportproblems bestimmt werden kann, muss zunächst das Flussvektorfeld $q : \overline{\mathbb{D}} \to \mathbb{R}^d$ berechnet werden.
Sei hierfür $p : D \to \mathbb{R}$ der hydrostatische Druck, $\kappa : D \to (\mathbb{R}_{\text{sym}})^{d\times d}$ der Permeabilitätstensor und $G = (0, 0, p_0 g_0)^\top$. Wie bereits in der Einleitung angedeutet, kann der Fluss des Regenwassers durch das Darcy-Gesetz $q = -\kappa(\nabla p + G)$ modelliert werden. Durch $u(x) := p(x) + p_0 g_0 x_3$ vereinfacht sich das Darcy-Gesetz zu $q = -\kappa\nabla u$.
Nehmen wir die physikalische Annahme hinzu, dass der Fluss q 'quellfrei' sein soll, also an keiner Stelle Masse verschwinden oder erscheinen kann, erhalten wir das Potentialströmungsproblem:

$$\text{Bestimme } u : \overline{\mathbb{D}} \to \mathbb{R} \text{ und } q : \overline{\mathbb{D}} \to \mathbb{R}^2 \text{ mit}$$

$$\text{(PS)} \begin{cases} \operatorname{div} q = 0 & ,\text{in } \mathbb{D} \\ q = -\kappa\nabla u & ,\text{in } \mathbb{D} \\ u = u_D & ,\text{auf } \Gamma_D \\ -q \cdot n = g_N & ,\text{auf } \Gamma_N \end{cases}$$

Bemerkung. Wir wollen aus verschiedenen Gründen direkt die sogenannte gemischte Formulierung des Potentialströmungsproblem nutzen. Näheres dazu findet sich im nächsten Abschnitt.

Anschließend suchen wir die Dichteverteilung $\rho : \mathbb{D} \times \mathbb{T} \to \mathbb{R}_{\geq 0}$ einer transportierten Substanz (in unserem Modell das Öl).
Gegeben sei dazu die Anfangsverteilung $\rho_0 : \mathbb{D} \to \mathbb{R}_{\geq 0}$ und der Einfluss der Substanz über die Zeit $\rho_{\text{in}} : \Gamma_{\text{in}} \times \mathbb{T} \to \mathbb{R}_{\geq 0}$ mit $\Gamma_{\text{in}} := \{z \in \partial\mathbb{D} : q(z) \cdot n(z) \leq 0\} \subset \partial\mathbb{D}$. Dabei ist $n(z)$ der äußere Normalenvektor im (Rand-)Punkt z. Wir bedienen uns wieder der Physik und fordern die Erfüllung der Bilanzgleichung

$$\forall K \subseteq \mathbb{D}, t \in \mathbb{T} : \frac{d}{dt}\int_K \rho(x,t)\,\mathrm{d}x + \int_{\partial K} \rho(x,t)q(x)\cdot n(x)\,\mathrm{d}a = 0.$$

Wenden wir für ein zulässiges $K \subseteq \mathbb{D}$ und $\rho, q \in C^1(\mathbb{D})$ den Satz von Gauß an erhalten wir

$$\int_K \partial_t \rho(x,t) + \operatorname{div}(\rho q)(x,t)\,\mathrm{d}x = 0$$

und können so die lineare Transportgleichung ableiten:

$$\partial_t \rho + \operatorname{div}(\rho q) = 0 \text{ in } \mathbb{D} \times (0,T]$$

Mit den entsprechenden Rand- und Anfangswerten erhalten wir so:

$$\text{Bestimme } \rho : \overline{\mathbb{D}} \times \mathbb{T} \to \mathbb{R}_{\geq 0}, \text{ sodass}$$

$$\text{(TP)} \begin{cases} \partial_t \rho + \operatorname{div}(\rho q) = 0 & \text{,in} \quad \mathbb{D} \times (0,T) \\ \rho(x,t) = \rho_{\text{in}}(x,t) & \text{,auf} \quad \Gamma_{\text{in}} \times (0,T) \\ \rho(x,0) = \rho_0(x) & \text{,auf} \quad \mathbb{D} \end{cases}$$

5.1.2 Probabilistisches Problem

In dem letzten Unterabschnitt sind wir bereits bei der Lösung des Potentialströmungsproblems davon ausgegangen, sämtliche benötigten Randwerte sowie den Permeabilitätstensor κ exakt für das gesamte Gebiet $\mathbb{D}$ zu kennen. Wir wollen uns von dieser durchaus starken Annahme lösen und deshalb zusätzlich die Permeabilität κ mit Mitteln der Stochastik modellieren. Sei dazu $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und ab nun $d = 2$, also $\mathbb{D} \subseteq \mathbb{R}^2$.

Bemerkung. Grundsätzlich funktionieren die vorgestellten Verfahren auch für $d = 3$, wir wollen uns aber der Anschaulichkeit halber auf zwei Dimensionen beschränken. Das so betrachtete Gebiet $\mathbb{D}$ lässt sich so z.B. als Querschnitt einer Bodenschicht interpretieren.

Weiter sei nun $\kappa(\cdot, x) : \Omega \to \mathbb{R}_{\geq 0}$ die (vom Zufall abhängige) Permeabilität. Wie schon an anderer Stelle (z.B. in [21]) wollen wir die Permeabilität als lognormal-Feld modellieren. Unser so entstehendes Problem fällt somit in den Bereich der Uncertainty Quantification und gegeben durch:

Für $\omega \in \Omega$, bestimme $u(\omega, \cdot) : \overline{\mathbb{D}} \to \mathbb{R}$ und $q(\omega, \cdot) : \overline{\mathbb{D}} \to \mathbb{R}^2$ mit

$$\text{(PS)} \begin{cases} \operatorname{div}(q(\omega, x)) &= 0 & \text{, für } x \in \mathbb{D} \\ q(\omega, x) &= -\kappa(\omega)\nabla u(\omega, x) & \text{, für } x \in \mathbb{D} \\ -q(\omega, x) \cdot n &= g_N(x) & \text{, für } x \in \Gamma_N \\ u(\omega, x) &= u_D(x) & \text{, für } x \in \Gamma_D \end{cases}$$

Für $\omega \in \Omega$ und $q(\omega, \cdot) : \overline{\mathbb{D}} \to \mathbb{R}^2$, bestimme $\rho(\omega, \cdot) : \overline{\mathbb{D}} \times \mathbb{T} \to \mathbb{R}_{\geq 0}$ mit

$$\text{(TP)} \begin{cases} \partial_t \rho(\omega, x, t) + \operatorname{div}(\rho(\omega, x, t) q(\omega, x)) &= 0 & \text{, für } (x, t) \in \mathbb{D} \times (0, T] \\ \rho(\omega, x, t) &= \rho_{\text{in}}(x, t) & \text{, für } (x, t) \in \Gamma_{\text{in}} \times \mathbb{T} \\ \rho(\omega, x, 0) &= \rho_0(x) & \text{, für } x \in \mathbb{D} \end{cases}$$

für die Anfangs- und Randwerte:

$g_N : \Gamma_{\text{N}} \to \mathbb{R}$
$u_D : \Gamma_{\text{D}} \to \mathbb{R}$
$\rho_{\text{in}} : \Gamma_{\text{in}} \times \mathbb{T} \to \mathbb{R}_{\geq 0}$
$\rho_0 : \mathbb{D} \to \mathbb{R}_{\geq 0}$

wobei $\partial\mathbb{D} = \Gamma_{\text{D}} \dot{\cup} \Gamma_{\text{N}}$ und $\Gamma_{\text{in}} \coloneqq \{z \in \partial\mathbb{D} : q(z) \cdot n(z) \leq 0\} \subset \partial\mathbb{D}$

Dabei stellen wir uns die Aufgabe, den Erwartungswert eines gegebenen Zielfunktionals $Q(\rho)$ zu berechnen, etwa dem Ausfluss der transportierten Substanz über den Rand. An dieser Stelle können wir dann, nachdem wir uns in den nächsten zwei Unterabschnitten damit beschäftigt haben, wie wir obige Probleme numerisch lösen, die MLMC Methode nutzen, um diesen Erwartungswert zu berechnen.

5.2 Numerische Lösung des Potentialströmungsproblem

Bemerkung. Die beiden folgenden Abschnitte bauen im Wesentlichen auf den beiden Vorlesungen 'Einführung in das Wissenschaftliche Rechnen' (SS 2019) und 'Finite Elemente Methoden' (WS 2019/2020) von Herrn Prof. Dr. Wieners auf. Dem entsprechend sind als Quellen neben [3], [2] und [15] vor allem die Mitschriebe zu den oben genannten Vorlesungen, sowie die Berichte zum Rechnerpraktikum mit M++ [1] zu nennen.

Wie bereits in obigem Abschnitt erwähnt, sollen sich die nächsten beiden Abschnitte damit beschäftigen, wie wir die oben beschriebenen Probleme für ein festes $\omega \in \Omega$ numerisch lösen können. Wir wollen dabei im Folgenden auf eine Möglichkeit eingehen, diese Berechnung numerisch durchzuführen. Insbesondere werden dabei jene Verfahren beschrieben, welche wir auch später innerhalb der MLMC Methode in M++ nutzen wollen. Da wir in diesen beiden Abschnitten $\omega \in \Omega$ ohnehin fest halten, genügt es zudem das deterministische Problem zu betrachten.
Sowohl das hybride Finite Elemente Verfahren, welches wir zur Lösung des Potentialströmungsproblem nutzen wollen, als auch das Discontinuous Galerkin Vefahren, mit dessen Hilfe wir das Transportproblem lösen wollen, bauen auf der Finite Elemente Theorie auf. Diese ist im Wesentlichen in der zweiten Hälfte des 20. Jahrhunderts entstanden, ist aber bis heute in praktischer wie auch in theoretischer Sicht aktuell. Die Grundidee ist hierbei, die vorliegenden Rand-Anfangswertaufgaben in einem passenden endlichen Unterraum zu lösen. Dabei löst man sich auf analytischer Seite zunächst oft von einzelnen Regularitäts- und Differenzierbarkeitsbedingungen und führt einen sogenannten schwachen Lösungsbegriff ein (vergleiche Abschnitt 2.1). Statt nun aber solch eine schwache Lösung in einem unendlich dimensionalen Funktionenraum, wie beispielsweise in den Sobolevräumen $H^1(\mathbb{D})$ oder $H_0^1(\mathbb{D})$ zu bestimmen, zieht man sich auf endlich dimensionale Unterräume zurück.
Die folgende Definition entstammt [3] und geht ursprünglich (1978) auf Ciarlet zurück.

Definition 5.1. Sei

- $K \subseteq \mathbb{R}^d$ eine beschränkte abgeschlossene Menge mit einem nichtleeren Inneren und stückweise stetig differenzierbarem Rand
- $\mathcal{P}$ ein endlich dimensionaler Funktionenraum auf K
- $\mathcal{N} = \{N_1, N_2, \ldots, N_k\}$ eine Basis für $\mathcal{P}'$

Dann heißt $(K, \mathcal{P}, \mathcal{N})$ ein finites Element.

Wir wollen im Folgenden diese theoretische Definition zwar im Hinterkopf behalten, aber wie in [2] meist nur mit den sogenannten Finite-Elemente-Räumen arbeiten. Dabei wird eine geeignete Zerlegung $\mathcal{T} = \{K_1, K_2, \ldots, K_M\}$ von $\mathbb{D}$ in endlich viele Teilgebiete gewählt. Anschließend betrachten wir einen endlichen Raum von Funktionen, die eingeschränkt auf diese Teilgebiete von einfacher Gestalt sind, beispielsweise bieten sich oft polynomielle Darstellungen niedrigen Grades an. Ein solches Teilgebiet $K \in \mathcal{T}$ nennen

wir Finites Element oder auch Zelle und fordern implizit, verbunden mit dem betrachteten Funktionenraum, die Erfüllung der obigen Definition.
Im Falle $\mathbb{D} \subseteq \mathbb{R}^2$ kommen so z.B. Dreiecke oder Vierecke in Frage, in $\mathbb{D} \subseteq \mathbb{R}^3$ können Tetraeder, Würfel, Quader und andere genutzt werden.
Sei nun $\mathbb{D} \subseteq \mathbb{R}^2$ zudem ein polygonales Gebiet, um eine einfache Zerlegung in Dreiecke oder Vierecke zu gewährleisten.

Definition 5.2. 1. Eine Zerlegung $\mathcal{T} = \{K_1, K_2, \dots, K_M\}$ von $\mathbb{D}$ in Dreiecks- oder Viereckselemente heißt zulässig, wenn folgende Eigenschaften erfüllt sind:

- $\overline{\mathbb{D}} = \bigcup_{i=1}^{M} K_i$
- Für $i \neq j$ ist $K_i \cap K_j$
  a) ein gemeinsamer Eckpunkt von K_i und K_j
  b) eine gemeinsame Kante von K_i als auch von K_j
  c) oder $K_i \cap K_j = \emptyset$

2. Wir schreiben oft $\mathcal{T}_h$ anstatt $\mathcal{T}$, wenn jedes Element einen Durchmesser von höchstens h besitzt .

3. Eine Familie von Zerlegungen $\{\mathcal{T}_h\}$ heißt uniform, wenn ein $\delta > 0$ existiert, sodass jedes $K \in \mathcal{T}_h$ einen Kreis mit Radius r_K enthält mit $r_K \geq \frac{h}{\delta}$.

Abbildung 1: Zulässige Zerlegung und unzulässige Zerlegung mit hängendem Knoten

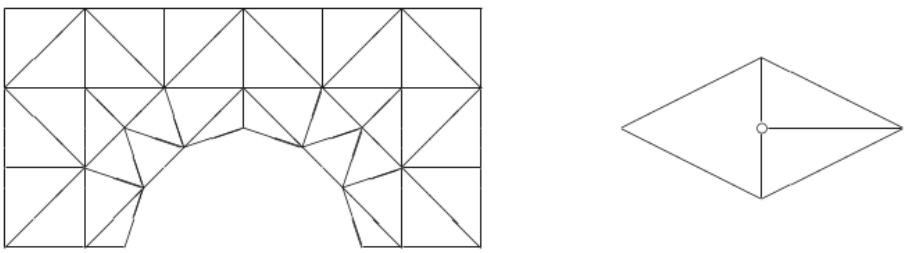

Abbildung aus [2] Seite 58

Wir werden außerdem im Laufe der Thesis dazu übergehen, ähnlich wie bereits im Abschnitt über die Multilevel Monte Carlo Methode auch bei Zerlegungen von 'Leveln' zu sprechen. Dabei betrachten wir stets eine uniforme Familie zulässiger Zerlegungen $\{\mathcal{T}_h\}_{h\in\mathcal{H}}$ und fordern dabei, dass die Indexmenge $\mathcal{H}$ eine ganz bestimmte Form hat. Genauer soll

$$\mathcal{H} = \{h_0, h_1 \coloneqq \frac{h_0}{2}, h_2 \coloneqq \frac{h_1}{2} = \frac{h_0}{4}, \dots\} \text{ für ein } h_0 > 0$$

gelten. Insbesondere gelte also $\overline{\mathcal{H}} \ni 0$. Sprechen wir dann von Level i meinen wir damit die Zerlegung $\mathcal{T}_{h_i} \in \{\mathcal{T}_h\}$. Zudem führen wir für alle Zerlegungen folgende Bezeichnungen ein:

- ein $K \in \mathcal{T}$ nennen wir Zelle
- ein $z \in \mathcal{V}_K \coloneqq \{z_{K,0}, z_{K,1}, z_{K,2}\} \subset \mathbb{R}^2$ nennen wir Knoten und $\mathcal{V}_K$ die Menge der Knoten von K
- $\mathcal{V}_\mathcal{T} \coloneqq \bigcup_{K\in\mathcal{T}} \mathcal{V}_K$ sei die Menge aller Knoten
- $\mathcal{F} \coloneqq (\{\partial K_1 \cap \partial K_2 : K_1, K_2 \in \mathcal{T}\} \cup \{\partial K_1 \cap \partial\mathbb{D} : K_1 \in \mathcal{T}\}) \setminus \{\emptyset\}$ sei die Menge aller Seiten
- $\mathcal{F}_K \coloneqq (\{\partial K \cap \partial K' : K' \in \mathcal{T}\} \cup \{\partial K \cap \partial\mathbb{D}\}) \setminus \{\emptyset\}$ sei die Menge aller Seiten von K
- $\partial\mathbb{D}_h \coloneqq \bigcup_{F\in\mathcal{F}} F$ sei der Rand von $\mathbb{D}_h$.

5.2.1 Schwache Formulierung

Betrachten wir also die deterministische Version des Potentialströmungsproblem:

$$\text{Bestimme } u : \overline{\mathbb{D}} \to \mathbb{R} \text{ und } q : \overline{\mathbb{D}} \to \mathbb{R}^2 \text{ mit}$$

$$\text{(PS)} \begin{cases} \operatorname{div} q = 0 & \text{,in } \mathbb{D} \quad (1) \\ q = -\kappa\nabla u & \text{,in } \mathbb{D} \quad (2) \\ u = u_D & \text{,auf } \Gamma_D \\ -q \cdot n = g_N & \text{,auf } \Gamma_N \end{cases}$$

Satz 2.6 sagt uns , dass wir in obiger Formulierung Gleichung (1) mit Testfunktionen $\phi \in H^1(\mathbb{D})$ und Gleichung (2) mit Testfunktionen $\psi \in H^1(\operatorname{div}, \mathbb{D})$ multiplizieren und anschließend über $\mathbb{D}$ integrieren können und so eine äquivalente schwache Formulierung herleiten:

$$\int_\mathbb{D} \operatorname{div}(q)\phi \,\mathrm{d}x = 0 \text{ für alle Testfunktionen } \phi : \mathbb{D} \to \mathbb{R}$$
$$\int_\mathbb{D} (q + \kappa\nabla u) \cdot \psi \,\mathrm{d}x = 0 \text{ für alle Testfunktionen } \psi : \mathbb{D} \to \mathbb{R}^2$$

Da κ weiter symmetrisch positiv definit ist, lässt sich letztere Gleichung zu

$$\begin{aligned} & \int_\mathbb{D} \kappa^{-1}(q + \kappa\nabla u) \cdot \psi \,\mathrm{d}x = 0 \\ \Leftrightarrow \qquad & \int_\mathbb{D} \nabla u \cdot \psi \,\mathrm{d}x = -\int_\mathbb{D} (\kappa^{-1} q) \cdot \psi \,\mathrm{d}x \qquad (\star) \end{aligned}$$

umformen. Außerdem wollen wir nun noch die Dirichlet-Randbedingungen $u = u_D$ auf Γ_D einfließen lassen. Dazu verwenden wir den Satz von Gauß:

$$\int_{\partial\Omega} (u\psi) \cdot n \,\mathrm{d}a \overset{\text{Gauß}}{=} \int_\Omega \operatorname{div}(u\psi) \,\mathrm{d}x = \int_\Omega \nabla u \cdot \psi \,\mathrm{d}x + \int_\Omega u \operatorname{div}(\psi) \,\mathrm{d}x \quad (\psi : \Omega \to \mathbb{R}^2)$$

Wählen wir nun unseren Ansatzraum so, dass für die Funktion ψ gilt $\psi \cdot n = 0$ auf Γ_N. Damit folgt

$$\int_{\Gamma_D} (u_D\psi)\cdot n \,\mathrm{d}a \overset{\substack{\psi\cdot n|_{\Gamma_N}=0 \\ u|_{\Gamma_D}=u_D}}{=} \int_{\partial\Omega} (u\psi)\cdot n \,\mathrm{d}a = \underbrace{\int_\Omega \nabla u\cdot\psi \,\mathrm{d}x}_{\overset{(\star)}{=} -\int_\Omega (\kappa^{-1}q)\cdot\psi \,\mathrm{d}x} + \int_\Omega u \,\mathrm{div}(\psi)\,\mathrm{d}x.$$

Die Neumann-Randbedingung $(\kappa\nabla u)\cdot n = g_N$ auf Γ_N wird durch die Wahl des Lösungsraumes erfüllt.

Wir erhalten so folgende schwache Formulierung:

Bestimme (q,u) mit $q\cdot n = -g_N$ auf Γ_N und

$$\text{(sPS)} \begin{cases} \int_{\mathbb{D}} \kappa^{-1}q\cdot\psi \,\mathrm{d}x - \int_{\mathbb{D}} u \,\mathrm{div}(\psi)\,\mathrm{d}x &= -\int_{\Gamma_D}(u_D\psi)\cdot n\,\mathrm{d}a \\ \int_{\mathbb{D}} \mathrm{div}(q)\,\phi\,\mathrm{d}x &= 0 \end{cases}$$

für alle (ψ,ϕ) in einem geeigneten Testraum mit $\psi\cdot n = 0$ auf Γ_N

5.2.2 Diskretisierung

Sei $\mathcal{T}$ eine zulässige Zerlegung von $\mathbb{D}$ und alle Bezeichnungen wie oben. Wir nummerieren zunächst die Zellen und die Seiten durch:

$$\begin{aligned} \mathcal{F} &= \{F_1,\dots,F_{|\mathcal{F}|}\} \qquad \text{globale Seitennummerierung} \\ \mathcal{T} &= \{K_1,\dots,K_{|\mathcal{T}|}\} \qquad \text{globale Zellennumerierung} \end{aligned}$$

Dabei sei im Weiteren $N \coloneqq |\mathcal{F}|$ und $\mathrm{M} \coloneqq |\mathcal{T}|$ Als Nächstes soll es nun Ziel sein, eine Lösung der im letzten Abschnitt erklärten schwachen Formulierung in einem endlich dimensionalen Finite Elemente Ansatzraum zu bestimmen. Um aber hierfür genau diese Räume definieren zu können, benötigen wir zuerst sogenannte Basisfunktionen, genauer die Seiten- und die Zellenbasis.

Definition 5.3. (Seiten- und Zellenbasis)

(a) $\{\psi_i\}_{i=1}^N$ heißt Seitenbasis und ist definiert durch

$$\forall i,j\in\{1,\dots,N\} : \int_{F_j} \psi_i\cdot n^K \,\mathrm{d}a = \pm\delta_{i,j} \text{ und } \psi_i|_K \in \mathbb{P}_1(K,\mathbb{R}^2)\cap C(\overline{\mathbb{D}}) \ (K\in\mathcal{T})$$

(b) $\{\mu_i\}_{m=1}^{|\mathcal{K}|}$ heißt Zellenbasis und ist gegeben durch

$$\forall m\in\{1,\dots,M\} : \mu_m \coloneqq \mathbb{1}_{K_m}.$$

Anschließend können wir mithilfe dieser Basisfunktionen die Testräume bzw. Finite Elemente Räume definieren:

Definition 5.4. (Ansatzräume)

(a) $W_h \coloneqq \operatorname{span}\{\psi_1, \dots, \psi_N\}$ (Seitenansatzraum/ Raum für ψ und q_h)

(b) $W_h(g) \coloneqq \{\psi_h \in W_h : \int_F \psi_h \cdot n \,\mathrm{d}a = \int_F g \,\mathrm{d}a \;\text{ für alle } F \subseteq \Gamma_\mathrm{N})\}$

(c) $Q_h \coloneqq \operatorname{span}\{\mu_1, \dots, \mu_M\}$ (Zellenansatzraum/ Raum für ϕ und u_h)

Zusammen mit der schwachen Formulierung (5.2.1) erhalten wir so das nun diskretisierte Problem:

$$\begin{aligned}
&\text{Bestimme } (q_h, u_h) \in W_h(-g_N) \times Q_h \text{ mit} \\
&\begin{cases} \int_\Omega \kappa^{-1} q_h \cdot \psi_h \,\mathrm{d}x - \int_\Omega u_h \operatorname{div}(\psi_h) \,\mathrm{d}x &= -\int_{\Gamma_D} (u_D \psi_h) \cdot n \,\mathrm{d}a \\ \int_\Omega \operatorname{div}(q_h)\, \phi_h \,\mathrm{d}x &= 0 \end{cases} \\
&\text{für alle } (\psi_h, \phi_h) \in W_h(0) \times Q_h
\end{aligned}$$

5.3 Formulierung als LGS

Wir können nun damit beginnen, das so entstandene endlich dimensionale Problem in ein Lineares Gleichungs System umzuformulieren. Dazu definieren wir:

$$\begin{aligned}
\underline{A} \in \mathbb{R}^{N \times N} \text{ mit } \underline{A}[n,k] &\coloneqq \int_\Omega \kappa^{-1} \psi_n \cdot \psi_k \,\mathrm{d}x \\
\underline{B} \in \mathbb{R}^{M \times N} \text{ mit } \underline{B}[m,k] &\coloneqq -\int_\Omega \mu_m \operatorname{div}(\psi_k) \,\mathrm{d}x \\
\underline{b} \in \mathbb{R}^{N} \text{ mit } \underline{b}[k] &\coloneqq -\int_{\Gamma_D} u_D \psi_k \cdot n \,\mathrm{d}a
\end{aligned}$$

und (für die Randbedingungen)

$$\underline{W}(g) \coloneqq \left\{ \underline{q} \in \mathbb{R}^N : \underline{q}[k] = \int_{F_k} g \,\mathrm{d}a \;(\text{für } k \text{ mit } F_k \subseteq \Gamma_N) \right\}$$

Unser zu lösendes Problem lässt sich so mit $q_h = \sum_{n=1}^{N} \underline{q}[n]\psi_n$ und $u_h = \sum_{m=1}^{M} \underline{u}[m]\mu_m$ umformen zu

$$\begin{aligned}
&\text{Bestimme} (\underline{q}, \underline{u}) \in \underline{W}(-g_N) \times \mathbb{R}^M \text{ mit} \\
&\text{(L gFE)} \begin{cases} \underline{A}\underline{q} + \underline{B}^T \underline{u} &= \underline{b} \\ \underline{B}\underline{q} &= 0 \end{cases}
\end{aligned}$$

oder anders geschrieben

$$\text{Bestimme}(\underline{q}, \underline{u}) \in \underline{W}(-g_N) \times \mathbb{R}^M \text{ mit}$$
$$\text{(dgPS)} \left\{ \begin{pmatrix} \underline{A} & \underline{B}^T \\ \underline{B} & 0 \end{pmatrix} \begin{pmatrix} \underline{q} \\ \underline{u} \end{pmatrix} = \begin{pmatrix} \underline{b} \\ 0 \end{pmatrix}. \right.$$

Wir haben so eine diskrete gemischte Formulierung des Potentialströmungsproblem hergeleitet und können mit dieser aus gegebenen Rand- und Anfangswerten ein Flussvektorfeld q erzeugen, welches der obigen Differentialgleichung genügt. Es handelt sich hierbei um das gemischte Finite Elemente Verfahren. In M++ selbst lösen wir das Potentialströmungsproblem durch eine Abwandlung dieses Verfahrens. Wir diskretisieren dazu eine äquivalente Formulierung von (sPS) und erhalten so mit dem hybriden Finite Elemente Verfahren die gleichen Ergebnisse, die auch der vorgestellte gemischte Ansatz liefern würde, bei besserer Effizienz und guter Parallelisierbarkeit. Da das Potentialströumgsproblem in dieser Thesis primär dazu genutzt werden soll, das Vektorfeld q zu bestimmen, soll uns aus theoretischer Sicht aber obige Formulierung genügen und wir verweisen hinsichtlich der Lösung mit hybriden gemischten Finiten Elementen, neben einem kleinen, Überblick verschaffendem Abschnitt im Appendix 9.2, auf die Literatur, wie etwa [4] oder [27].

5.4 Numerische Lösung des Transportproblem

In diesem Abchnitt soll nun, nachdem wir $q(\omega, \cdot)$ als Finite-Elemente-Lösung des Potentialströmungsproblems erhalten haben, die numerische Lösung des linearen Transportproblems behandelt werden:

Für $\omega \in \Omega$ und $q(\omega, \cdot) : \overline{\mathbb{D}} \to \mathbb{R}^2$, bestimme $\rho(\omega, \cdot) : \overline{\mathbb{D}} \times \mathbb{T} \to \mathbb{R}_{\geq 0}$ mit

$$\text{(pTP)} \begin{cases} \partial_t \rho(\omega, x, t) + \operatorname{div}(\rho(\omega, x, t) q(\omega, x)) &= 0 & \text{, für } (x,t) \in \mathbb{D} \times (0, T] \\ \rho(\omega, x, t) &= \rho_{\text{in}}(x, t) & \text{, für } (x,t) \in \Gamma_{\text{in}} \times \mathbb{T} \\ \rho(\omega, x, 0) &= \rho_0(x) & \text{, für } x \in \mathbb{D} \end{cases}$$

Insbesondere wollen wir an dieser Stelle wieder $\omega \in \Omega$ fest halten und betrachten deshalb zunächst nur das deterministische Problem wie in 5.1.1:

Bestimme $\rho : \overline{\mathbb{D}} \times \mathbb{T} \to \mathbb{R}_{\geq 0}$, sodass

$$\text{(dTP)} \begin{cases} \partial_t \rho(x, t) + \operatorname{div}(\rho(x, t) q(x)) = 0 & \text{,in} \quad \mathbb{D} \times (0, T) \\ \rho(x, t) = \rho_{\text{in}}(x, t) & \text{,auf } \Gamma_{\text{in}} \times (0, T) \\ \rho(x, 0) = \rho_0(x) & \text{,auf } \mathbb{D} \end{cases}$$

Wir greifen dabei auf ein sogenanntes discontinuous Galerkin Verfahren zurück, welches für diese Problemklasse bereits an anderen Stellen (z.B. in [9]) erprobt wurde. Ursprünglich geht das Discontinuous Galerkin Verfahren auf Reed und Hill [25] zurück. Einen guten (wenn auch mittlerweile etwas in die Jahre gekommenen) Überblick über die Anwendung von discontinuous Galerkin Verfahren bietet [7]. Grundätzlich handelt es sich beim discontinuous Galerkin Verfahren ebenfalls um einen FEM Ansatz, der zwar Änhlichkeiten zum Finite Elemente Verfahren aufweist, welches wir im letzten Abschnitt gesehen hatten, aber auch einige bedeutende Unterschiede aufweist, auf welche wir im Folgenden besonders eingehen wollen. Anders als zuvor das Potentialströmungsproblem ist die lineare Transportgleichung nämlich sowohl Orts- als auch Zeitabhängig. Daher werden wir die lineare Transportgleichung zunächst im Ort diskretisieren. Wir erhalten so eine Semidiskretisierung, welche wir anschließend mit einem Zeitintegrator, wie beispielsweise der impliziten Mittelpunktsregel, in eine Volldiskretisierung überführen.

5.4.1 Diskretisierung

Wie bereits weiter oben beschrieben werden wir im Folgenden zunächst den Raum diskretisieren und anschließend die so entstandene Semidiskretisierung in eine Volldiskretisierung auflösen. Insgesamt wollen wir das discontinuous Galerkin Verfahren mit einem Zeitintegrator, wie der impliziten Mittelpunktsregel oder einem klassischen Runge-Kutta-Verfahren nutzen. Zunächst führen wir die analytische Flussfunktion ein.

Definition 5.5. (Flussfunkion)
Zu einem gegebenen Flussvektorfeld $q : \mathbb{D} \to \mathbb{R}^2$ ist die Flussfunktion Υ definiert als:

$$\Upsilon : \mathrm{Abb}(\mathbb{D} \times \mathbb{T}, \mathbb{R}) \to \mathrm{Abb}(\mathbb{D} \times \mathbb{T}, \mathbb{R}^2)$$
$$\rho \mapsto \rho q$$

Für eine klassische Lösung ρ von (dTP) gilt dann insbesondere $\partial_t \rho = -\operatorname{div}(\Upsilon(\rho))$ auf $\mathbb{D} \times (0, T]$.
Halten wir also zunächst $t \in \mathbb{T}$ und leiten so die Semidiskretisierung her.
Sei nun $\mathcal{T}$ eine zulässige Triangulierung von $\mathbb{D}$ aus Dreiecken wie in 5.2 und $(\cdot, \cdot)_A$ das $L^2(A)-$Skalarprodukt. Wir wählen als Lösungs-/Testraum $Q_h = \prod_{K \in \mathcal{T}} \mathbb{P}_p(K, \mathbb{R})$ für ein festes $p \geq 1$. Anders als zuvor fordern wir für unsere Lösungs- und Testfunktionen diesmal aber nicht die Stetigkeit auf $\mathbb{D}$. Da so Q_h nicht im betrachteten analytischen Lösungs- und Testraum liegt, etwa $Q_h \nsubseteq H^1(\mathbb{D})$, nennt man Q_h auch einen nicht-konformen Ansatzraum. Außerdem lässt sich im Allgemeinen auch die später bestimmt Lösung $\rho_h \in Q_h$ (definiert auf $\mathbb{D}_h = \bigcup_{K \in \mathcal{T}} K$) nicht stetig auf $\mathbb{D}$ fortsetzen, denn für eine beliebige innere Kante F kann der Grenzwert von ρ_h auf den anliegenden Zellen K, K' ($\overline{F} = \partial K \cap \partial K'$) unterschiedlich sein.
Trotzdem müssen wir auch auf den inneren Kanten $\mathcal{F}^0 \subset \mathcal{F}$ festlegen, welcher Grenzwert in einem solchen Falle gewählt wird.
dazu führen wir als Pendant zur analytischen Flussfunktion (vgl. 5.5) auch eine numerische Flussfunktion ein. Grundsätzlich kommen mehrere solche Flussfunktionen in Frage, welche direkten Einfluss auf Eigenschaften des entstehenden Verfahrens besitzen. Wir entscheiden uns an dieser Stelle für den weit verbreiteten sogenannten upwind flux:

Definition 5.6. (upwind flux)
Sei $K \in \mathcal{T}$ eine beliebige Zelle und $F \in \mathcal{F}_K$ eine Kante von K. Dann ist

$$\Upsilon^\star : \mathrm{Abb}(\mathbb{D} \times \mathbb{T}, \mathbb{R}) \to \mathrm{Abb}(\mathbb{D} \times \mathbb{T}, \mathbb{R}^2)$$
$$\rho_h \mapsto \begin{cases} \Upsilon(\rho_h|_K), & \text{für } q \cdot n_F^K \geq 0 \\ \Upsilon(\rho_h|_{K'}), & \text{für } q \cdot n_F^K < 0 \text{ und } \overline{F} = \partial K \cap \partial K' \end{cases}$$

Sei nun also ρ klassische Lösung von (dTP) mit $\partial_t \rho = -\operatorname{div}(\Upsilon(\rho))$ auf $\mathbb{D}$. Dann gilt nach Satz von Gauß:

$$\int_{\partial \mathbb{D}} \rho q \cdot n \phi \, \mathrm{d}a = \int_{\partial \mathbb{D}} \Upsilon(\rho) \cdot n \phi \, \mathrm{d}a = \int_{\mathbb{D}} \operatorname{div}(\Upsilon(\rho)\phi) \, \mathrm{d}x \tag{5.1}$$

Das Integral über den Rand von $\mathbb{D}$ können wir nach der folgenden kleinen Vorüberlegung auch als Integral über alle Kanten der gewählten Zerlegung $\mathcal{T}$ ausdrücken:
Es gilt nämlich für alle inneren Kanten, also solche Kanten F für die zwei Zellen K und K' existieren, sodass $\overline{F} = K \cap K'$ ist, dass $\int_F \Upsilon^\star(\rho) \cdot n^K \phi \, \mathrm{d}a = -\int_F \Upsilon^\star(\rho) \cdot n^{K'} \phi \, \mathrm{d}a$ stets erhalten ist.
Summieren wir also zunächst über alle Zellen, addieren anschließend die Integrale über die Kanten und ersetzen dabei den analytischen durch den numerischen Fluss, erhalten

wir gerade wieder obiges Randintegral. Es gilt also:

$$\sum_{K\in\mathcal{T}}\sum_{F\in\mathcal{F}_K}\int_F \Upsilon^\star(\rho)\cdot n^K\phi\,\mathrm{d}a = \int_{\partial\mathbb{D}}\Upsilon(\rho)\cdot n\phi\,\mathrm{d}a \overset{5.1}{=} \int_{\mathbb{D}}\operatorname{div}(\Upsilon(\rho)\phi)\,\mathrm{d}x$$

Nach der Produktregel der Divergenz lässt sich das letzte Integral auswerten zu:

$$\int_{\mathbb{D}}\operatorname{div}(\Upsilon(\rho)\phi)\,\mathrm{d}x = \int_{\mathbb{D}}\phi\operatorname{div}(\Upsilon(\rho)) + \Upsilon(\rho)\cdot\nabla\phi\,\mathrm{d}x \overset{\text{Vor.}}{=} -\int_{\mathbb{D}}\partial_t\rho\phi\,\mathrm{d}x + \int_{\mathbb{D}}\Upsilon(\rho)\cdot\nabla\phi\,\mathrm{d}x$$

Durch Umstellen und das Zusammenfassen der obigen Resultate erhalten wir so:

$$\sum_{K\in\mathcal{T}}\int_K\partial_t\rho\phi\,\mathrm{d}x = \sum_{K\in\mathcal{T}}\int_K\Upsilon(\rho)\cdot\nabla\phi\,\mathrm{d}x - \sum_{K\in\mathcal{T}}\sum_{F\in\mathcal{F}_K}\int_F\Upsilon^\star(\rho)\cdot n^K\phi\,\mathrm{d}a$$

Dabei wurde zusätzlich ausgenutzt, dass es sich bei den Kanten um Nullmengen handelt und wir so das Integral über $\mathbb{D}$ als Summe der Integrale über alle Zellen auffassen können. Nutzen wir nun noch aus, dass für den Fluss $\rho(x,t) = \rho_{\text{in}}(x,t)$ für $x\in\Gamma_{\text{in}}$ gilt, kommen wir so auf

$$\sum_{K\in\mathcal{T}}\int_K\partial_t\rho\phi\,\mathrm{d}x = \sum_{K\in\mathcal{T}}\int_K\Upsilon(\rho)\cdot\nabla\phi\,\mathrm{d}x - \sum_{K\in\mathcal{T}}\left(\sum_{\substack{F\in\mathcal{F}_K\\F\not\subseteq\Gamma_{\text{in}}}}\int_F\Upsilon^\star(\rho)\cdot n^K\phi\,\mathrm{d}a - \sum_{\substack{F\in\mathcal{F}_K\\F\subseteq\Gamma_{\text{in}}}}\rho_{\text{in}}q\cdot n^K\phi\,\mathrm{d}a\right) \tag{5.2}$$

Sei nun $\mathcal{T} = \{K_1,\dots,K_N\}, N \coloneqq |\mathcal{T}|$. Die Semidiskretisierung ist motiviert durch (5.2) und lautet: Bestimme $\rho_h\in Q_h$, sodass für alle $\phi_h\in Q_h$ gilt:

$$\sum_{i=1}^N(\partial_t\rho_h,\phi_h)_{K_i} = \sum_{i=1}^N\left((\Upsilon(\rho_h),\nabla\phi_h)_{K_i} - \sum_{\substack{F\in\mathcal{F}_K\\F\not\subseteq\Gamma_{\text{in}}}}(\Upsilon^\star(\rho_h)\cdot n^K,\phi_h)_F - (\rho_{\text{in}}q\cdot n^K,\phi_h)_{\partial K_i\cap\Gamma_{\text{in}}}\right) \tag{5.3}$$

Durch Einsetzen der Zellenbasis $\{\mu_i\}_{i=1}^N$ mit $\mu_i = \mathbb{1}_{K_i}$ und mit $\operatorname{supp}(\mu_i)\subseteq\overline{K_i}(\ i\in\{1,\dots,N\})$ ergibt sich für alle $i\in\{1,\dots,N\}$

$$(\partial_t\rho_h,\mu_i)_{K_i} = \left(\underbrace{(\Upsilon(\rho_h),\nabla\mu_i)_{K_i}}_{=0,\text{ da }\nabla\mu_i=0} - \sum_{\substack{F\in\mathcal{F}_{K_i}\\F\not\subseteq\Gamma_{\text{in}}}}\left(\Upsilon^*(\rho_h)\cdot n^{K_i},\mu_i\right)_F - \left(\rho_{\text{in}}q\cdot n^{K_i},\mu_i\right)_{\partial K_i\cap\Gamma_{\text{in}}}\right)$$

Wir erhalten so folgende Darstellung:

$$\underbrace{(\partial_t \rho_h, \mu_i)_{K_i}}_{(5.4.1)} = \underbrace{-\sum_{F \in \mathcal{F}_{K_i}, F \not\subseteq \Gamma_{\text{in}}} \left(\Upsilon^\star(\rho_h) \cdot n^{K_i}, \mu_i\right)_F}_{(5.4.2)} \underbrace{- \left(\rho_{\text{in}} q \cdot n^{K_i}, \mu_i\right)_{\partial K_i \cap \Gamma_{\text{in}}}}_{(5.4.3)} \tag{5.4}$$

Zusammen mit der Basisdarstellung von ρ_h in $\{\mu_i\}_{i=1}^N$, $\rho_h = \sum_{i=1}^N \underline{\rho}[i]\ \mu_i$, und $\Upsilon^\star(\rho_h) = \begin{cases} \rho_h|_K \text{ , falls } q \cdot n^K|_F \geq 0 \\ \rho_h|_{K'} \text{ , falls } q \cdot n^K|_F < 0 \end{cases}$ können wir 5.4 in eine gewöhnliche Differentialgleichung erster Ordnung umformulieren. Dazu klammern wir jeweils $\underline{\rho}$ aus und definieren mithilfe von (5.4.1) die Massenmatrix $\underline{M} \in \mathbb{R}^{N \times N}$

$$\underline{M}[K, K'] \coloneqq \begin{cases} \int_K |\mu_K|^2 \,\mathrm{d}x & \text{, für } K = K' \\ 0 & \text{, sonst} \end{cases}$$

mit (5.4.2) die Flussmatrix $\underline{A} \in \mathbb{R}^{N \times N}$

$$\underline{A}[K, K'] \coloneqq \begin{cases} -\sum\limits_{\substack{F \in \mathcal{F}_K \\ F \text{ mit } q \cdot n^K_{|F} > 0}} \int_F \mu_K^2 q \cdot n^K \,\mathrm{d}a & \text{, für } K = K' \\ -\int_F \mu_K \mu_{K'} q \cdot n^K \,\mathrm{d}a & \text{, für } q \cdot n^K < 0 \text{ und } \overline{F} = \partial K \cap \partial K' \\ 0 & \text{, sonst} \end{cases}$$

und mit (5.4.3) den Lastvektor $\underline{b} \in \mathbb{R}^N$

$$\underline{b}[K] \coloneqq \int_{\partial K \cap \Gamma_{\text{in}}} \rho_{\text{in}} q \cdot n \,\mathrm{d}a$$

So ergibt sich die Differentialgleichung

$$\begin{cases} \underline{M} \partial_t \underline{\rho}(t) = \underline{A} \underline{\rho}(t) + \underline{b}(t) \\ \underline{\rho}(0) = \underline{\rho_0} \end{cases}$$

Da dies nun eine gewöhnliche Differentialgleichung ist, können wir die Lösung

$$\underline{\rho}(t) = \exp(t \underline{M}^{-1} \underline{A}) \left(\underline{\rho_0} + \int_0^t \exp(-s \underline{M}^{-1} \underline{A}) \underline{b}(s) \,\mathrm{d}s \right) \tag{5.5}$$

explizit angeben. Es handelt sich hierbei aber immer noch um eine semidiskrete Formulierung. Wir wollen deshalb zuletzt noch auf die Herleitung der Zeitintegratoren eingehen. Diese nutzen wir um unter Verwendung der oben hergeleiteten Semidiskretisierung die numerische Lösung $\underline{\rho}$ sowohl orts- al auch zeitdiskret zu berechnen. Der Ansatz leitet

sich hierbei direkt aus dem aus dem Resultat (5.5) ab und besteht aus der Integration der Differentialgleichung $\underline{\underline{M}}\partial_t\underline{\rho} = \underline{\underline{A}}\underline{\rho} + \underline{b}$ über die Zeit t im Intervall $[t_i, t_{i+1}]$. Dabei ist $t_i = i\delta t$. Hiermit folgt:

$$\underline{\underline{M}}\underline{\rho}(t_{i+1}) - \underline{\underline{M}}\underline{\rho}(t_i) = \int_{t_i}^{t_{i+1}} \underline{\underline{M}}\partial_t\underline{\rho}(t)dt = \int_{t_i}^{t_{i+1}} \underline{\underline{A}}\underline{\rho}(t) + \underline{b}(t)dt.$$

Mithilfe der Anwendung verschiedener Quadraturformeln lässt sich daraus ein Runge-Kutta Verfahren herleiten. Über die Rechteckformel

$$\int_{t_i}^{t_{i+1}} \underline{\underline{A}}\underline{\rho}(t) + \underline{b}dt \approx (t_{i+1} - t_i)(\underline{\underline{A}}\underline{\rho}(t_{i+1}) + \underline{b}(t_{i+1})) = \delta t(\underline{\underline{A}}\underline{\rho}(t_{i+1}) + \underline{b}(t_{i+1}))$$

ergibt sich z.B. das implizite Euler Verfahren

$$\underline{\rho}(t_{i+1}) = \underline{\rho}(t_i) + \delta t\underline{\underline{M}}^{-1}(\underline{\underline{A}}\underline{\rho}(t_{i+1}) + \underline{b}(t_{i+1})).$$

Ein weiteres Verfahren dieser Art, welches wir an dieser Stelle verwenden werden, ist die implizite Mittelpunktsregel (der Übersicht wegen für $\underline{b} \equiv 0$):

$$\underline{\rho}(t_{i+1}) = \underline{\rho}(t_i) + \delta t\underline{\underline{M}}^{-1}(\underline{\underline{A}}\frac{1}{2}(\underline{\rho}(t_i) + \underline{\rho}(t_{i+1})) + \frac{1}{2}(\underline{\rho}(t_i) + \underline{\rho}(t_{i+1}))).$$

Das so entstehenden Gesamtverfahren ist aufgrund der Kombination von Discontinuous Galerkin Verfahren und Runge-Kutta-Zeitintegratoren in der Literatur oft auch unter dem Namen 'Runge–Kutta discontinuous Galerkin Methods' zu finden. Einen schönen Überblick über diese Verfahrensklasse bietet der Artikel [10]. Nachdem wir nun das Discontinuous Galerkin Verfahren für die lineare Transportgleichung eingeführt und erklärt haben, sollen nun noch auf einige Eigenschaften des Verfahrens verwiesen werden. Dabei wollen wir uns aber beschränken einige grundlegende Resultate zu nennen und so eher einen groben Überblick mit Referenzen zur Literatur zu geben. Mehr zur numerischen Analyse des Discontinuous Galerkin Verfahren findet sich zum einen in Standardwerken, wie [13], eine schöne Zusammenstellung bietet aber auch [16].
Ebenfalls findet sich in [16] eine grundlegende numerische Analyse des Discontinuous Galerkin Verfahrens angewandt auf die stationäre (TODO stimmt das? fehlende Zeitabhängigkeit) lineare Transportgleichung. Dabei werden unter anderem die Konsistenz, die sogenannte Galerkin-Orthogonalität sowie die Stabilität und Konvergenz des Verfahrens behandelt. Mit der numerischen Analyse des Discontinuous Galerkin Verfahrens an sich befassten sich unter anderem LeSaint und Raviart [23], Peterson [24] und Richter [26]. Runge-Kutta DG verfahren für der linearen Transportgleichung ähnliche Problemstellungen betrachteten Cockburn und Shu in einer 5 teiligen Serie von Arbeiten. Besonders zu nennen sind dabei in unserem Kontext [8] und [6].

5.5 Eigenschaften des Discontinuous Galerkin Verfahren

Nachdem wir nun das Discontinuous Galerkin Verfahren für die lineare Transportgleichung eingeführt und erklärt haben, sollen noch einige Eigenschaften des Verfahrens beleuchtet werden. Dabei wollen wir uns aber darauf beschränken, einige grundlegende Resultate zu nennen und so eher einen groben Überblick mit Referenzen zur Literatur zu geben. Mehr zur numerischen Analyse des Discontinuous Galerkin Verfahren findet sich zum einen in Standardwerken, wie [13], eine schöne Zusammenstellung bietet aber auch [16]. Ebenfalls findet sich in [16] eine grundlegende numerische Analyse des Discontinuous Galerkin Verfahrens auf die stationäre lineare Transportgleichung.

5.5.1 Lösungsbegriffe

Wie zuvor bereits beim Potentialströmungsproblem können wir auch für das Transportproblem eine sogenannte schwache Formulierung bestimmen. Diese hängt im Fall des Transportproblems eng mit der Semidiskretisierung zusammen und lautet mit $\partial\mathbb{D} = \Gamma_{\text{in}} \dot\cup \Gamma_{\text{out}}$ und $\rho(x,t) = \rho_{\text{in}}(x,t)$ für $(x,t) \in \Gamma_{\text{in}} \times (0,T)$:

Definition 5.7. $\rho \in L_1(\mathbb{D} \times (0,T))$ heißt schwache Lösung des linearen Transportproblems, falls es für ein gegebenes $q : \overline{\mathbb{D}} \to \mathbb{R}^2$ folgende Bedingungen erfüllt:

$$
\begin{array}{lllll}
\text{(swTP)} & B(\rho,\phi) & = \langle l,\phi\rangle \quad \forall \phi \in H^1(\mathbb{D}\times\mathbb{T}) \text{ mit} & \phi(\cdot,T) = 0 \text{ und } \phi|_{\Gamma_{\text{out}}} = 0 \\
\text{Dabei sind :} & & & & \\
 & B(\rho,\phi) & := \int_0^T \int_{\mathbb{D}} \rho(\partial_t\phi + q\nabla\phi)\,\mathrm{d}x\,\mathrm{d}t & - \int_{\Gamma_{\text{out}}} \rho q\cdot n\phi\,\mathrm{d}a\,\mathrm{d}t \\
 & \langle l,\phi\rangle & := \int_{\Gamma_{\text{in}}} \rho_{\text{in}} q\cdot n\phi\,\mathrm{d}a\,\mathrm{d}t & - \int_{\mathbb{D}} \rho_0\phi(0)\,\mathrm{d}x \\
 & \Gamma_{\text{out}} & := \{z\in\partial\mathbb{D} : q(z)\cdot n(z) > 0\} & \\
 & \Gamma_{\text{in}} & := \{z\in\partial\mathbb{D} : q(z)\cdot n(z) \le 0\} &
\end{array}
$$

Es gilt an dieser Stelle außerdem:

Lemma 5.8. (Zusammenhang der Lösungsbegriffe)

1. Ist ρ eine klassische Lösung, so ist ρ auch eine schwache Lösung.
2. Ist $\rho \in C^2(\mathbb{D}\times\mathbb{T},\mathbb{R})$ und eine schwache Lösung, so ist ρ auch eine klassische Lösung.

Beweis. Sei $\phi : \mathbb{D}\times\mathbb{T} \to \mathbb{R}$ eine beliebige Testfunktion aus $H^1(\mathbb{D}\times\mathbb{T})$, für die $\phi(\cdot,T) = 0$ und $\phi|_{\Gamma_{\text{out}}} = 0$ gelte. Wir halten zunächst fest, dass der Raum $H_0^1(\mathbb{D}\times\mathbb{T})$ vollständig in dem so betrachteten Testraum enthalten. Wir beginnen nun mit der Differentialgleichung $\partial_t\rho(x,t) + \mathrm{div}(\rho(x,t)q(x)) = 0$, multiplizieren zunächst mit einer Testfunktion ϕ aus dem

Testraum und integrieren anschließend über den Raum-Zeitzylinder $\mathbb{D} \times \mathbb{T}$:

$$\int_{\mathbb{D}\times\mathbb{T}} \big(\partial_t \rho(x,t) + \operatorname{div}(\rho(x,t)q(x))\big)\phi(x,t)\ \mathrm{d}(x,t) = \underbrace{\int_{\mathbb{D}\times\mathbb{T}} \partial_t \rho(x,t)\phi(x,t)\ \mathrm{d}(x,t)}_{(1)} + \underbrace{\int_{\mathbb{D}\times\mathbb{T}} \operatorname{div}(\rho(x,t)q(x))\phi(x,t)\ \mathrm{d}(x,t)}_{(2)}$$

Betrachten wir nun zunächst Integral (1), so folgt mit partieller Integration:

$$\begin{aligned}
\int_{\mathbb{D}\times\mathbb{T}} \partial_t \rho(x,t)\phi(x,t)\ \mathrm{d}(x,t) &= \int_{\mathbb{D}}\int_{\mathbb{T}} \partial_t \rho(x,t)\phi(x,t)\,\mathrm{d}t\,\mathrm{d}x \\
&= \int_{\mathbb{D}} \Big(-\int_{\mathbb{T}} \rho(x,t)\partial_t\phi(x,t)\,\mathrm{d}t + \big[\rho(x,t)\phi(x,t)\big]_0^T \Big)\,\mathrm{d}x \\
&= -\int_{\mathbb{D}\times\mathbb{T}} \rho(x,t)\partial_t\phi(x,t)\ \mathrm{d}(x,t) + \int_{\mathbb{D}} \underbrace{\rho(x,T)\phi(x,T)}_{=0} - \underbrace{\rho(x,0)}_{=\rho_0(x) \text{ auf } \mathbb{D}}\ \phi(x,0)\,\mathrm{d}x \\
&= -\int_{\mathbb{D}\times\mathbb{T}} \rho(x,t)\partial_t\phi(x,t)\ \mathrm{d}(x,t) - \int_{\mathbb{D}} \rho_0(x)\phi(x,0)\,\mathrm{d}x
\end{aligned}$$

Außerdem können wir Integral (2) mit 2.3 wie folgt ausdrücken:

$$\begin{aligned}
\int_{\mathbb{D}\times\mathbb{T}} \operatorname{div}(\rho(x,t)q(x))\phi(x,t)\ \mathrm{d}(x,t) &= \int_{\mathbb{T}}\int_{\mathbb{D}} \operatorname{div}(\rho(x,t)q(x))\phi(x,t)\,\mathrm{d}x\,\mathrm{d}t \\
&= \int_{\mathbb{T}} \Big(-\int_{\mathbb{D}} \rho(x,t)q(x)\nabla\phi(x,t)\,\mathrm{d}x + \int_{\partial\mathbb{D}} \rho(x,t)q(x)\cdot n\ \phi(x,t)\,\mathrm{d}a \Big)\,\mathrm{d}t \\
&= -\int_{\mathbb{D}\times\mathbb{T}} \rho(x,t)q(x)\nabla\phi(x,t))\ \mathrm{d}(x,t) + \int_{\mathbb{T}}\int_{\partial\mathbb{D}} \rho(x,t)q(x)\cdot n\phi(x,t)\,\mathrm{d}a\,\mathrm{d}t
\end{aligned}$$

Wenn ρ eine klassische Lösung ist, dann existieren insbesondere $\partial_t\rho$ und $\operatorname{div}(\rho q)$ und obige Umformungen sind zulässig, d.h. ρ erfüllt auch die schwache Formulierung.
Ist ρ hingegen eine schwache Lösung die zusätzlich in $C^2(\mathbb{D}\times\mathbb{T},\mathbb{R})$ liegt, so lassen sich alle oben durchgeführten Umformungen auch in die andere Richtung durchführen und wir erhalten:

$$\int_{\mathbb{D}\times\mathbb{T}} (\partial_t\rho + \operatorname{div}(\rho q))\phi\ \mathrm{d}(x,t) = 0 \quad \forall \phi \in H_0^1(\mathbb{D}\times\mathbb{T})$$

Dann folgt mit 2.6, dass ρ auch die ursprüngliche Differentialgleichung
$\partial_t\rho(x,t) + \operatorname{div}(\rho(x,t)q(x)) = 0$ erfüllt. Somit ist ρ also auch klassische Lösung.

□

Auch die Semidiskretisierung können wir in ähnlicher Form, wie eben noch die schwa-

che Formulierung, ausdrücken. Sei dazu

$$B_h(\rho_h, \phi) \coloneqq \sum_{i=1}^{N} (\partial_t \rho_h, \phi_h)_{K_i} - \left(\sum_{i=1}^{N} (\Upsilon(\rho_h), \nabla \phi_h)_{K_i} - \sum_{\substack{F \in \mathcal{F}_{K_i} \\ F \not\subseteq \Gamma_{\text{in}}}} (\Upsilon^\star(\rho_h) \cdot n^K, \phi_h)_F \right)$$

$$\langle l_h, \phi_h \rangle \coloneqq (\rho_{\text{in}} q \cdot n^K, \phi_h)_{\partial K_i \cap \Gamma_{\text{in}}}$$

Dann löst $\rho_h \in Q_h$ die Semidiskretisierung 5.3, wenn $B_h(\rho_h, \phi) = \langle l_h, \phi_h \rangle$ für alle $\phi_h \in Q_h$ erhalten ist.

5.5.2 Konsistenz

Der letzte Abschnitt hat sich damit beschäftigt, aus dem ursprünglich unendlich dimensionalen Problem letztendlich eine volldiskretisierte Verfahrensvorschrift in einem endlichen Ansatzraum herzuleiten. Fragen wir nun nach der Konsistenz des Verfahren, stellen wir damit zugleich die Frage, ob wir immer noch die richtige Gleichung lösen. Genauer heißt das Verfahren genau dann konsistent, wenn eine analytische Lösung ρ des ursprünglichen Problems (dTP) auch die hergeleitete Verfahrensvorschrift erfüllt. Wir betrachten zunächst etwas abstrakter den formalen Prozess der Diskretisierung einer abstrakten Gleichung $T\rho = 0$:

Abbildung 2: Diskretisierungsprozess

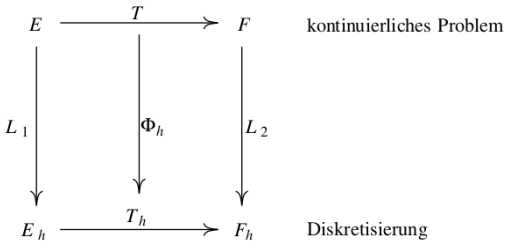

Abbildung aus [5] Seite 392

Das diskretisierte Problem $T_h \rho_h = 0$ heißt genau dann konsistent, wenn für eine analytische Lösung $\rho^\star \in E$ gilt, dass $\lim_{h \to 0} \|\Phi_h(T) L_1 \rho^\star - L_2 T \rho^\star\|_{F_h} = 0$. Betrachten wir zunächst in einem Zwischenschritt die Semidiskretisierung (5.3). Die gewählte Herleitung dieser Formulierung soll dabei nahelegen, dass für eine exakte (und glatte) Lösung ρ die Konsistenz für die Semidiskretisierung erfüllt ist. In oben eingeführter Notation gilt also gerade $B_h(\rho^\star, \phi)$ für alle Testfunktionen ϕ. Auch wenn es sich bei Q_h um einen nicht konformen Ansatzraum handelt, ist der Herleitung der Semidiskretisierung dennoch zu entnehmen, dass auch $B_h(\rho^\star, \phi_h)$ für alle $\phi_h \in Q_h$ gilt $(\star)$. Wichtig ist dabei unter anderem auch die Wahl der numerischen Flussfunktion, der upwind flux erhält aber gerade gewünschten Eigenschaften. Mehr dazu findet sich in [16]. Für die Zeitdiskretisierung in Form der impliziten Mittelpunktsregel gilt als einstufige Gauß-Quadratur

$\|\Phi_h(T)L_1\rho^\star - L_2T\rho^\star\|_{F_h} = O(h^2)$, womit sie insbesondere konsistent ist.
Daher gilt so für das kombinierte Verfahren, dass eine klassische Lösung $\rho^\star$ somit auch die Volldiskretisierung in obigem Sinne erfüllt.

5.5.3 Galerkin Orthogonalität

Eine direkte Folgerung aus der Konsistenz und $(\star)$ stellt die Galerkin Orthogonalität dar. Für eine Lösung der Semidiskretisierung ρ_h und eine analytische Lösung im klassischen Sinne $\rho^\star$ gilt dann nämlich:

$$B_h(\rho^\star - \rho_h, \phi_h = 0) \quad \forall \phi_h \in Q_h$$

5.5.4 Stabilität und Konvergenz

Die Stabilität ist bei der numerischen Lösung der Transportgleichung einer der wesentlichen Gründe, weswegen wir das Discontinuous Galerkin Verfahren einem (Standard-) Finite-Elemente-Ansatz vorziehen. Es zeigt sich nämlich, dass ein normales Finite Elemente Verfahren, wie wir es zuvor bei der Lösung des Potentialströmungsproblems genutzt hatten, beim Transportproblem instabil ist. Grund dafür ist, dass $\|q \cdot \nabla\rho_h\|$ beliebig groß werden kann. Für die DG-Diskretisierung mit upwind flux kann hingegen gezeigt werden, dass die Lösung des Transportproblems stabil ist. Auf eine theoretische Stabilitätsanalyse möchten wir aber an dieser Stelle verzichten und verweisen z.B. auf [16]. Ebenso wollen wir bezüglich der Konvergenz des Verfahrens auf entsprechende Literatur verweisen. Wir werden später Konvergenzannahmen stellen, und diese in entsprechenden Experimenten verifizieren. In der Literatur finden sich aber auch theoretische Konvergenzbeweise, meist unter Verwendung vielfältiger funktionalanalytischer Grundlagen und unter gewissen Regularitätsvoraussetzungen an die bestimmte Lösung. Die theoretische Betrachtung von Discontinuous Galerkinverfahren ist bereits Gegenstand vielfältiger wissenschaftlicher Arbeiten und aber zugleich ein breites Feld, welches noch lange nicht vollständig erforscht ist.
Im nächsten Abschnitt sollen nun schließlich die Überlegungen der letzten Abschnitte gebündelt werden und wir wollen die Multilevel Monte Carlo Methode bei der speziellen Anwendung auf das Transportproblem betrachten.

6 Anwendung der Multilevel Monte Carlo Methode auf das Transportproblem

6.1 Noch einmal Monte Carlo

6.2 Multilevel Monte Carlo

7 Experiment

8 Ausblick und Fazit

9 Appendix

9.1 Zusammenhang zwischen multivariater Normalverteilung und Normalverteilung

Satz 9.1. Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $X = (X_1, \ldots, X_n)$ ein Zufallsvektor mit (nicht entarteter) multivariater Normalverteilung mit Parametern $\mu = (\mu_1, \ldots, \mu_n) \in \mathbb{R}^n$ und $C = (\sigma_{ij})_{1 \le i,j \le n} \in \mathbb{R}^{n \times n}$.
Dann ist $\mathbb{E}[X] = \mu$ und für alle $i, j \in 1, \ldots, n$ gelten:

$$X_j \sim N(\mu_j, \sigma_{jj}) \text{ und } \sigma_{ij} = \mathrm{Cov}(X_i, X_j)$$

Beweis. (fasst mehrere Resultate aus [5] zusammen)

Da C symmetrisch positiv definit ist, existiert ein invertierbares $A \in \mathbb{R}^{n \times n}$ mit $C = AA^\top$ (Cholesky-Zerlegung). Weiter sei $Y = (Y_1, \ldots, Y_n)^\top$ ein Zufallsvektor, wobei die einzelnen $Y_1, \ldots, Y_n$ unabhängige und je $N(0,1)$-verteilte Zufallsvariablen sind. Durch $T(x) \coloneqq Ax + \mu$ erhalten wir somit für $x \in \mathbb{R}^k$ eine stetig differenzierbare Abbildung die den $\mathbb{R}^k$ auf sich selbst abbildet und die Funktionaldeterminante $\det A$ besitzt. Ist Y nun ein n-dimensionaler Zufallsvektor mit Dichte f, so besitzt der Zufallsvektor $Z \coloneqq AY + \mu$ nach dem Transformationssatz die Dichte

$$g(y) = \frac{f(A^{-1}(y-\mu))}{|\det A|}, \quad y \in \mathbb{R}^k.$$

Wir erhalten also mit

$$f(x) = \prod_{j=1}^{n} \left(\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x_j^2}{2} \right) \right) = \frac{1}{(2\pi)^{\frac{n}{2}}} \exp\left(-\frac{x^\top x}{2} \right), \text{ für } x \in \mathbb{R}^n$$

$$g(y) = \frac{1}{(2\pi)^{\frac{n}{2}} |\det A|} \exp\left(-\frac{1}{2} \left(A^{-1}(y-\mu) \right)^\top \left(A^{-1}(y-\mu) \right) \right), \text{ für } y \in \mathbb{R}^n.$$

Wegen $C = AA^\top$, $(A^{-1})^\top = (A^\top)^{-1}$ und $|\det A| = \sqrt{\det C}$ ist somit

$$g(y) = \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{\det C}} \exp\left(-\frac{1}{2} (y-\mu)^\top C^{-1} (y-\mu) \right), \text{ für } y \in \mathbb{R}^n.$$

. Insbesondere ist also $Z \sim X \sim N_n(\mu, C)$.
Seien nun $A = (a_{ij})_{1 \le i,j \le n}$, dann folgt

$$X_j \sim \sum_{l=1}^{n} a_{jl} Y_l + \mu_j \ .$$

Wegen $K_l \coloneqq a_{ij} Y_l \sim N(0, a_{jl}^2)$ und der Unabhängigkeit der Y_l (mit dem sogennanten Blockungslemma folgt somit Unabhängigkeit der K_l) gilt nach dem Additionsgesetz für die Normalverteilung

$$X_j \sim N\left(\mu_j, \sum_{l=1}^{n} a_{jl}^2\right).$$

Aus $C = AA^\top$ folgt schließlich $\sigma_{jj} = \sum_{l=1}^{n} a_{jl}^2$. Es bleibt nun also noch zu zeigen, dass $\mathbb{E}[X] = \mu$ und $\sigma_{ij} = \mathrm{Cov}(X_i, X_j)$. Wir bezeichnen mit $\mathrm{Cov}(X) \coloneqq (\mathrm{Cov}(X_i, X_j))_{1 \le i,j \le n}$ die Kovarianzmatrix. Es ist $\mathbb{E}[Y] = 0$ und $\mathrm{Cov}(Y) = I_n$. Es gilt also:

$$\mathbb{E}[X] = \mathbb{E}[AY + \mu] = (\mathbb{E}[\sum_{l=1}^{n} K_l + \mu_j]) = (\mathbb{E}[\sum_{l=1}^{n} a_{jl} Y_l + \mu_j]) = A\mathbb{E}[Y] + \mu = \mu$$

$$\mathrm{Cov}(X) = \mathrm{Cov}(AY + \mu) = \mathrm{Cov}(AY) = A\mathrm{Cov}(Y)A^\top = AA^\top = C$$

□

9.2 Referenzzelle und Hybridisierung

9.2.1 Referenzzelle

Bevor wie uns an dieser Stelle der Hybridisierung widmen können, wollen wir noch kurz auf einen wichtigen Aspekt der Implementierung finiter Elemente eingehen. An dieser Stelle hat es sich nämlich bereits oft als nützlich erwiesen, eine sogenannte Referenzzelle einzuführen. Statt sich die Daten jeder Zelle statisch zu speichern und dann darauf zuzugreifen, gehen wir dabei stets von der Referenzzelle aus und können über je eine linear affine Abbildung in den tatsächlichen Zellen operieren.

Definition 9.2. Das Referenzdreieck $\triangle$ ist definiert als

$$\hat{K} \coloneqq \mathrm{conv}\{\hat{\mathcal{V}}\}, \text{ wobei } \hat{\mathcal{V}} \coloneqq \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}$$

Die Seiten des Referenzdreiecks sind

$$\hat{F}_0 \coloneqq \mathrm{conv} \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\}$$

$$\hat{F}_1 \coloneqq \mathrm{conv} \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}$$

$$\hat{F}_2 \coloneqq \mathrm{conv} \left\{ \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right\}$$

die Seitenbasis ist gegeben durch

$$\hat{\psi}_0 : \hat{K} \to \mathbb{R}^2, \hat{\psi}_0(\xi) \coloneqq \begin{pmatrix} \xi_1 \\ \xi_2 - 1 \end{pmatrix}$$
$$\hat{\psi}_1 : \hat{K} \to \mathbb{R}^2, \hat{\psi}_1(\xi) \coloneqq \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}$$
$$\hat{\psi}_2 : \hat{K} \to \mathbb{R}^2, \hat{\psi}_2(\xi) \coloneqq \begin{pmatrix} \xi_1 - 1 \\ \xi_2 \end{pmatrix}.$$

und $\hat{n}$ sei der äußere Normalenvektor von $\hat{K}$

Abbildung 3: Referenzzelle

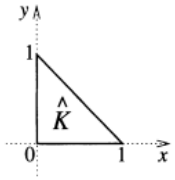

Abbildung aus [20] Seite 51

Bemerkung. $\forall i, j \in \{1, 2, 3\} : \int_{F_j} \hat{\psi}_i \cdot \hat{n} \, \mathrm{d}a = \delta_{i,j}$ und $\hat{\psi}_i \in \mathbb{P}_1(\hat{K}, \mathbb{R}^2)$.

Weiter setzen wir noch

$$\text{die Menge der Seiten} \qquad \hat{\mathcal{F}} \coloneqq \left\{\hat{F}_0, \hat{F}_1, \hat{F}_2\right\}$$
$$\text{und den Seitenansatzraum} \qquad \hat{W} \coloneqq \operatorname{span}\{\psi_0, \psi_1, \psi_2\}.$$

Transformation von $\hat{K}$ zu K: Für ein beliebiges $K \in \mathcal{K}$ wollen wir jetzt eine Seitenbasis $\{\psi_1^K, \psi_2^K, \psi_3^K\}$ berechnen (Wie bisher gegeben durch $\forall i \in \{1, 2, 3\} : \psi_i^K \in \mathbb{P}_1(K, \mathbb{R}^2)$ und $\int_{F_j^K} \psi_i^K \cdot n^K \, \mathrm{d}a = \delta_{i,j}$, wobei n^K äußere Normale von K und F_j^K beliebige Seite von K). Dazu betrachten wir die affine Transformationsabbildung φ_K von $\hat{K}$ zu K:

$$\varphi_K : \hat{K} \to K, \varphi_K(\xi) = z_{0,K} + B_K \xi \text{ mit passenden } B_K \in \mathbb{R}^{2\times 2} \text{ und}$$
$$J_K \coloneqq \det(B_K) > 0.$$

Abbildung 4: Affine lineare Transformation von $\hat{K}$ nach K

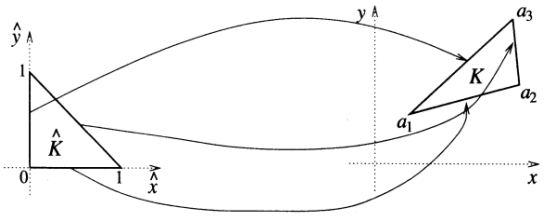

Abbildung aus [20] Seite 53

Lemma 9.3. Es gilt: $\tilde{n}^K = \frac{1}{|B_K^{-T}\hat{n}|} B_K^{-T}\hat{n}$ ist Normale zu ∂K.

Die Seitenbasis auf K ist dann gegeben durch

$$\psi_i^K = J_K^{-1} B_K \hat{\psi}_i \circ \varphi_K \ (i \in \{1,2,3\})$$

Die globale Seitenbasis $\{\psi_j\}_{j=1}^{|\mathcal{F}|}$ auf $\mathbb{D}$ erhalten wir dann mithilfe einer weiteren Abbildung l, die zwischen der Seitennummerierung in einer Zelle K und der globalen Seitennummerierung vermittelt. Es ist dabei

$$l : \mathcal{K} \times \{1,2,3\} \to \{1,\dots,|\mathcal{F}|\}, (K,i) \mapsto l(K,i).$$

Wir setzen nun also $\psi_j (j \in \{1,\dots,|\mathcal{F}|\})$ durch

$$\psi_j(x) = \begin{cases} \psi_i^K(x), & \text{falls } j = l(K,i) \\ 0, & \text{sonst.} \end{cases}$$

Bemerkung. Für alle Zellen $K \in \mathcal{K}$ von denen F_j eine anliegende Seite ($\overline{K} \cap F_j \neq \emptyset$) und F_j lokal mit $i \in \{1,2,3\}$ nummeriert ist, gilt:

$$\psi_j|_K = \psi_i^K.$$

9.2.2 Hybridisierung

Wir betrachten die Räume

$$W_K \coloneqq \left\{ \psi_K : K \to \mathbb{R}^2 : \psi_K = J_K^{-1} B_K \hat{\psi} \circ \varphi_K^{-1}, \hat{\psi} \in \hat{W} \right\}$$

$$W_{\mathcal{K}} \coloneqq \prod_{K \in \mathcal{K}} W_K, \qquad M_h \coloneqq \prod_{F \in \mathcal{F}} \mathbb{P}_0(F)$$

$$M_h(u_D) \coloneqq \left\{ \mu_h \in M_h : \forall F \subset \Gamma_D \int_F \mu_h \,\mathrm{d}a = \int_F u_D \,\mathrm{d}a \right\}$$

Bemerkung.

$$\psi_h \in W_h \iff \left[\psi_h \in W_{\mathcal{K}} \text{ und } (\psi_{K_1} - \psi_{K_2}) \cdot n^F = 0 \ (F = \partial K_1 \cap \partial K_2 \in \mathcal{F}^\circ)\right]$$

Und untersuchen folgendes Problem:

$$\text{Bestimmme } (q_h, u_h, \lambda_h) \in W_{\mathcal{K}} \times Q_h \times M_h(u_D) \text{ mit}$$
$$\begin{cases} (1) \displaystyle\int_K \kappa^{-1} q_h \psi_K \,\mathrm{d}x - \int_K u_h \operatorname{div}(\psi_K) \,\mathrm{d}x = -\int_{\partial K} \lambda_h \psi_K \cdot n^K \,\mathrm{d}a \\ (2) \displaystyle\int_K \operatorname{div}(q_h) \phi_K \,\mathrm{d}x = 0 \\ (3) \displaystyle\sum_{K \in \mathcal{K}} \int_{\partial K} q_h \cdot n \mu_h \,\mathrm{d}a = -\int_{\Gamma_N} g_N \mu_h \,\mathrm{d}a \end{cases}$$
$$\text{für alle } K \in \mathcal{K}, \psi_K \in W_K, \phi_K \in Q_h \text{ und } \mu_h \in M_h(0)$$

Dieses Problem ist äquivalent zu dem diskreten gemischten FE-Problem, welches wir zuvor betrachtet haben:

$$\text{Bestimme } (q_h, u_h) \in W_h(-g_N) \times Q_h \text{ mit}$$
$$\begin{cases} \int_\Omega \kappa^{-1} q_h \cdot \psi_h \,\mathrm{d}x - \int_\Omega u_h \operatorname{div}(\psi_h) \,\mathrm{d}x = -\int_{\Gamma_D} u_D \psi_h \cdot n \,\mathrm{d}a \\ -\int_\Omega \operatorname{div}(q_h) \phi_h \,\mathrm{d}x = 0 \end{cases}$$
$$\text{für alle } (\psi_h, \phi_h) \in W_h(0) \times Q_h$$

Für ein festes $K \in \mathcal{K}$ ergibt sich mit der Wahl einer Basis von W_K, Q_h und M_h eine Formulierung als LGS mit Nebenbedingung, wobei $\underline{q}_K \coloneqq \underline{R}_K \underline{q}$, $\underline{u}_K \coloneqq \underline{R}_K \underline{u}$.

$$\text{Bestimme } \underline{q}, \underline{u} \text{ und } \underline{\lambda} \text{ mit}$$
$$\begin{cases} (1) \begin{pmatrix} \underline{A}_K & \underline{B}_K \\ \underline{B}_K^T & 0 \end{pmatrix} \begin{pmatrix} \underline{q}_K \\ \underline{u}_K \end{pmatrix} = \begin{pmatrix} -\underline{C}_K \underline{R}_K \underline{\lambda} \\ 0 \end{pmatrix} \\ (2) \displaystyle\sum_{K \in \mathcal{K}} (\underline{R}_K \underline{\mu})^T \underline{C}_K \underline{q}_K \ = \underline{\mu}^T \underline{b} \end{cases}$$
$$\text{für alle } \underline{\mu} \text{ mit } \underline{\mu}[F] = 0 \text{ für } F \in \Gamma_D \cap \mathcal{F}$$

$$\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 1 \\ \underline{\mu} \end{pmatrix}^T \underbrace{\left(\begin{array}{cccccc|c} \underline{A}_{K_1} & \underline{B}_{K_1} & & & & & \underline{C}_{K_1}\underline{R}_{K_1} \\ \underline{B}_{K_1} & 0 & & & & & 0 \\ & & \underline{A}_{K_2} & \underline{B}_{K_2} & & & \underline{C}_{K_2}\underline{R}_{K_2} \\ & & \underline{B}_{K_2} & 0 & & & 0 \\ & & & & \ddots & & \\ \hline \underline{R}_{K_1}^T\underline{C}_{K_1}^T & 0 & \underline{R}_{K_2}^T\underline{C}_{K_2}^T & 0 & & & 0 \end{array}\right)}_{=:\left(\begin{array}{c|c} \underline{D} & \underline{E} \\ \hline \underline{E}^T & 0 \end{array}\right)} \underbrace{\begin{pmatrix} \underline{q}_{K_1} \\ \underline{u}_{K_1} \\ \underline{q}_{K_2} \\ \underline{u}_{K_2} \\ \vdots \\ \underline{\lambda} \end{pmatrix}}_{=:\begin{pmatrix} \begin{pmatrix} \underline{q}_{K_i} \\ \underline{u}_{K_i} \end{pmatrix}_{K_i \in \mathcal{K}} \\ \underline{\lambda} \end{pmatrix}} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ \underline{\mu}^T\underline{b} \end{pmatrix}$$

Mit dem Schurkomplement $\underline{S} \coloneqq \underline{E}^T\underline{D}^{-1}\underline{E}$ folgt

$$\underline{\mu}^T\underline{S}\,\underline{\lambda} = \underline{\mu}^T\underline{b} \text{ für alle } \underline{\mu} \text{ mit } \underline{\mu}[F] = 0 \text{ für } F \in \Gamma_D \cap \mathcal{F}$$

Sobald wir $\underline{\lambda}_k \coloneqq \underline{R}_K\underline{\lambda}$ bestimmt haben, können wir auch das obere LGS (1) lösen um $\underline{q}_K$ und $\underline{u}_K$ zu erhalten.

Literatur

[1] M++ (meshes, multigrid and more). `http://www.math.kit.edu/ianm3/page/mplusplus/de`. Accessed: 2019-10-17.

[2] D. Braess. *Finite Elemente: Theorie, schnelle Löser und Anwendungen in der Elastizitätstheorie*. Springer-Verlag, 2013.

[3] S. Brenner and R. Scott. *The mathematical theory of finite element methods*, volume 15. Springer Science & Business Media, 2007.

[4] F. Brezzi and M. Fortin. *Mixed and hybrid finite element methods*, volume 15. Springer Science & Business Media, 2012.

[5] M. Brokate, N. Henze, F. Hettlich, A. Meister, G. Schranz-Kirlinger, and T. Sonar. *Grundwissen Mathematikstudium*. Springer, 2016.

[6] B. Cockburn, S. Hou, and C.-W. Shu. The runge-kutta local projection discontinuous galerkin finite element method for conservation laws. iv. the multidimensional case. *Mathematics of Computation*, 54(190):545–581, 1990.

[7] B. Cockburn, G. E. Karniadakis, and C.-W. Shu. The development of discontinuous galerkin methods. In *Discontinuous Galerkin Methods*, pages 3–50. Springer, 2000.

[8] B. Cockburn and C.-W. Shu. Tvb runge-kutta local projection discontinuous galerkin finite element method for conservation laws. ii. general framework. *Mathematics of computation*, 52(186):411–435, 1989.

[9] B. Cockburn and C.-W. Shu. The runge–kutta discontinuous galerkin method for conservation laws v: multidimensional systems. *Journal of Computational Physics*, 141(2):199–224, 1998.

[10] B. Cockburn and C.-W. Shu. Runge–kutta discontinuous galerkin methods for convection-dominated problems. *Journal of scientific computing*, 16(3):173–261, 2001.

[11] G. De Marsily. Quantitative hydrogeology. Technical report, Paris School of Mines, Fontainebleau, 1986.

[12] M. Dobrowolski. *Angewandte Funktionalanalysis: Funktionalanalysis, Sobolev-Räume und elliptische Differentialgleichungen*. Springer-Verlag, 2010.

[13] A. Ern and J.-L. Guermond. Theory and practice of finite elements. 2004. *Applied Mathematical Sciences*, 2004.

[14] L. C. Evans. *Partial differential equations*. American Mathematical Society, Providence, R.I., 2010.

[15] M. Hanke-Bourgeois. *Grundlagen der numerischen Mathematik und des wissenschaftlichen Rechnens*, volume 1. Springer, 2002.

[16] R. Hartmann. Numerical analysis of higher order discontinuous Galerkin finite element methods. In H. Deconinck, editor, *VKI LS 2008-08: CFD - ADIGMA course on very high order discretization methods, Oct. 13-17, 2008*. Von Karman Institute for Fluid Dynamics, Rhode Saint Genèse, Belgium, 2008.

[17] S. Heinrich. *Random approximation in numerical analysis*. Universität Kaiserslautern. Fachbereich Informatik, 1992.

[18] S. Heinrich. Multilevel monte carlo methods. In *International Conference on Large-Scale Scientific Computing*, pages 58–67. Springer, 2001.

[19] A. Klenke. *Wahrscheinlichkeitstheorie*, volume 1. Springer, 2006.

[20] P. Knabner and L. Angermann. *Numerik partieller Differentialgleichungen: eine anwendungsorientierte Einführung*. Springer-Verlag, 2013.

[21] P. Kumar, P. Luo, F. J. Gaspar, and C. W. Oosterlee. A multigrid multilevel monte carlo method for transport in the darcy–stokes system. *Journal of Computational Physics*, 371:382–408, 2018.

[22] B. Lapeyre, E. Pardoux, and R. Sentis. *Introduction to Monte Carlo methods for transport and diffusion equations*, volume 6. Oxford University Press on Demand, 2003.

[23] P. Lesaint and P.-A. Raviart. On a finite element method for solving the neutron transport equation. *Publications mathématiques et informatique de Rennes*, (S4):1–40, 1974.

[24] T. E. Peterson. A note on the convergence of the discontinuous galerkin method for a scalar hyperbolic equation. *SIAM Journal on Numerical Analysis*, 28(1):133–140, 1991.

[25] W. H. Reed and T. Hill. Triangular mesh methods for the neutron transport equation. Technical report, Los Alamos Scientific Lab., N. Mex.(USA), 1973.

[26] G. R. Richter. An optimal-order error estimate for the discontinuous galerkin method. *Mathematics of Computation*, 50(181):75–88, 1988.

[27] J. E. Roberts and J.-M. Thomas. Mixed and hybrid methods. 1991.

[28] T. J. Sullivan. *Introduction to uncertainty quantification*, volume 63. Springer, 2em015.

Erklärung

Ich versichere wahrheitsgemäß, die Arbeit selbstständig verfasst, alle benutzten Hilfsmittel vollständig und genau angegeben und alles kenntlich gemacht zu haben, was aus Arbeiten anderer unverändert oder mit Abänderungen entnommen wurde, sowie die Satzung des KIT zur Sicherung guter wissenschaftlicher Praxis in der jeweils gültigen Fassung beachtet zu haben.

Ort, den Datum