

Multiskalen-basierte
Finite-Differenzen–Verfahren
auf adaptiven dünnen Gittern

Dissertation

zur

Erlangung des Doktorgrades (Dr. rer. nat.)

der

Mathematisch–Naturwissenschaftlichen Fakultät

der

Rheinischen Friedrich–Wilhelms–Universität Bonn

vorgelegt von

Frank Koster

aus

Neu-Buch

Bonn, November 2001

Angefertigt mit Genehmigung der Mathematisch–Naturwissenschaftlichen Fakultät der
Rheinischen Friedrich–Wilhelms–Universität Bonn

1. Referent: Prof. Dr. Michael Griebel

2. Referent: Prof. Dr. Angela Kunoth

Tag der Promotion:

Für meine Eltern.

Inhaltsverzeichnis

1	Einleitung	1
2	Univariate Multiskalenfunktionen	10
2.1	Biorthogonale Multiskalenfunktionen	10
2.2	Glattheit, polynomiale Exaktheit und Normäquivalenzen	13
2.3	Verallgemeinerte Hierarchische Basen und Interpolets	16
2.4	Interpolets auf dem Intervall	18
2.5	Lifting-Wavelets	23
3	Adaptive Algorithmen in 1D	26
3.1	Multiskalentransformation	27
3.2	Bedingungen an \mathcal{T}	29
3.3	Lifting-Wavelettransformation	33
3.4	Petrov-Galerkin-Diskretisierung	37
3.5	Randbedingungen beim Petrov-Galerkin-Verfahren	41
3.6	Finite-Differenzen	43
4	Multivariate Multiskalenfunktionen	48
4.1	Isotrope Verfeinerung	49
4.2	Anisotrope Verfeinerung	51
5	Approximation von Funktionen	56
5.1	Sobolev-Räume und lineare Approximation	57
5.2	Besov-Räume und nichtlineare Approximation	58
6	Adaptive Diskretisierungen	66
6.1	Kollokationsmethode	67
6.2	Randbedingungen bei der Kollokationsmethode	68
6.3	Finite-Differenzen-Diskretisierung	71
6.4	Vergleich verschiedener Diskretisierungstechniken	72
6.5	Konsistenz für partielle Ableitungsoperatoren D^{re_i}	74
7	Vorkonditionieren von elliptischen Gleichungssystemen	86
7.1	Diagonal- und Lifting-Vorkonditionierer	86
7.2	Numerische Experimente	93
8	Elliptische Gleichungen	97

8.1	Verfeinerungsstrategien für Galerkin-Verfahren	97
8.2	Verfeinerungsstrategien für Dünngitter-Diskretisierungen	101
8.3	Numerische Experimente I	104
8.4	Numerische Experimente II	107
9	Erhaltungsgleichungen	110
9.1	Zeitschritt-Verfahren und Verfeinerungsstrategien	111
9.2	Numerische Experimente	112
10	Navier-Stokes Gleichungen	118
10.1	Druck-Korrektur-Verfahren	119
10.2	Stabile Diskretisierung der Druck-Poisson-Gleichung	121
10.3	Numerische Experimente: 2D Scherströmung	126
10.4	Numerische Experimente: 3D Scherströmung	128
11	Zusammenfassung und Ausblick	134
A	Beweise und Nebenrechnungen	137
A.1	Beweis für Theorem 2.1	137
A.2	Beweis für Korollar 2.1	138
A.3	Beweis für Theorem 3.1	138
A.4	Beweis für Lemma 3.2	139
A.5	Beweis für Lemma 3.3	140
A.6	Beweis für Lemma 3.4	140
A.7	Beweis für Lemma 3.5	141
A.8	Beweis für Lemma 3.8	142
A.9	Beweis für Lemma 3.9	142
A.10	Beweis für Lemma 4.3	143
A.11	Sobolev-/Besov-Regularität von $ \mathbf{x} ^\beta$	144
A.12	Abschätzungen für die Punktsingularität	145
A.13	Beweis für Lemma 6.2	145
A.14	Beweis für Lemma 6.5	147
A.15	Beweis für Lemma 6.6	147
A.16	Beweis für Theorem 7.1	148
A.17	Beweis für Lemma 8.1	151
B	Notation	152
	Literaturverzeichnis	156

Kapitel 1

Einleitung

Viele interessante Vorgänge, die in den Natur-, Wirtschafts- und Sozialwissenschaften untersucht werden, können durch partielle Differentialgleichungen beschrieben werden. Einige wichtige Beispiele dafür sind das zeitliche Verhalten des Erdklimas [DF97], die Ausbreitung von Nervenimpulsen [Pes76] oder die Bewertung von Aktienoptionen in finanzmathematischen Modellen [WDH97]. Eine numerische Simulation solcher Vorgänge besteht darin, eine approximative Lösung der zugrundeliegenden partiellen Differentialgleichung (PDGL) zu bestimmen. Dazu wird bei vielen numerischen Verfahren die Lösung direkt oder indirekt durch eine endliche Linearkombination von Ansatzfunktionen approximiert. Die PDGL wird auf dem entsprechenden Ansatzraum z.B. mittels eines Galerkin-Verfahrens diskretisiert, so daß man nach Lösen eines oder mehrerer linearer oder nichtlinearer Gleichungssysteme die gesuchte numerische Lösung erhält. Klarerweise ist der dafür notwendige Rechenaufwand mindestens proportional zur Dimension des Ansatzraumes. Ein Problem ist nun, daß bei vielen Anwendungen die gesuchten Lösungen sehr komplex sind, so daß eine große Anzahl von Ansatzfunktionen notwendig ist, um eine genaue Approximation zu erlauben. Größenordnungen von $10^6 \dots 10^9$ für die Zahl der Freiheitsgrade sind bei heutigen numerischen Verfahren durchaus üblich. Entsprechend hoch ist der Rechenaufwand, so daß selbst auf den schnellsten Computern Rechenzeiten von Wochen oder Monaten notwendig sind. Um den Rechenaufwand entscheidend zu verringern, bieten sich im wesentlichen zwei Lösungsmöglichkeiten an:

Man kann versuchen, schnelle Verfahren zu finden, bei denen der Rechenaufwand möglichst langsam mit der Zahl der Freiheitsgrade zunimmt. Dies ist insbesondere wichtig für das Lösen der linearen Gleichungssysteme, welche in der Regel den mit Abstand größten Anteil am Rechenaufwand stellen. In den letzten 30 Jahren sind hier mit den Mehrgitter-Verfahren entscheidende Fortschritte gemacht worden [Bra72, Hac78, BPX90, DK92, GO95a]. Mit diesen Verfahren ist es möglich, die besonders häufigen elliptischen linearen Gleichungssysteme mit einem Aufwand zu lösen, der proportional zur Anzahl der Unbekannten – also asymptotisch optimal – ist.

Die zweite Möglichkeit besteht darin, durch eine geschickte Auswahl der Ansatzfunktionen bereits mit relativ wenigen Freiheitsgraden die gewünschte Genauigkeit zu erreichen. Das klassische Beispiel sind hier die adaptiven Finite-Element-Methoden, bei denen die Ansatzfunktionen in Bereichen geringer Glattheit der Lösung konzentriert werden [BR78, ZKGB82, BJR95]. Da man in der Regel a priori nicht weiß, wo diese Bereiche sind, muß die Verteilung der Ansatzfunktionen durch einen iterativen Prozeß gesteuert werden. Dabei werden im einfachsten Fall entsprechend eines Verfeinerungskriteriums, das auf einer aktuellen Lösung basiert, zum bestehenden Ansatzraum neue Basisfunktionen hinzugefügt.

Auf diese Weise erhält man ganz automatisch eine gewisse Hierarchie von Ansatzfunktionen. Diese Beobachtung war einer der Ansatzpunkte für die Entwicklung von Multiskalenbasen [ZISC70, ZKGB82, Yse86]. Diese folgen einem ähnlichen Gedanken, wie auch die Mehrgitter-Verfahren. Bei Mehrgitter-Verfahren betrachtet man das Problem des *Lösens des Gleichungssystems* auf einer Hierarchie von Vergrößerungsleveln. Ausgangspunkt von Multiskalenbasen ist die *Approximation von Funktionen* auf unterschiedlichen Leveln [Mal89a].

Am Beispiel der Approximation einer Funktion $u : [0, 1] \rightarrow \mathbb{R}$ können wir die wesentlichen Eigenschaften von Multiskalenfunktionen und Unterschiede zu klassischen, nichtadaptiven Finite-Element-Ansatzfunktionen erläutern. Benutzt man für die Approximation von u z.B. Splines $\{\phi_i\}_{0 \leq i \leq n}$ $(N - 1)$ ten Grades auf einem uniformen Gitter mit Maschenweite n^{-1} , dann erhält man für $u \in \mathcal{H}^s$ mit $s \leq N$ bekanntlich die Abschätzung [Bra92]

$$\inf_{u_n \in \text{span}\{\phi_i\}} \|u - u_n\|_{\mathcal{L}^2} \lesssim n^{-s} \|u\|_{\mathcal{H}^s} .$$

Ist u sehr glatt, dann ist die Konvergenzrate der Approximation somit genau N . Enthält u aber Singularitäten, d.h. es gibt niederdimensionale Mannigfaltigkeiten, auf denen Funktionswerte oder Gradienten gegen unendlich gehen, dann hat u nur eine geringe Sobolev-Regularität, so daß die Konvergenzrate deutlich kleiner als N ist. In diesem Fall sind also für uniforme Ansatzräume viele Freiheitsgrade notwendig, um genaue Approximationen zu ermöglichen.

Multiskalenbasen sind hingegen gleichermaßen für die Approximation von glatten Funktionen und Funktionen mit Singularitäten geeignet. Im einfachsten Fall sind die Multiskalenfunktionen gestauchte und gestreckte Repliken $\psi_{lt}(x) = \psi(2^l x - t)$, $l \in \mathbb{N}, t \in \mathbb{N}$, einer Mutterfunktion ψ . Dabei ist l das Verfeinerungslevel und t der Translationsindex. In Bild 1.1 sind die Funktionen der ersten drei Level für zwei wichtige univariate Multiskalenbasen skizziert, die Haar-Basis [Haa10] (links) sowie die hierarchische Basis [Fab09] (rechts).

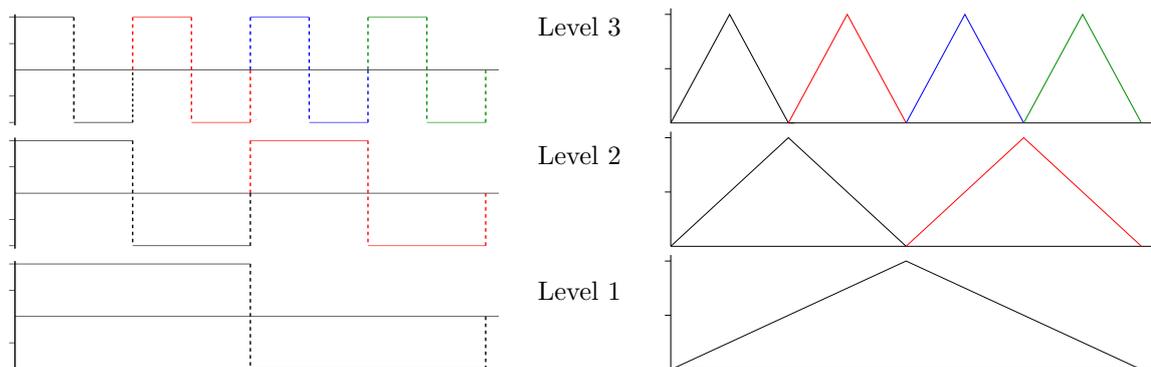


Abbildung 1.1: Links: die Haar-Basis. Rechts: die hierarchische Basis.

Die Approximationskraft von Multiskalenfunktionen beruht darauf, daß bereits mittels weniger Ansatzfunktionen von niedrigen Leveln, die einen großen Träger haben, die glatten Bereiche von u gut approximiert werden können. Wesentlich dafür ist, daß, ähnlich wie bei Splines, Polynome eines bestimmten Grades, z.B. $N - 1$, bereits mit den größten Basisfunktionen dargestellt werden können. Man sagt dann, daß die Basis den Grad N an polynomialer Exaktheit hat. Z.B. ist $N = 1$ für die Haar-Basis und $N = 2$ für die hierarchische Basis.

Eine gute Approximation von eventuellen Singularitäten wird durch zusätzliche Ansatzfunktionen auf höheren Leveln, die in Nähe der Singularitäten liegen, erreicht. Da Singularitäten auf niederdimensionalen Mannigfaltigkeiten leben, ist die Zahl der zusätzlich notwendigen Ansatzfunktionen nicht allzugroß. Bei Multiskalenbasen hat man also die Freiheit, die lokale Auflösung an die lokale Glattheit von u anzupassen, d.h. zu *adaptieren*. Für den Fall, daß dies bei vorgegebener Anzahl von Freiheitsgraden bestmöglich gelingt, spricht man von der Best- n -Term-Approximation. Die damit erzielbare Approximationsrate ist durch die Abschätzung

$$\inf_{\mathcal{T}:|\mathcal{T}|=n} \inf_{u_n \in \text{span}\{\psi_{it}\}_{it \in \mathcal{T}}} \|u - u_n\|_{\mathcal{L}^2} \lesssim n^{-s} \|u\|_{\mathcal{B}_{\gamma^0}^s}, \quad s \leq N$$

gegeben [Don92, FJ90]. Der Term $\|u\|_{\mathcal{B}_{\gamma^0}^s}$ bezeichnet die Norm von u in einem sogenannten *Besov*-Raum [Tri78], siehe auch Theorem 5.3. Der Punkt ist nun, daß Funktionen mit Singularitäten zwar eine kleine Sobolev-Regularität haben, aber eine sehr hohe Besov-Regularität haben können [DeV98, DD97]. In solchen Fällen ist die Best- n -Term-Approximationsrate gleich $N -$ genau wie bei glatten Funktionen! Abschätzungen für die optimalen Approximationsraten waren schon lange bekannt, bevor man den Zusammenhang zu Besov-Räumen erkannt hatte. Insbesondere für h - p -adaptive Finite-Element-Methoden [SB91], bei denen sowohl die Maschenweite h als auch der Polynomgrad p der Ansatzfunktionen lokal adaptiert wird, war man daran interessiert, wie die optimalen Ansatzräume für bestimmte Typen von Funktionen aussehen [GB86, BS94]. Für die entsprechenden Rechnungen wurden dabei lokale Taylor-Entwicklungen benutzt. Nichts anderes verbirgt sich hinter den Besov-Raum-Abschätzungen, wie wir in Kapitel 5 sehen werden.

Speziell für einfache mehrdimensionale Gebiete, wie den Einheitskubus $[0, 1]^d$, gibt es im wesentlichen zwei große Gruppen von Multiskalenbasen, die sich bezüglich der Approximationseigenschaften deutlich voneinander unterscheiden. Zum einen gibt es Basisfunktionen, deren Träger in jede Koordinatenrichtung ungefähr die gleiche Länge hat, welche durch das Verfeinerungslevel bestimmt ist. Ein Beispiel ist in Bild 1.2 dargestellt. Wir nennen

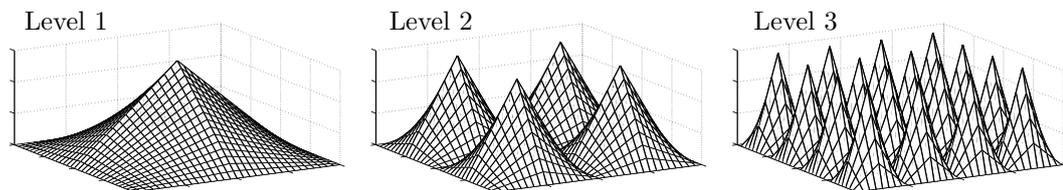


Abbildung 1.2: *Isotrope Variante der hierarchischen Basis auf den ersten drei Leveln.*

diese Basisfunktionen 'isotrop', obgleich sie natürlich nicht wirklich isotrop sind. Isotrope Multiskalenfunktionen werden von den meisten Autoren für z.B. das Lösen von PDGLen benutzt [Yse86, BK97, Dah97, Har94, Kun01, Sch98b]. Vom Standpunkt der Approximationstheorie sind sie besonders gut untersucht, insbesondere ihre Beziehung zu Besov-Räumen [DeV98, Hoc99].

Auf der anderen Seite gibt es Multiskalenfunktionen, die durch Tensorprodukte von univariaten Multiskalenfunktionen von beliebigen Leveln gebildet werden. Ein Beispiel für diese Konstruktion ist in Bild 1.3 zu sehen. Dabei wurde die univariate hierarchische Basis als Ausgangspunkt benutzt. Bei dieser Tensorprodukt-Konstruktion gibt es viele stark anisotrope Basisfunktionen, so daß wir die so konstruierte Basis etwas verallgemeinernd als 'anisotrop' bezeichnen können. Solche Basen haben zum Teil dramatisch bessere Approximationseigenschaften als die entsprechenden isotropen Basen. Dies ist zum Beispiel der Fall,

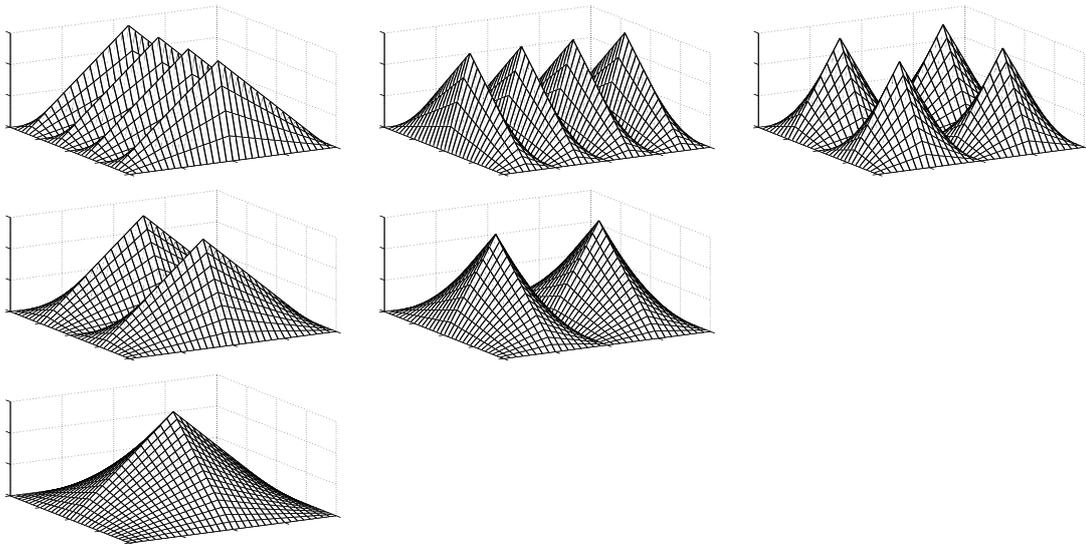


Abbildung 1.3: *Anisotrope Variante der multivariaten hierarchischen Basis.*

falls die zu approximierende Funktion eine anisotrope Glattheit hat [GOS99, Hoc99], z.B. bei Kanten- oder Flächensingularitäten, oder falls u beschränkte gemischte Ableitungen hat [GK99, Tem93, Zen91]. Angenommen für $u : [0, 1]^2 \rightarrow \mathbb{R}$ ist $\|\partial_{xxyy}u\|_{\mathcal{L}^2}$ beschränkt, dann erhält man für die isotrope Basis aus Bild 1.2 im allgemeinen die Abschätzung $\|u - u_n\|_{\mathcal{L}^2} \lesssim n^{-1}$ für die Best- n -Term Approximation. Für die anisotrope Basis aus Bild 1.3 kann jedoch $\|u - u_n\|_{\mathcal{L}^2} \lesssim (n/\log n)^{-2}$ gezeigt werden, d.h. die Approximationsrate ist nahezu doppelt so hoch. Um den Unterschied einmal graphisch zu veranschaulichen, tragen wir für eine bestimmte vorgegebene Genauigkeit $\epsilon > 0$ die Mittelpunkte der Träger der für die Best- n -Term-Approximation benutzten Ansatzfunktionen in einem Gitter ein. Im isotropen Fall erhält man im wesentlichen ein uniformes Gitter mit $O(\epsilon^{-1})$ Gitterpunkten (Bild 1.4 links), wogegen im anisotropen Fall ein ausgedünntes Gitter (Bild 1.4 rechts) mit $O(\epsilon^{-1/2} |\log \epsilon|)$ Punkten entsteht. Die entsprechenden Gitter nennt man deshalb auch *dünne Gitter* [Zen91] und die Approximationsräume *Dünngitter-Approximationsräume* [Kna00].

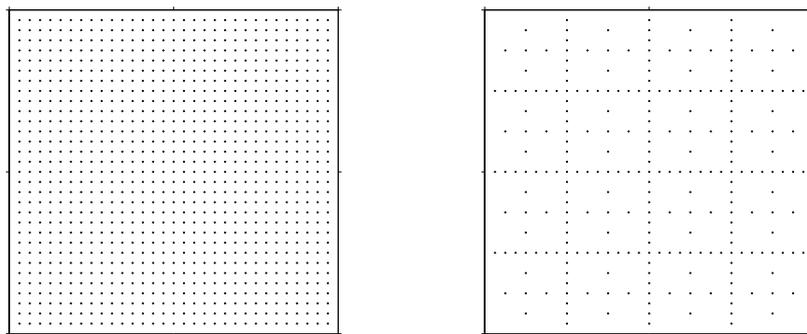


Abbildung 1.4: *Gitter zu den optimierten Ansatzräumen für isotrope Multiskalenfunktionen (links) bzw. anisotrope Tensorprodukt-Basen (rechts). Das rechte Gitter ist ein dünnes Gitter.*

Für allgemeine Funktionen aus Besov-Räumen ist bekannt, daß die anisotropen Basen asymptotisch mindestens fast genauso gut approximieren, wie die isotropen Basen [Hoc99]. Im

schlechtesten Fall taucht in den Approximationsabschätzungen gegenüber den isotropen Basen ein zusätzlicher, dimensionsabhängiger logarithmischer Faktor auf. In der Summe ihrer Eigenschaften sind die anisotropen Tensorprodukt-Multiskalenfunktionen aber wohl doch die aussichtsreicheren Ansatzfunktionen.

Unabhängig davon, welche Ansatzfunktionen man benutzt, gibt es in praktischen Anwendungen natürlich das Problem, wie die optimalen oder nahezu optimale Ansatzräume bestimmt werden können. Dies ist insbesondere beim Lösen von PDGLen der Fall, wo die zu approximierende Funktion selbst nicht bekannt ist. Die Idee ist hier, ausgehend von einer schon bekannten Näherung der Lösung mittels bestimmter Verfeinerungsstrategien einen neuen Ansatzraum zu bestimmen, der für die Approximation der wirklichen Lösung besser geeignet ist, als der bisherige. Mit dem neuen Ansatzraum wird eine neue numerische Lösung bestimmt und das Schema kann von vorn beginnen. Obgleich adaptive Verfahren und Verfeinerungsstrategien schon lange in Gebrauch sind [BR78, BS93, Ver96, BEK96] ist es erst in jüngster Zeit gelungen zu zeigen, daß die entsprechenden iterativen Schemen tatsächlich konvergieren [Doe96] bzw. sogar im Sinne der Best- n -Term-Approximation mit der optimalen Rate konvergieren [CDD01, DDU01]. Das wichtigste Hilfsmittel für diese Aussagen sind sogenannte Normäquivalenzen zwischen bestimmten Sobolev-Normen von Funktionen und gewichteten Folgen-Normen der zugehörigen Multiskalenkoeffizienten. Diese Beziehungen haben die Form

$$\left\| \sum_{l,t} u_{l,t} \psi_{l,t} \right\|_{\mathcal{H}^s}^2 \sim \sum_{l,t} |u_{l,t}|^2 \|\psi_{l,t}\|_{\mathcal{H}^s}^2. \quad (1.1)$$

Der Bereich an Sobolev-Räumen, für den solche Normäquivalenzen gelten, hängt von der Multiskalenbasis ab. Beispielsweise ist die Haar-Basis orthogonal bezüglich des \mathcal{H}^0 -Skalarproduktes, kann aber mangels \mathcal{H}^1 -Regularität natürlich (1.1) nicht für $s = 1$ erfüllen. Die hierarchische Basis wiederum erfüllt (1.1) für die \mathcal{H}^1 -Semi-Norm, jedoch nicht für \mathcal{H}^0 . In gewissem Sinne ist der Grund für die fehlende Normäquivalenz für $s = 0$, daß der Mittelwert der hierarchischen Basisfunktionen nicht verschwindet, wie wir z.B. in Abschnitt 2.2 sehen werden. Vor etwa 15 Jahren begann man parallel zu den Finite-Element-Entwicklungen mit der Untersuchung von Multiskalenfunktionen vom Standpunkt der Signalverarbeitung [GGM84] bzw. der Funktionalanalysis [LM86, Mey92]. Eine der zentralen Forderungen an die Basisfunktionen war dabei der verschwindende Mittelwert. Die untersuchten bzw. neu konstruierten Multiskalenfunktionen hatten somit eine oszillierende, wellenartige Form und wurden deshalb *Wavelets* (Wellchen) genannt. Letztlich hat sich die Forderung an den Mittelwert aber als restriktiv und auch nicht wirklich notwendig erwiesen, wie man am Beispiel der sehr nützlichen hierarchischen Basis sehen kann. Aus diesem Grund scheint es uns statthaft, hin und wieder den eingängigen Begriff Wavelets als Synonym für beliebige Multiskalenfunktionen zu benutzen.

Normäquivalenzen spielen weiterhin eine wichtige Rolle beim schnellen Lösen der linearen Gleichungssysteme, die bei der Diskretisierung von PDGLen entstehen. Man kann sich leicht überlegen, daß bei einer Galerkin-Diskretisierung der Poisson-Gleichung mit einer auf $\|\psi_{lt}\|_{\mathcal{H}^1} = 1$ normierten Multiskalenbasis die Steifigkeitsmatrix eine Konditionszahl von der Größenordnung 1 hat, falls (1.1) für $s = 1$ erfüllt ist. Mit z.B. einem CG-Verfahren benötigt man dann bekanntlich nur wenige Iterationen, um das Gleichungssystem zu lösen [Hac91, SB86]. In Kapitel 7 der vorliegenden Arbeit wird gezeigt, daß sich die Normäquivalenzen auch bei anderen Diskretisierungstechniken in diesem Sinne ausnutzen lassen. Multiskalenbasen führen also nicht nur zu sehr effizienten Approximationen sondern gegebenenfalls auch

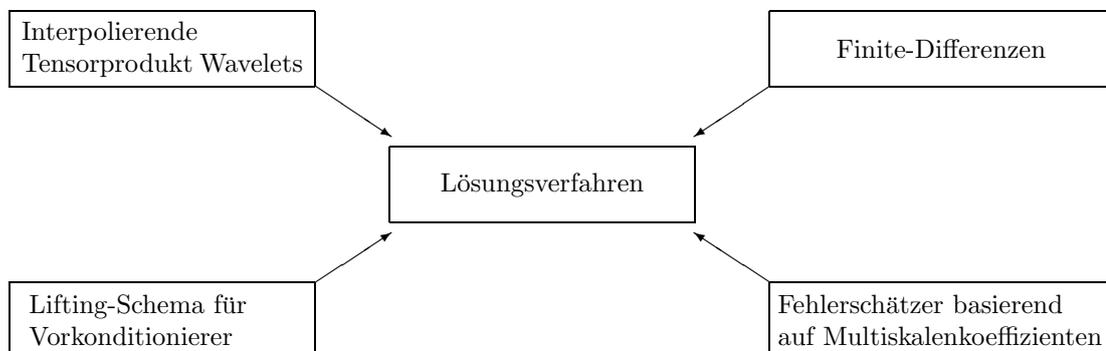
zu effizienten Verfahren für die zu lösenden Gleichungssysteme.

Neben den Approximationseigenschaften der Ansatzfunktionen ist beim numerischen Lösen von PDGLen die Diskretisierung entscheidend. Die wichtigsten Anforderungen sind hierbei:

- die stabile und genaue Diskretisierung der Differentialoperatoren,
- die schnelle Auswertung der diskreten Operatoren,
- das schnelle Lösen der entstehenden Gleichungssysteme,
- die schnelle und genaue Auswertung von nichtlinearen Funktionalen.

Ein früher bei Wavelets recht oft verfolgter Weg war, möglichst viele dieser Forderungen mittels speziell konstruierter Basisfunktionen zu erfüllen. Ein Beispiel dafür sind die sogenannten Wavelet-Vaguelette-Paare [Tch87, LPR94, Sch97], die aus zwei unterschiedlichen Multiskalenbasen (Wavelets und Vaguelettes) bestehen, die z.B. bezüglich des \mathcal{H}^1 -Sklarproduktes orthogonal zueinander sind. Bei einer Petrov-Galerkin-Diskretisierung der Poisson-Gleichung mit diesen beiden Basen als Ansatz- bzw. Testfunktionen entspricht der diskrete Operator einer Diagonalmatrix, die natürlich einfach zu invertieren ist. Jedoch wird die Auswertung aller anderen Operatoren und nichtlinearen Funktionale sehr teuer, da die Vaguelettes sehr breite Träger haben [Sch97]. Ein anderes Beispiel sind die in [Urb95] angegebenen divergenzfreien Wavelets. Hier resultieren aus der Konstruktion viele unterschiedliche Typen von Vektorwertigen Wavelets, die bei jeder Operation separat behandelt werden müssen. Generell scheint die Konstruktion angepaßter Wavelets nur für Operatoren mit konstanten Koeffizienten auf dem Einheitskubus möglich zu sein, so daß dieser Weg nur für eine sehr kleine Problemklasse möglich ist. Deshalb verfolgen wir in dieser Arbeit eine andere Richtung.

Das Ziel dieser Arbeit ist die Entwicklung und Analyse eines effizienten Lösungsverfahrens, das flexibel auf eine große Klasse von Differentialgleichungen angewendet werden kann. Das Verfahren basiert auf anisotropen Tensorprodukten von Verallgemeinerungen der hierarchischen Basis mit einem höheren Grad an polynomialer Exaktheit [DD89]. Die Diskretisierung der Operatoren beruht auf finiten Differenzen und für die Verfeinerungsstrategie werden einfache Fehlerschätzer benutzt, die auf den Multiskalenkoeffizienten basieren. Speziell für das Laplace-Problem werden zwei Vorkonditionierer vorgestellt, wobei einer auf dem sogenannten Lifting-Schema [CDP96, Swe97] beruht.



Mit der Wahl der Ansatzfunktionen vereinen wir die guten Approximationseigenschaften von anisotropen Tensorprodukt-Basen mit der algorithmischen Einfachheit und Effizienz

von hierarchischen Basen. Diese kommt insbesondere bei der Multiskalentransformation zum Tragen, die bei der Auswertung von nichtlinearen Funktionalen oder Differentialoperatoren benötigt wird. Durch die Finite-Differenzen-Schemen steht unmittelbar eine große Zahl von effizienten und hochoptimierten Diskretisierungstechniken für diverse Differentialgleichungen zur Verfügung. Beispiele sind die oszillationsdämpfenden WENO-Verfahren für Konvektionsprobleme [Shu99] oder spezielle Diskretisierungen für div-grad-Operatoren, die bei Sattelpunktproblemen auftreten können.

Bei unserem Verfahren können die Anforderungen an die Diskretisierung der Operatoren weitgehend von den Ansatzfunktionen entkoppelt werden. Dadurch kann bei jedem Teilproblem die jeweils günstigste Lösung gewählt werden. Ein weiterer Vorteil der Multiskalen-basierten Finite-Differenzen-Technik ist, daß die Auswertung der diskreten Operatoren einen deutlich geringeren Rechenaufwand benötigt, als z.B. bei einer Galerkin-Diskretisierung. Auch die Auswertung von rechten Seiten ist einfacher, da keine adaptiven Quadraturen notwendig sind.

Finite-Differenzen-Diskretisierungen sind schon früher in Kombination mit anisotropen Tensorprodukt-Basen benutzt worden [Gri98, Sch98a]. Der Beitrag der vorliegenden Arbeit besteht in den folgenden Punkten:

- Das Konzept aus [Gri98, Sch98a] wird auf hierarchische Basen höherer Ordnung und beliebige explizite Finite-Differenzen-Schemen verallgemeinert. Dadurch gewinnt man wegen der besseren Approximationseigenschaften deutlich an Effizienz. Außerdem zeigt sich, daß für Multiskalenfunktionen höherer Ordnung die elliptischen Gleichungssysteme auch deutlich besser vorkonditioniert werden können.
- Es werden die Grundzüge einer einheitlichen Konvergenztheorie für das Finite-Differenzen- sowie Petrov-Galerkin-Verfahren entwickelt, die als Ansatz- und Testfunktionen primale bzw. duale Multiskalenfunktionen benutzen. Die Theorie beruht auf der sogenannten $(s, s - r)$ -Stabilität der diskreten Operatoren [BA72, Hac96, Sch98b]. Mittels dieses Stabilitätsbegriffes kann der numerische Fehler durch den (relativ einfach zu kontrollierenden) Interpolationsfehler und einen Konsistenzfehler abgeschätzt werden:

$$\|u - u_{\mathcal{T}}\|_{\mathcal{H}^s} \lesssim \|u - I_{\mathcal{T}}u\|_{\mathcal{H}^s} + \|A_{\mathcal{T}}I_{\mathcal{T}}u - I_{\mathcal{T}}Au\|_{\mathcal{H}^{s-r}} \quad (\text{siehe Abschnitt 6.4}).$$

Erstmals werden für den allgemeinen adaptiven Fall Abschätzungen für den Konsistenzfehler gezeigt.

- Für Laplace-Gleichungssysteme werden ein einfacher diagonaler Vorkonditionierer und ein auf dem Lifting-Schema basierender Vorkonditionierer vorgestellt. Für einen Spezialfall gelingt es, Abschätzungen für die l_2 -Konditionszahl κ herzuleiten. In Abhängigkeit von der Ordnung N der (Lifting-)Interpolets und dem maximalen Level L erhalten wir die folgenden asymptotischen Abschätzungen für κ :

	Diagonal-Vorkonditionierer	Lifting-Vorkonditionierer
$N = 2$	$\sim L2^L$	$\sim \sqrt{2^L}$
$N \geq 4$	$\sim \sqrt{L2^L}$	~ 1

Mit dem Lifting-Vorkonditionierer gelingt es damit erstmalig, nahezu optimal vorkonditionierte lineare Gleichungssysteme zu erhalten, die aus der Finite-Differenzen-Diskretisierung in Kombination mit anisotropen Tensorprodukt-Multiskalenfunktionen

resultieren. Interessanterweise sind die Konditionszahlen weitgehend unabhängig von der Raumdimension des Gebietes. Das wäre bei einer Galerkin-Diskretisierung für den Diagonal-Vorkonditionierer nicht der Fall [GO95b].

Diese Arbeit ist folgendermaßen gegliedert:

In *Kapitel 2* werden grundlegende Begriffe zu univariaten biorthogonalen Multiskalenfunktionen definiert und analytische Eigenschaften wie der Grad der polynomialen Exaktheit oder Normäquivalenzen erläutert. Als spezielle Basen betrachten wir verallgemeinerte hierarchische Basisfunktionen, sogenannte Interpolets, und daraus abgeleitete Lifting-Wavelets.

In *Kapitel 3* werden für den univariaten Fall adaptive Algorithmen für die Elementaroperationen (Multiskalentransformation, inverse Multiskalentransformation, Finite-Differenzen-Diskretisierung, das biorthogonale Petrov-Galerkin-Schema und den Basiswechsel von/zu den Lifting-Wavelets) vorgestellt. Ein Schwerpunkt sind dabei algorithmisch motivierte Bedingungen an die adaptiven Basen sowie Voraussetzungen für den Basiswechsel bei den Lifting-Wavelets.

In *Kapitel 4* beschreiben wir zwei alternative Konstruktionen von multivariaten Wavelets. Der Schwerpunkt liegt dabei auf der anisotropen Tensorprodukt-Konstruktion.

In *Kapitel 5* untersuchen wir die Approximationseigenschaften der verschiedenen multivariaten Wavelets. Die wichtigsten in der Literatur bekannten Aussagen werden zusammengetragen und an Hand von Beispielen erläutert.

In *Kapitel 6* werden die Finite-Differenzen- und die biorthogonale Petrov-Galerkin-Diskretisierung für den multivariaten Fall vorgestellt. Neben Algorithmen wird eine Analyse des Konsistenzfehlers angegeben.

In *Kapitel 7* beschäftigen wir uns mit dem Vorkonditionieren von Laplace-Gleichungssystemen. Für einen einfachen Fall gelingt es, die Konditionszahlen abzuschätzen. Numerische Experimente zeigen, daß die gefundenen Abschätzungen wohl auch im allgemeinen Fall gültig sind.

In *Kapitel 8* werden Verfeinerungstechniken für das adaptive Lösen von elliptischen PD-GLen behandelt. Eine einfache, Multiskalenkoeffizienten-basierte Verfeinerungsstrategie wird in zwei numerischen Beispielen für das adaptive Lösen einer Poisson-Gleichung getestet.

Gegenstand von *Kapitel 9* ist die Behandlung von Erhaltungsgleichungen mittels Zeitschritt-Verfahren. Diese erfordern ebenfalls speziell angepaßte Verfeinerungsstrategien. Numerische Experimente zeigen, daß unser Verfahren deutlich effizienter als vergleichbare nichtadaptive Verfahren ist.

In *Kapitel 10* betrachten wir die Navier-Stokes-Gleichungen. Dabei können wir analog wie in Kapitel 9 vorgehen. Jedoch erfordert die Druck-Poisson-Gleichung stabilisierte Finite-Differenzen-Diskretisierungen, die hier konstruiert werden. Anhand zweier Beispiele für zwei- bzw. dreidimensionale turbulente Scherströmungen wird gezeigt, daß unser Verfahren auch auf komplizierte Probleme angewendet werden kann.

Den Abschluß dieser Arbeit bildet *Kapitel 11* in dem die wesentlichen Resultate und Konzepte für zukünftige Weiterentwicklungen diskutiert werden. Die in dieser Arbeit verwendeten Bezeichnungen und Symbole sind im Anhang B aufgelistet.

An dieser Stelle möchte ich allen danken, die mich bei dieser Arbeit unterstützt haben. Meinem Betreuer Prof. Dr. Michael Griebel möchte ich für die anregende Arbeitsatmosphäre und für die wertvollen Diskussionen und Ideen danken. Zudem danke ich Frau Prof. Dr. Angela Kunoth für die Übernahme des Zweitgutachtens. Weiterhin möchte ich mich bei allen meinen Kollegen bedanken, die mir zu zahllosen Gelegenheiten geholfen haben. Besonders hervorheben möchte ich Dr. Gerhard Zumbusch und Alex Schweitzer, die die erstklassige Rechnerinfrastruktur aufgebaut haben. Finanziert wurde meine Tätigkeit von der Deutschen Forschungsgemeinschaft unter der Projektnummer GR-1144/7-2.

Bonn, im November 2001

Kapitel 2

Univariate Multiskalenfunktionen

In diesem Kapitel werden wir zunächst univariate biorthogonale Multiskalenfunktionen beschreiben und die wesentlichen Begriffe einführen. In einem ergänzenden Abschnitt werden dann Hilfsmittel vorgestellt, mit deren Hilfe sich Aussagen über so wichtige Eigenschaften wie Glattheit oder verschwindende Momente der Basisfunktionen treffen lassen. Anschließend gehen wir auf sogenannte verallgemeinerte hierarchische Basen und insbesondere auf Interpolets ein. Dabei werden wir jene algorithmischen bzw. analytischen Eigenschaften hervorheben, welche für unsere adaptiven Multiskalenverfahren von Bedeutung sind.

Aus Gründen der Einfachheit werden wir uns in diesen ersten drei Abschnitten auf Multiskalenfunktionen beschränken, die auf ganz \mathbb{R} definiert sind. Eine Konstruktion von Interpolets auf dem Intervall $[0, 1]$ ist letzten Abschnitt dieses Kapitels angegeben.

2.1 Biorthogonale Multiskalenfunktionen

Ausgangspunkt für die Definition biorthogonaler Multiskalenfunktionen ist eine aufsteigende Folge von Unterräumen $V^l \subset \mathcal{L}^2(\mathbb{R})$, deren Vereinigung dicht in $\mathcal{L}^2(\mathbb{R})$ liegt

$$V^{l_0} \subset V^{l_0+1} \subset \dots \subset \mathcal{L}^2(\mathbb{R}), \quad \overline{\bigcup_{l=l_0}^{\infty} V^l} = \mathcal{L}^2(\mathbb{R}).$$

Hierbei stellt l_0 ein zunächst noch willkürlich gewähltes größtes Level dar, dessen Bedeutung in Abschnitt 2.4 im Zusammenhang mit Multiskalenfunktionen auf dem Intervall geklärt werden wird. In den von uns betrachteten Fällen werden die Räume V^l von Dilaten/Translaten

$$\phi^{(l,s)}(x) := \phi(2^l x - s), \quad s \in \mathbb{Z}$$

einer Mutterfunktion ϕ aufgespannt. Da die Räume V^l ineinander geschachtelt sind, muß ϕ für charakteristische Koeffizienten h_s , $s \in \mathbb{Z}$, der Skalierungsgleichung

$$\phi(x) = \sum_{s \in \mathbb{Z}} h_s \phi(2x - s) \tag{2.1}$$

genügen. Deshalb nennen wir ϕ sowie $\phi^{(l,s)}$ *Skalierungsfunktionen*. Die Funktion ϕ sei normiert:

$$\int_{\mathbb{R}} \phi(x) dx = 1.$$

Die eigentlichen Multiskalenfunktionen $\psi_{(l,t)}$ sind nun die Basisfunktionen von Komplementräumen W_l von V^{l-1} in V^l für $l > l_0$, d.h.

$$V^l = V^{l-1} \oplus W_l \quad \text{und} \quad W_l = \text{span}\{\psi_{(l,t)}\}_{t \in \mathbb{Z}} .$$

Sowohl die Komplementräume als auch deren Basisfunktionen sind *nicht* eindeutig bestimmt. Die Wahl der Komplementräume hat enormen Einfluß auf algorithmische und analytische Eigenschaften der Multiskalenzerlegung. Auf diese Problematik wird u.a. im Abschnitt 2.2 detailliert eingegangen.

Hier nehmen wir an, daß auch die Multiskalenfunktionen Dilate/Translate einer Mutterfunktion ψ sind. Wegen $W_l \subset V^l$ folgt daraus die Existenz von Koeffizienten g_t , $t \in \mathbb{Z}$, so daß

$$\psi_{(l,t)} = \sum_{s \in \mathbb{Z}} g_{s-2t} \phi^{(l,s)} . \quad (2.2)$$

Die Zerlegungen $V^l = V^{l_0} \oplus \bigoplus_{k=l_0+1}^l W_k$ bzw. $\mathcal{L}^2 = \overline{V^{l_0} \oplus \bigoplus_{k=l_0+1}^{\infty} W_k}$ werden *Multiskalenanalysen* von V^l bzw. \mathcal{L}^2 genannt [Mal89a].

Zu der obigen Multiskalenanalyse gibt es eine weitere, die sogenannte duale Multiskalenanalyse mit Räumen \tilde{V}^l und \tilde{W}_l sowie den Basisfunktionen $\tilde{\phi}^{(l,s)}$ und $\tilde{\psi}_{(l,t)}$. Die dualen Basisfunktionen erfüllen für $l > l_0$ die folgenden Biorthogonalitätsrelationen [CDF92]

$$\begin{aligned} \langle \tilde{\phi}^{(l-1,s)}, \phi^{(l-1,t)} \rangle &= \delta_{s,t} \quad , \quad \langle \tilde{\phi}^{(l-1,s)}, \psi_{(l,t)} \rangle = 0 \quad , \\ \langle \tilde{\psi}_{(l,s)}, \psi_{(l,t)} \rangle &= \delta_{s,t} \quad , \quad \langle \tilde{\psi}_{(l,s)}, \phi^{(l-1,t)} \rangle = 0 \quad . \end{aligned} \quad (2.3)$$

Falls die primalen und dualen Funktionen mindestens in \mathcal{L}^2 sind, ist $\langle \cdot, \cdot \rangle$ das \mathcal{L}^2 -Skalarprodukt. Wir werden jedoch in Abschnitt 2.3 auch den Fall kennenlernen, daß die dualen Funktionen Linearkombinationen von Dirac-Funktionalen sind. Dann ist $\langle \cdot, \cdot \rangle$ als duale Paarung zu verstehen. Die dualen Skalierungs- bzw. Multiskalenfunktionen erfüllen ebenfalls Skalierungsgleichungen analog zu (2.1) und (2.2). Es gilt [CDF92]

$$\begin{aligned} \tilde{\phi}(x) &= \sum_{s \in \mathbb{Z}} \tilde{h}_s \tilde{\phi}(2x - s) \quad \text{wobei} \quad \tilde{h}_s := (-1)^s g_{1-s} \\ \tilde{\psi}(x) &= \sum_{s \in \mathbb{Z}} \tilde{g}_s \tilde{\psi}(2x - s) \quad \text{wobei} \quad \tilde{g}_s := (-1)^{1-s} h_{1-s} . \end{aligned} \quad (2.4)$$

Die duale Multiskalenanalyse tritt bereits bei dem einfachsten Algorithmus, der Multiskalentransformation, in Erscheinung. Das Ziel dieses Algorithmus ist, für eine Funktion $u \in V^L$, die als Linearkombination der Skalierungsfunktionen

$$u = \sum_s u^{(L,s)} \phi^{(L,s)} \quad (2.5)$$

gegeben ist, die alternative Multiskalendarstellung

$$u = \sum_{l=l_0}^L \sum_t u_{(l,t)} \psi_{(l,t)} \quad (2.6)$$

zu finden. Dabei treten die dualen Funktionen nicht direkt auf, sondern nur indirekt über die Skalierungskoeffizienten \tilde{h}_s und \tilde{g}_s .

Für die Darstellung des Algorithmus benutzen wir die folgende Notation, von der wir in dieser Arbeit noch oft Gebrauch machen werden. Mit

$$\Phi^l := \{\phi^{(l,s)}\}_{s \in \mathbb{Z}} \quad \text{und} \quad \Psi_l := \{\psi_{(l,t)}\}_{t \in \mathbb{Z}}$$

werden *Zeilenvektoren* der Basisfunktionen bezeichnet. Weiterhin sind

$$\mathbf{u}^l := \{u^{(l,s)}\}_{s \in \mathbb{Z}} \quad \text{und} \quad \mathbf{u}_l := \{u_{(l,t)}\}_{t \in \mathbb{Z}}$$

Spaltenvektoren von Koeffizientenfolgen bzgl. der Skalierungs- bzw. Multiskalenfunktionen. Dementsprechend ist beispielsweise

$$\Phi^l \cdot \mathbf{u}^l = \sum_{s \in \mathbb{Z}} u^{(l,s)} \phi^{(l,s)} \quad \text{und} \quad \Psi_l \cdot \mathbf{u}_l = \sum_{t \in \mathbb{Z}} u_{(l,t)} \psi_{(l,t)} .$$

Hochgestellte Indizes, wie in $u^{(l,s)}$, werden für Darstellungen mittels Skalierungsfunktionen benutzt und tiefgestellte Indizes, wie in $u_{(l,s)}$, für Darstellungen in der Multiskalenbasis. Fett gedruckte Symbole wie \mathbf{u}^l oder Ψ_l bezeichnen Folgen oder Vektoren. Um die lästige Unterscheidung zwischen $l = l_0$ und $l > l_0$ zu vermeiden, setzen wir

$$W_{l_0} := V^{l_0} \quad \text{und} \quad \Psi_{l_0} := \Phi^{l_0} .$$

Mittels Φ^l und Ψ_l kann auch eine kompakte, algebraische Form der Skalierungsbeziehungen (2.1) bzw. (2.2) angegeben werden:

$$\Phi^{l-1} = \Phi^l \cdot H^l, \quad \Psi_l = \Phi^l \cdot G^l . \quad (2.7)$$

Dabei sind H^l und G^l $\mathbb{Z} \times \mathbb{Z}$ -Matrizen mit $H_{r,s}^l := h_{r-2s}$ bzw. $G_{r,s}^l := g_{r-2s}$ für $r, s \in \mathbb{Z}$. In analoger Weise sind $\tilde{\Phi}^l$, $\tilde{\Psi}_l$, \tilde{H}^l , \tilde{G}^l für das duale System definiert. In dieser Notation lauten die Biorthogonalitätsbeziehungen (2.3) demnach

$$\langle \tilde{\Phi}^l, \Phi^l \rangle = I_{\mathbb{Z}}, \quad \langle \tilde{\Psi}_l, \Psi_l \rangle = I_{\mathbb{Z}}, \quad \langle \tilde{\Phi}^{l-1}, \Psi_l \rangle = 0, \quad \langle \tilde{\Psi}_l, \Phi^{l-1} \rangle = 0 . \quad (2.8)$$

Eine unmittelbare Folgerung aus (2.7) und (2.8) ist

$$\begin{pmatrix} \tilde{H}^{lT} \\ \tilde{G}^{lT} \end{pmatrix} \cdot (H^l, G^l) = \begin{pmatrix} I_{\mathbb{Z}} & 0 \\ 0 & I_{\mathbb{Z}} \end{pmatrix} . \quad (2.9)$$

Beispielsweise ist

$$I_{\mathbb{Z}} = \langle \tilde{\Phi}^{l-1}, \Phi^{l-1} \rangle = \langle \tilde{\Phi}^l \cdot \tilde{H}^l, \Phi^l \cdot H^l \rangle = \tilde{H}^{lT} \cdot \langle \tilde{\Phi}^l, \Phi^l \rangle \cdot H^l = \tilde{H}^{lT} \cdot H^l . \quad (2.10)$$

Das Ziel der Multiskalentransformation ist es, ausgehend von der Darstellung $u = \Phi^L \cdot \mathbf{u}^L$ die Koeffizienten \mathbf{u}_l der Multiskalendarstellung $u = \sum_{l=l_0}^L \Psi_l \cdot \mathbf{u}_l$ zu finden. Für z.B. den ersten Schritt der Transformation nutzen wir, daß es wegen $V^L = V^{L-1} \oplus W_L$ Koeffizienten \mathbf{u}^{L-1} und \mathbf{u}_L geben muß, mit $u = \Phi^{L-1} \cdot \mathbf{u}^{L-1} + \Psi_L \cdot \mathbf{u}_L$. Wegen (2.8) erhält man diese Koeffizienten durch Testen von u mit $\tilde{\Phi}^{L-1}$ bzw. $\tilde{\Psi}_L$. Also

$$\mathbf{u}^{L-1} = \langle \tilde{\Phi}^{L-1}, \Phi^L \cdot \mathbf{u}^L \rangle = \tilde{H}^{LT} \cdot \mathbf{u}^L \quad \text{bzw.} \quad \mathbf{u}_L = \langle \tilde{\Psi}_L, \Phi^L \cdot \mathbf{u}^L \rangle = \tilde{G}^{LT} \cdot \mathbf{u}^L . \quad (2.11)$$

Dieses Schema kann nun rekursiv auf \mathbf{u}^{L-1} angewandt werden. Nachfolgend ist der komplette Algorithmus dargestellt. Dabei verwenden wir eine funktionale Notation, wie man sie für ein Computerprogramm verwenden würde. Dies erlaubt eine einfache Beschreibung zum Beispiel für das Ergebnis der Hintereinanderausführung mehrerer Algorithmen. Außerdem sieht man mit der funktionalen Beschreibung auch sofort, welche Werte Input und welche Output sind.

Funktion $\{\mathbf{u}_l\} = \mathbf{WT}(\mathbf{u}^L)$ for $l = L$ downto $l_0 + 1$ $\mathbf{u}_l = \tilde{G}^{lT} \cdot \mathbf{u}^l$ $\mathbf{u}^{l-1} = \tilde{H}^{lT} \cdot \mathbf{u}^l$ end $\mathbf{u}_{l_0} = \mathbf{u}^{l_0}$

Für die inverse Multiskalenttransformation haben wir in jedem Schritt eine Darstellung der Form $\Phi^l \cdot \mathbf{u}^l$ des Terms

$$\Phi^{l-1} \cdot \mathbf{u}^{l-1} + \Psi_l \cdot \mathbf{u}_l$$

zu finden. Wegen (2.7) und (2.8) ist

$$\mathbf{u}^l = \langle \tilde{\Phi}^l, \Phi^{l-1} \cdot \mathbf{u}^{l-1} + \Psi_l \cdot \mathbf{u}_l \rangle = H^l \cdot \mathbf{u}^{l-1} + G^l \cdot \mathbf{u}_l.$$

Der komplette Algorithmus für die inverse Transformation ist somit:

Funktion $\{\mathbf{u}^l\} = \mathbf{IWT}(\{\mathbf{u}_l\})$ $\mathbf{u}^{l_0} = \mathbf{u}_{l_0}$ for $l = l_0 + 1$ to L $\mathbf{u}^l = G^l \cdot \mathbf{u}_l + H^l \cdot \mathbf{u}^{l-1}$ end
--

Falls die Skalierungsmatrizen H^l, G^l, \tilde{H}^l und \tilde{G}^l gleichmäßig dünnbesetzt sind, d.h. die Anzahl von nichtverschwindenden Einträgen pro Matrixzeile ist gleichmäßig in l beschränkt, dann ist der Rechenaufwand für WT bzw. IWT proportional zur Anzahl von nichtverschwindenden Einträgen in \mathbf{u}^L . Diese Forderung an die Skalierungsmatrizen ist keineswegs trivial. Selbst wenn man die Skalierungs- bzw. Multiskalenfunktionen so wählt, daß H^l und G^l gleichmäßig dünnbesetzt sind, dann folgt daraus noch lange nicht, daß dies auch für \tilde{H}^l und \tilde{G}^l gilt. Formal kann man dann zwar die in WT notwendigen Multiplikationen mit \tilde{H}^{lT} bzw. \tilde{G}^{lT} auf das Lösen von Gleichungssystemen mit (H^l, G^l) als Koeffizientenmatrix zurückführen, vergleiche (2.9), jedoch ist dieser Ansatz im adaptiven Fall viel zu kompliziert.

2.2 Glattheit, polynomiale Exaktheit und Normäquivalenzen

Nachdem wir einen ersten Eindruck von den algorithmischen Aspekten von Multiskalenfunktionen bekommen haben, wenden wir uns jetzt den analytischen Eigenschaften zu. Die wichtigsten dabei sind die Glattheit der Skalierungsfunktionen, der Grad der polynomialen Exaktheit der Räume V^l sowie die Zahl der verschwindenden Momente der Multiskalenfunktionen. Diese Eigenschaften sind essentiell für die Approximationseigenschaften der Basen sowie für die Äquivalenz von bestimmten Sobolev-Normen und Normen der Multiskalenkoeffizienten. Zunächst sollen aber einige der Begriffe genauer definiert werden.

Definition 2.1

- (i) Die Skalierungsfunktion ϕ hat den Grad N der polynomialen Exaktheit, falls sich jedes Polynom p vom Grade kleiner als N als Linearkombination von Translaten von ϕ darstellen läßt. D.h. es gibt Koeffizienten $\{p_s\}_{s \in \mathbb{Z}}$ mit

$$p(x) = \sum_{s \in \mathbb{Z}} p_s \phi(x - s) \quad \forall x$$

- (ii) Die Multiskalenfunktionen haben M verschwindende Momente, falls für $t \in \mathbb{Z}$ und $l > l_0$ gilt

$$\int_{\mathbb{R}} \psi_{(l,t)}(x) x^i dx = 0 \quad \forall 0 \leq i < M .$$

- (iii) Die Parameter γ^* bzw. θ^* bezeichnen obere Schranken für die Sobolev- bzw. Lipschitz-Regularität von ϕ .

$$\gamma^* = \sup\{s \mid \phi \in \mathcal{H}^s\}, \quad \theta^* = \sup\{s \mid \phi \in \mathcal{C}^s\}$$

Analog sind \tilde{N} , \tilde{M} , $\tilde{\gamma}^*$ und $\tilde{\theta}^*$ für die dualen Skalierungs- bzw. Multiskalenfunktionen definiert.

Einer der Gründe, warum wir diese Eigenschaften bereits an dieser Stelle erwähnen, ist ihre Anschaulichkeit. Durch sie bekommt man eine konkretere Vorstellung von den einzelnen Funktionen, die im ersten Abschnitt doch recht abstrakt eingeführt wurden. Weiterhin brauchen wir z.B. die Glattheit um festzustellen, ob bestimmte Ansatzfunktionen für unsere Zwecke geeignet sind oder nicht: Eine der Diskretisierungstechniken, die wir vorstellen werden, ist eine Kollokationsmethode, so daß die Multiskalenfunktionen mindestens in C^2 sein müssen, um für die Lösung der Poisson-Gleichung in Frage zu kommen.

Die obigen Eigenschaften haben eine Gemeinsamkeit: sie lassen sich analytisch bzw. numerisch über das Symbol

$$h(z) = \frac{1}{2} \sum_{s \in \mathbb{Z}} h_s z^s$$

bzw. die analog definierten Symbole $g(z)$, $\tilde{h}(z)$, $\tilde{g}(z)$ erschließen. Dies ist Gegenstand des nachfolgenden Theorems, das die wichtigsten Aussagen zusammenfaßt, vgl. [Dau92].

Theorem 2.1

- (i) Der Grad der polynomialen Exaktheit der Skalierungsfunktion ϕ entspricht dem Grad der Nullstelle des Symbols $h(z)$ in $z = -1$. D.h. es gibt eine rationale Funktion $L(z)$, so daß

$$h(z) = \left(\frac{1+z}{2} \right)^N L(z) \quad \text{wobei} \quad 0 < |L(-1)| < \infty . \quad (2.12)$$

- (ii) Die Zahl der verschwindenden Momente von ψ entspricht dem Grad der Nullstelle von $g(z)$ in $z = 1$.
- (iii) Der Grad der polynomialen Exaktheit der dualen Skalierungsfunktionen ist gleich der Zahl der verschwindenden Momente der primalen Multiskalenfunktionen, d.h. $M = \tilde{N}$ und $N = \tilde{M}$.
- (iv) Sei $L(z)$ wie in (2.12) definiert. Weiterhin sei

$$q_j := \sup_{\zeta \in \mathbb{R}} \left| \prod_{k=0}^{j-1} L(e^{-i2^{-k}\zeta}) \right|$$

$$K_j := \frac{1}{j} \log_2 q_j, \quad K := \inf_{j \in \mathbb{N}} K_j .$$

Dann gilt

$$\theta^* \geq N - K - 1 \quad \text{und} \quad \gamma^* \geq N - K - \frac{1}{2} .$$

Die Beweise sind im Anhang A.1 aufgeführt. Teil (iv) des obigen Theorems ist Grundlage der numerischen Schätzung von θ^* bzw. γ^* . Oft sind die Koeffizienten h_s von $h(z)$ explizit als rationale Zahlen bekannt. Damit kann durch Polynomdivision eine geschlossene Darstellung von $L(z)$ gefunden werden, so daß die Koeffizienten q_j bzw. K_j exakt berechnet werden können. In der Regel erhält man bereits für $j < 10$ sehr gute Abschätzungen der gesuchten Werte.

Die polynomiale Exaktheit, sowie das Prinzip, die Skalierungs- bzw. Multiskalenfunktionen als Dilate/Translate zu definieren, bilden die Grundlage für die Charakterisierung der Sobolev-Regulartät von Funktionen mittels des Abfallverhaltens ihrer Multiskalenkoeffizienten. Da die einzelnen Zwischenschritte zum Beweis solcher Beziehungen auch für sich interessant sind, sollen diese nachfolgend skizziert werden. Die Beweise findet man in Kapitel 6 von [Sch98b] bzw. der darin zitierten Literatur.

Lemma 2.1 (Jackson-Ungleichung) Für $u \in \mathcal{H}^t$ und $s < \gamma^*$, $s \leq t \leq N$ gilt

$$\inf_{v \in V^l} \|u - v\|_{\mathcal{H}^s} \lesssim 2^{-l(t-s)} \|u\|_{\mathcal{H}^t} .$$

Diese Aussage kann allein mit der polynomialen Exaktheit der Skalierungsfunktionen $\phi^{(l,s)}$ und dem Bramble-Hilbert-Lemma sowie Interpolationstechniken bewiesen werden.

Lemma 2.2 (Bernstein-Ungleichung) Für $v \in V^l$ und $s \leq t < \gamma^*$ gilt

$$\|v\|_{\mathcal{H}^t} \lesssim 2^{l(t-s)} \|v\|_{\mathcal{H}^s} .$$

Diese Aussage kann man in unserem Fall aus der Konstruktion $\phi^{(l,s)} = \phi(2^l x - s)$ folgern. Eine weitere Folgerung daraus ist

Lemma 2.3 (\mathcal{L}^2 -Stabilität auf einem Level) Es gilt für $l \geq l_0$ mit von l unabhängigen Äquivalenzkonstanten

$$\|\Psi_l \cdot \mathbf{u}_l\|_{\mathcal{L}^2}^2 \sim 2^{-l} \sum_t |u_{(l,t)}|^2 . \quad (2.13)$$

Der Vorfaktor 2^{-l} resultiert aus der Normierung $\|\psi_{(l,t)}\|_{\mathcal{L}^2}^2 \sim 2^{-l}$. Die obige Relation ist bereits ein Prototyp für die eigentlichen Normäquivalenzen. Gegenüber dem allgemeinen Fall hat man hier die Einschränkung auf \mathcal{L}^2 und auf ein Level. Aus den obigen drei Aussagen folgt

Theorem 2.2 (Einseitige Norm-Abschätzungen) Für $-M < s < \gamma^*$ gibt es eine Konstante c_s , so daß für alle $L \geq l_0$ und alle $u = \sum_{l=l_0}^L \Psi_l \cdot \mathbf{u}_l$ gilt:

$$\|u\|_{\mathcal{H}^s}^2 \leq c_s \sum_{l=l_0}^L 4^{ls} \|\Psi_l \cdot \mathbf{u}_l\|_{\mathcal{L}^2}^2 \lesssim \sum_{l=l_0}^L 2^{-l} \sum_t 4^{ls} |u_{(l,t)}|^2 . \quad (2.14)$$

Für $s = -M$ gilt

$$\|u\|_{\mathcal{H}^{-M}}^2 \leq c_{-M} L \sum_{l=l_0}^L 2^{-l} \sum_t 4^{-lM} |u_{(l,t)}|^2 . \quad (2.15)$$

Bemerkung 2.1 Die linke Abschätzung in (2.14) wird ausschließlich mit den Jackson- und Bernstein-Ungleichungen bewiesen. Für rechte Abschätzung in (2.14) bzw. (2.15) benötigt man darüber hinaus nur den \lesssim Teil von (2.13).

Analog kann man auch für die dualen Skalierungs- bzw. Multiskalenfunktionen vorgehen. Dabei ist für $\tilde{\gamma}^* < 0$ die \mathcal{L}^2 -Stabilität aus Lemma 2.3 durch eine \mathcal{H}^r -Stabilität für ein $r < \tilde{\gamma}^*$ in der Form $\|\tilde{\Psi}_l \cdot \mathbf{u}_l\|_{\mathcal{H}^r} \sim \sum_t 4^{rl} |u_{(l,t)}|^2 2^{+l}$ zu ersetzen. Der Faktor 2^{+l} resultiert hier wieder aus der Normierung der primalen Basisfunktionen zu $\|\psi_{(l,t)}\|_{\mathcal{L}^2}^2 \sim 2^{-l}$. In der linken Ungleichung in (2.14) wird dann zuerst $\|u\|_{\mathcal{H}^s}^2$ durch $\sum_{l=l_0}^L \sum_t 4^{l(s-r)} \|\tilde{\Psi}_l \cdot \mathbf{u}_l\|_{\mathcal{H}^r}^2$ abgeschätzt. Man erhält somit Abschätzungen für \mathcal{H}^s -Normen für $s \in [-\tilde{M}, \tilde{\gamma}^*]$. Zusammen mit einem Dualitätsargument [Sch98b] folgt dann damit die zentrale Aussage dieses Abschnittes.

Theorem 2.3 (Normäquivalenz) Für $-\tilde{\gamma}^* < s < \gamma^*$ und $u = \sum_{l=l_0}^L \Psi_l \cdot \mathbf{u}_l$ gilt

$$\|u\|_{\mathcal{H}^s} \sim \sum_{l=l_0}^L 2^{-l} \sum_t 4^{ls} |u_{(l,t)}|^2 .$$

2.3 Verallgemeinerte Hierarchische Basen und Interpolets

In diesem Abschnitt stellen wir eine besondere Klasse von biorthogonalen Systemen vor, die sehr angenehme algorithmische Eigenschaften hat: verallgemeinerte hierarchische Basen. Interpolets sind spezielle hierarchische Basen. Sie bilden die Grundlage für die in dieser Arbeit betrachteten numerischen Methoden. Wir werden deshalb detailliert auf die Eigenschaften von Interpolets eingehen.

Definition 2.2 (Verallgemeinerte Hierarchische Basen) Seien $\phi^{(l,s)}, l \geq l_0$, vorgegebene Skalierungsfunktionen. Dann sind für verallgemeinerte hierarchische Basen die zugehörigen Multiskalenfunktionen als

$$\psi_{(l,t)} := \begin{cases} \phi^{(l,2t+1)} & l > l_0 \\ \phi^{(l,t)} & l = l_0 \end{cases} .$$

definiert. Somit sind die Multiskalenfunktionen einfach nur hierarchisch angeordnete Skalierungsfunktionen von unterschiedlichen Leveln.

Durch die Wahl der Skalierungsfunktionen können einige Eigenschaften der Multiskalenfunktionen bestimmt werden. Wählt man zum Beispiel die orthogonalen Daubechies-Skalierungsfunktionen [Dau88, Dau93], dann sind die Multiskalenfunktionen eines Levels orthogonal, jedoch nicht zwischen verschiedenen Leveln. Algorithmisch sinnvoller ist es aber, interpolierende Skalierungsfunktionen mit kompaktem Träger zu wählen [Don92]. Eine Möglichkeit sind Splines. In diesem Fall erhält man hierarchische Splines als Multiskalenfunktionen. Eine andere Möglichkeit ist, die Fundamentalfunktionen des Deslauriers-Dubuc-Interpolationsprozesses [DD89] als Skalierungsfunktionen zu wählen. In diesem Fall erhält man Interpolets. Speziell für die beiden letztgenannten Beispiele, bei den die Skalierungsfunktionen interpolierend sind, d.h.

$$\phi(s) = \delta_{0,s} , \quad s \in \mathbb{Z} \tag{2.16}$$

folgt für verallgemeinerte hierarchische Basen, daß die dualen Skalierungsfunktionen Translate des Dirac-Funktional sind. Im allgemeinen ist die duale Skalierungsfunktion $\tilde{\phi}$ durch

die Biorthogonalitätsbeziehung $\langle \tilde{\phi}, \phi(\cdot - s) \rangle = \delta_{0,s}$ nicht eindeutig bestimmt. Wegen (2.16) erfüllt das Dirac-Funktional diese Bedingung. Es gibt aber auch andere Funktionen, die diese Bedingung erfüllen. Ein wichtiges Beispiel werden wir in Abschnitt 2.5 kennenlernen.

Eindeutig bestimmt werden die dualen Skalierungsfunktionen dann aber durch die Skalierungsgleichung, der sie genügen müssen. Für hierarchische Basen ist definitionsgemäß $g_s = \delta_{s,1}$, $s \in \mathbb{Z}$, so daß wegen (2.4) $\tilde{h}_s = (-1)^s g_{1-s} = \delta_{0,s}$ ist. Also muß $\tilde{\phi}$ die Gleichung $\tilde{\phi}(x) = \tilde{\phi}(2x)$ erfüllen. Als Lösung bleibt somit das Dirac-Funktional übrig (ohne Beweis). Für eine genauere Untersuchung von Lösungen von Skalierungsgleichungen mit weiteren Nebenbedingungen sei auf [Dau92] verwiesen. Dirac-Funktionale als duale Skalierungsfunktionen führen zu zwei wichtigen algorithmischen Eigenschaften.

Lemma 2.4 *Für verallgemeinerte hierarchische Basen mit interpolierenden Skalierungsfunktionen sind die dualen Skalierungsfunktionen verschobene Dirac-Funktionale $\tilde{\phi}^{(l,s)} = \delta(2^l \cdot -s)$. Daraus folgt*

- (i) *Die Koeffizienten $u^{(l,s)}$ einer Funktion u sind die Knotenwerte $u(2^{-l}s)$.*
- (ii) *Im Algorithmus WT werden im Teilschritt $\mathbf{u}^{l-1} = \tilde{H}^{lT} \cdot \mathbf{u}^l$ einfach nur Knotenwerte kopiert. Dieser Schritt ist somit eigentlich überflüssig. Die Multiskalenkoeffizienten $u_{(l,t)}$ können direkt, ohne Zwischenschritte aus Knotenwerten von u berechnet werden.*

Der Beweis ist trivial. Auf die Bedeutung dieser beiden Eigenschaften gerade für *adaptive* Verfahren wird in den Kapiteln 3 und 4 noch einmal besonders eingegangen.

Wir betrachten nun speziell die Interpolets, bei denen die Skalierungsfunktionen aus einem einfachen, aber sehr mächtigen Interpolationsprozeß stammen [DD89]. Zu gegebenen Daten $y(s)$, $s \in \mathbb{Z}$, ist eine möglichst glatte interpolierende Funktion $y : \mathbb{R} \rightarrow \mathbb{R}$ gesucht. In einem ersten Schritt werden Interpolationswerte in $\mathbb{Z} + \frac{1}{2}$ bestimmt. Für die Berechnung von z.B. $y(s + \frac{1}{2})$ wird zunächst das Lagrange-Polynom $p(x)$ vom Grad $2n-1$ berechnet, das die Daten in $s-n+1, \dots, s+n$ interpoliert. Dann ist $y(s + \frac{1}{2}) := p(s + \frac{1}{2})$. Der Parameter n bestimmt später die Glattheit des Interpolanten und den Grad der polynomialen Exaktheit. Nachdem die Werte $y(\mathbb{Z})$ und $y(\mathbb{Z} + \frac{1}{2})$ bekannt sind, werden nach dem gleichen Schema $y(\mathbb{Z} + \frac{1}{4})$ und $y(\mathbb{Z} + \frac{3}{4})$ berechnet. Dabei werden z.B. für $y(s + \frac{1}{4})$ die Werte $y(s + \frac{1}{2}(-n+1)), \dots, y(s + \frac{n}{2})$ benutzt. Auf diese Weise lassen sich also Interpolationswerte für y auf einer Menge finden, die dicht in \mathbb{R} liegt. Der Interpolant $y(x)$ ist dann als stetige Fortsetzung dieser Werte definiert. Weiterhin ist der Interpolant linear von den Daten abhängig, so daß es eine Fundamentalfunktion F gibt, mit

$$y(x) = \sum_{s \in \mathbb{Z}} y(s) F(x - s) .$$

F ist der Interpolant für die Daten $y(s) = \delta_{0,s}$, $s \in \mathbb{Z}$.

Definition 2.3 (Interpolets der Ordnung $2n$) *Interpolets der Ordnung $2n$ sind die aus Definition 2.2 resultierenden Multiskalenfunktionen für $\phi = F$.*

Im folgenden fassen wir die wichtigsten Eigenschaften von ϕ zusammen. Die Beweise dafür sind in der Originalarbeit [DD89] zu finden. Lediglich für die Glattheit wurde ein numerisches Verfahren benutzt, das auf dem Kriterium aus Theorem 1 basiert.

Theorem 2.4 (Eigenschaften der Interpolets)

(i) Die Funktion ϕ erfüllt die Skalierungsgleichung (2.1), wobei die Koeffizienten h_s , $s \in \mathbb{Z}$, durch $h_s = h_{-s}$, $h_0 = 1$ und

$$\begin{pmatrix} 1^0 & 3^0 & \dots & (2n-1)^0 \\ 1^2 & 3^2 & \dots & (2n-1)^2 \\ \vdots & \vdots & \vdots & \vdots \\ 1^{2n-2} & 3^{2n-2} & \dots & (2n-1)^{2n-2} \end{pmatrix} \cdot \begin{pmatrix} h_1 \\ h_3 \\ \vdots \\ h_{2n-1} \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

definiert sind. Alle anderen Koeffizienten sind 0.

(ii) Es gilt

$$\text{supp}\phi = [-2n + 1, 2n - 1]. \quad (2.17)$$

(iii) Der Grad der polynomialen Exaktheit ist $N = 2n$.

(iv) Die Funktion ϕ ist interpolierend in \mathbb{Z} , d.h. $\phi(s) = \delta_{0,s}$. Eine einfache Folgerung daraus ist, daß die Polynome x^i für $0 \leq i < N$ (punktweise) folgendermaßen dargestellt werden können

$$x^i = \sum_{s \in \mathbb{Z}} s^i \phi(x - s). \quad (2.18)$$

(v) Aus dem Kriterium in Theorem 2.1 ergeben sich die folgenden Schätzwerte für die Glattheit von ϕ .

N	θ^* (Lipschitz)	γ^* (Sobolev)
2	1.0	1.5
4	1.671	2.171
6	2.255	2.755

Tabelle 2.1: Lipschitz- und Sobolev-Glattheit von ϕ für verschiedene N .

In Abbildung 2.1 sind die Funktionen ϕ für verschiedene Parameter $N = 2n$ dargestellt. Man

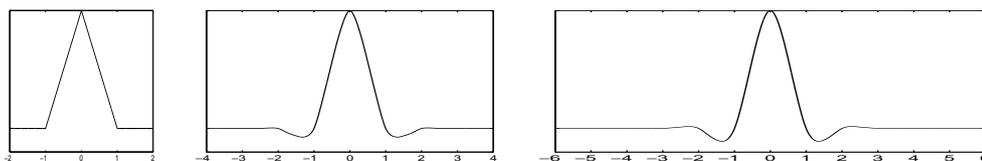


Abbildung 2.1: Die Funktionen ϕ für $N = 2, 4$ und 6 .

beachte, daß sich die Funktionen für $N = 4$ und $N = 6$ kaum unterscheiden, ähnlich wie beim Ansatz in [Bun98].

2.4 Interpolets auf dem Intervall

Die auf ganz \mathbb{R} definierten Interpolets sind nicht für die Approximation von Funktionen geeignet, die auf endlichen Gebieten definiert sind. Deshalb müssen die Definitionen/Konstruktionen der vorangegangenen Abschnitte entsprechend modifiziert werden. Der Einfachheit

halber betrachten wir das Gebiet $[0, 1]$. Das wesentliche Ziel unserer Konstruktion ist, Funktionen zu finden, die Skalierungsgleichungen analog zu (2.1) zu erfüllen. Weiterhin sollen diese Funktionen auf $[0, 1]$ alle Polynome vom Grad $< N$ aufspannen, um die Approximationseigenschaften zu erhalten. Drittens sollen sie interpolierend sein. Als eine erste Idee betrachten wir für $l \geq l_0 := \lceil \log_2 2N \rceil$ die Skalierungsfunktionen

$$\{\phi^{(l,s)}\}_{s=N}^{2^l-N}. \quad (2.19)$$

Wegen Theorem 2.4 (ii) haben diese Funktionen ihren Träger in $[0, 1]$ und sie lassen sich als Linearkombinationen von $\{\phi^{(l+1,s)}\}_{s=N}^{2^{l+1}-N}$ darstellen. Außerdem sind sie interpolierend in $2^{-l}\{N, \dots, 2^l - N\}$. Allerdings werden in der Nähe der Randpunkte 0 und 1 die Polynome x^i , $0 \leq i < N$, nicht durch (2.19) aufgespannt. Um dieses Problem zu lösen, fügen wir weitere Skalierungsfunktionen $\phi^{(l,0)}, \dots, \phi^{(l,N-1)}$ bzw. $\phi^{(l,2^l-N+1)}, \dots, \phi^{(l,2^l)}$ hinzu, die jedoch *keine* einfachen Dilate/Translate von ϕ sind, sondern geeignete Linearkombinationen von Dilaten/Translaten von ϕ . Wir beschreiben die Konstruktion dieser zusätzlichen Funktionen nur für den linken Rand $[0, \dots]$. Die Konstruktion für den rechten Rand $[\dots, 1]$ verläuft vollkommen analog. Zunächst fassen wir alle Translate $\phi(2^l x - s)$, die in 0 nicht verschwinden, zu einem Zeilenvektor zusammen $\Phi_0^l(x) := \{\phi(2^l x - s)\}_{-N < s < N}$. Diese Translate sind zusammen mit den inneren Skalierungsfunktionen aus (2.19) an der Rekonstruktion der Monome x^i für $x \geq 0$ beteiligt. Wegen (2.18) haben wir dann für $0 \leq i < N$ die folgende Darstellung der Monome

$$(2^l x)^i = \Phi_0^l(x) \cdot A_{\cdot,i} + \sum_{s \geq N} s^i \phi^{(l,s)}(x), \quad x \geq 0. \quad (2.20)$$

Dabei ist A die $\{-N+1, \dots, N-1\} \times \{0, \dots, N-1\}$ -Matrix mit Einträgen $A_{s,i} = s^i$. Für die Skalierungsgleichung für Φ_0^l benötigen wir die Skalierungsfunktionen

$$\Phi_+^{l+1}(x) := \{\phi(2^{l+1} x - s)\}_{N \leq s \leq 3N-3}$$

sowie die Teilmatrizen H_0^{l+1} und H_+^{l+1} von H^{l+1} ,

$$H_0^{l+1} := H_{\{-N+1, \dots, N-1\}, \{-N+1, \dots, N-1\}}^{l+1} \quad \text{bzw.} \quad H_+^{l+1} := H_{\{-N+1, \dots, N-1\}, \{N, \dots, 3N-3\}}^{l+1}.$$

Diese Matrizen und Vektoren sind genau so definiert, daß für $x \geq 0$ gilt

$$\Phi_0^l(x) = \Phi_0^{l+1}(x) \cdot H_0^{l+1} + \Phi_+^{l+1}(x) \cdot H_+^{l+1}. \quad (2.21)$$

Aus (2.18) läßt sich nun folgern, daß die Spalten $A_{\cdot,i}$ Eigenvektoren von H_0^{l+1} zum Eigenwert $\frac{1}{2^i}$ sind, [MP98] Lemma 3, so daß

$$H_0^{l+1} A = A D \quad \text{wobei} \quad D := (\delta_{i,j} 2^{-i})_{0 \leq i, j < N}. \quad (2.22)$$

Multiplizieren wir nun (2.21) mit A und nutzen (2.22), so erhält man

$$\Phi_0^l \cdot A = (\Phi_0^{l+1} \cdot A) D + \Phi_+^{l+1} \cdot (H_+^{l+1} A). \quad (2.23)$$

Das zeigt aber, daß $\Phi_0^l \cdot A$ als Linearkombination von $\Phi_0^{l+1} \cdot A$ und weiteren inneren Skalierungsfunktionen von Level $l+1$ geschrieben werden kann. D.h. es gilt eine Skalierungsgleichung. Somit stellen die Funktionen $\Phi_0^l \cdot A$ schon fast die gesuchten Ergänzungen von (2.19) dar. Wir

müssen uns nur noch um Randbedingungen (homogen Dirichlet oder Neumann) und um die Interpolationseigenschaft kümmern.

Wir beginnen mit der Behandlung von homogenen Dirichlet-Randbedingungen. Aus (2.20) folgt für $x = 0$ und $0 \leq i < N$

$$\delta_{0,i} = \Phi_0^l(0) \cdot A_{.,i} + 0.$$

D.h. durch Weglassen der Funktion $\Phi_0^l \cdot A_{.,0}$ erhält man einen Satz von randangepaßten Skalierungsfunktionen mit homogenen Dirichlet-Werten:

$$\Phi_0^l \cdot AU, \quad U := \begin{pmatrix} 0 \\ I_{\{1,\dots,N-1\}} \end{pmatrix}.$$

Das Weglassen einer Skalierungsfunktion entspricht der Elimination eines Freiheitsgrades, welcher durch die Randbedingungen vorgegeben ist. Trotz dieser Modifikation gilt eine Skalierungsgleichung analog zu (2.23). Man kann sich leicht überlegen, daß $DU = UU^T DU$, so daß

$$\Phi_0^l \cdot AU = \left(\Phi_0^{l+1} \cdot AU \right) \left(U^T DU \right) + \Phi_+^{l+1} \cdot \left(H_+^{l+1} \cdot AU \right).$$

In Bezug auf die polynomiale Exaktheit gibt es jedoch Einschränkungen. Wir haben beispielsweise nur noch für $1 \leq i < N$ die Rekonstruktion

$$(2^l x)^i = \Phi_0^l \cdot AU + \sum_{s \geq N} s^i \phi^{(l,s)}.$$

Zumindest in einer Umgebung 0 kann demnach das Polynom x^0 nicht dargestellt werden. Dies ist aber auch nicht nötig, um für Räume von Funktionen mit homogenen Dirichlet-Werten die Approximationseigenschaften (Jackson-Ungleichung) nachzuweisen. Deshalb stört der Verlust der Darstellbarkeit von x^0 nicht. Analoge Betrachtungen gelten für den Fall von homogenen Neumann-Randbedingungen. Hier wird

$$2^l \delta_{1,i} = \frac{\partial}{\partial x} \Phi_0^l(0) \cdot A_{.,i}$$

ausgenutzt, so daß $\Phi_0^l \cdot A_{.,1}$ weggelassen werden muß. Man verliert dabei die Darstellbarkeit von x^1 in Randnähe. In diesem Fall hat man als randangepaßte Skalierungsfunktionen $\Phi_0^l \cdot AU$, wobei

$$U = \left(\begin{array}{c|ccc} 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \hline 0 & I_{\{1,\dots,N-2\}} & & \end{array} \right) \in M(N, N-1).$$

Die mit der obigen Konstruktion erhaltenen N bzw. $N-1$ Skalierungsfunktionen $\Phi_0^l \cdot A$ bzw. $\Phi_0^l \cdot AU$ sind noch nicht interpolierend. Diese Eigenschaft kann dadurch erzielt werden, daß wir z.B. $\Phi_0^l \cdot A$ mit der Inversen der Matrix B der Knotenwerte von $\Phi_0^l \cdot A$ multiplizieren:

$$B = \begin{cases} (s^i)_{0 \leq s, i < N} & \text{(keine Randbed.)} \\ (s^i)_{1 \leq s, i < N} & \text{(homogene Dirichlet-RB.)} \\ (s^i)_{1 \leq s < N, i=0,2,\dots,N} & \text{(homogene Neumann-RB.)} \end{cases}.$$

Die finalen randangepaßten Skalierungsfunktionen sind dann

$$\{\phi^{(l,0)}, \dots, \phi^{(l,N-1)}\} := \Phi_0^l \cdot AB^{-1} \quad \text{bzw.} \quad \{\phi^{(l,1)}, \dots, \phi^{(l,N-1)}\} := \Phi_0^l \cdot AUB^{-1} \quad (2.24)$$

für den Fall homogener Dirichlet- oder Neumann-Randbedingungen. Es gilt

$$\phi^{(l,s)}(2^{-l}t) = \delta_{s,t}, \quad t \in \mathbb{N}, \quad 0 \leq s < N \quad \text{bzw.} \quad 0 < s < N.$$

Da B eine van der Monde Matrix ist – also invertierbar – wird durch die obige Definition weder die polynomiale Exaktheit noch die Gültigkeit einer Skalierungsgleichung beeinträchtigt. Aus (2.23) wird beispielsweise

$$\Phi_0^l \cdot AB^{-1} = (\Phi_0^{l+1} \cdot AB^{-1})(BDB^{-1}) + \Phi_+^{l+1} \cdot (H_+^{l+1}AB^{-1}).$$

Eine analoge Konstruktion kann auch für den rechten Rand des Intervalls $[0, 1]$ durchgeführt werden. Insgesamt erhalten wir somit auf Level $l \geq l_0$ die Skalierungsfunktionen $\phi^{(l,s)}$ für s aus der Indexmenge $S(l)$, welche nachfolgend definiert ist:

Randbed. in 0	Randbed. in 1	$S(l)$
nein	nein	$\{0, \dots, 2^l\}$
ja	nein	$\{1, \dots, 2^l\}$
nein	ja	$\{0, \dots, 2^l - 1\}$
ja	ja	$\{1, \dots, 2^l - 1\}$

Tabelle 2.2: Indexmenge für Skalierungsfunktionen auf Level l in Abhängigkeit von den Randbedingungen.

Wie im shift-invarianten Fall in Abschnitt 2.3 werden die primalen Multiskalenfunktionen nun als

$$\psi_{(l,t)} := \begin{cases} \phi^{(l,2t+1)} & l > l_0 \\ \phi^{(l,t)} & l = l_0 \end{cases} \quad \text{wobei} \quad t \in T(l) := \begin{cases} \{0, \dots, 2^{l-1} - 1\} & l > l_0 \\ S(l_0) & l = l_0 \end{cases} \quad (2.25)$$

definiert. Man kann nun zeigen, daß auch hier wieder die dualen Skalierungsfunktionen Translate des Dirac-Funktional sind. Bevor wir dies aber tun, soll noch die Notation aus Abschnitt 2.1 an die Randmodifikationen angepaßt werden. Diese Notation wird im Rest dieser Arbeit benutzt. Wir bezeichnen mit Φ^l bzw. $\tilde{\Phi}^l$ die Zeilenvektoren

$$\Phi^l := \{\phi^{(l,s)}\}_{s \in S(l)} \quad \text{bzw.} \quad \tilde{\Phi}^l := \{\tilde{\phi}^{(l,s)}\}_{s \in S(l)}.$$

Analog sind die Zeilenvektoren $\Psi_l := \{\psi_{(l,t)}\}_{t \in T(l)}$ bzw. $\tilde{\Psi}_l := \{\tilde{\psi}^{(l,t)}\}_{t \in T(l)}$ definiert.

Die Skalierungsmatrizen werden wiederum mit H^l , \tilde{H}^l , G^l und \tilde{G}^l bezeichnet. Allerdings handelt es sich diesmal um endliche Matrizen: H^l und \tilde{H}^l sind $S(l) \times S(l-1)$ -Matrizen und G^l , \tilde{G}^l sind $S(l) \times T(l)$ -Matrizen. Exemplarisch ist in Abbildung 2.2 das Besetzungsmuster von H^l für Interpolets zweiter Ordnung dargestellt. In der oberen linken und der unteren rechten Ecke stehen die Koeffizienten für die randangepaßten Skalierungsfunktionen und in der Mitte finden sich wieder die Koeffizienten h_s vom shift-invarianten Fall.

Theorem 2.5 (Skalierungsmatrizen) Sei $H^l \in M(S(l), S(l-1))$ die Skalierungsmatrix von Φ^l und $G^l \in M(S(l), T(l))$, $G_{s,t}^l = \delta_{2t+1,s}$, die Skalierungsmatrix für die hierarchischen Basisfunktionen. Dann ist \tilde{H}^l bzw. \tilde{G}^l durch

$$\begin{aligned} \tilde{H}_{s,t}^l &:= \delta_{2t,s}, \quad s \in S(l), \quad t \in S(l-1) \\ \tilde{G}^l &:= (I_{S(l)} - \tilde{H}^l H^{lT}) G^l \end{aligned}$$

Korollar 2.1 (Besetzungsmuster der Skalierungsmatrizen)

- (i) Für jede Spalte $\tilde{G}_{s,t}^l$ von \tilde{G}^l gibt es genau einen nicht verschwindenden Eintrag $\tilde{G}_{s,t}^l$ mit ungeradem s : $\tilde{G}_{2t+1,t}^l$.
- (ii) Jede gerade Zeile $H_{2s,\cdot}^l$ von H^l hat genau einen nicht verschwindenden Eintrag: $H_{2s,s}^l$.
- (iii) Es gilt $\tilde{G}_{s,t}^l \neq 0$ genau dann, falls entweder $s = 2t + 1$ oder falls s gerade ist und $H_{2t+1,s/2}^l \neq 0$.

Der Beweis ist im Anhang A.2 zu finden.

Die Aussagen von Theorem 2.5 und Korollar 2.1 gelten auch für periodische Randbedingungen. In diesem Fall ist $S(l) = \{0, \dots, 2^l - 1\}$ und $T(l) = \{0, \dots, 2^{l-1} - 1\}$. Die Skalierungsmatrizen H^l, G^l, \dots enthalten nur die Einträge h_s, g_s, \dots wie bei den auf ganz \mathbb{R} definierten Funktionen, allerdings zyklisch fortgesetzt. Z.B. ist

$$\tilde{H}_{r,s}^l = h_{\text{mod}(r-2s, 2^l)}, \quad r \in S(l), s \in S(l-1).$$

2.5 Lifting-Wavelets

Die in den vorangegangenen Abschnitten vorgestellten Interpolets sind dadurch gekennzeichnet, daß die dualen Skalierungsfunktionen Dirac-Funktionale sind. Für diese "Funktionen" ist $\tilde{\gamma}^* = -1/2$. Demnach gelten die Normäquivalenzen aus Abschnitt 2.2 nur für $s \in]1/2, \gamma^*[$. Für die in Kapitel 4 beschriebenen Tensorprodukt-Konstruktionen für multivariate Basisfunktionen ist es aber beispielsweise notwendig, daß die univariaten Multiskalenfunktionen \mathcal{H}^1 - und \mathcal{L}^2 -stabil sind, damit die multivariaten Basisfunktionen \mathcal{H}^1 -stabil sein können. Tensorprodukte von Interpolets können demnach niemals \mathcal{H}^1 -stabil sein. Dieser Nachteil läßt sich mit dem sogenannten Lifting-Schema [CDP96, Swe97] beheben. Dabei wird aus den instabilen Interpolets durch eine geschickte Modifikation eine stabilere Waveletbasis konstruiert. Das Interessante daran ist, daß man auch im adaptiven Fall sehr flexibel zwischen den beiden Basen hin- und herschalten kann. Während die Interpolets für die Diskretisierung von Operatoren geeignet sind, eignen sich die Lifting-Wavelets besser für das Vorkonditionieren von linearen Gleichungssystemen.

Sei nun $\{\psi_{(l,t)}\}$ eine Basis von Interpolets. Dann sind die Lifting-Wavelets (genauer Lifting-Interpolets) wie folgt definiert:

$$\bar{\psi}_{(l,t)} := \begin{cases} \psi_{(l,t)} + \sum_{s \in S(l-1)} Q_{s,t}^l \phi^{(l-1,s)} & l > l_0 \\ \phi^{(l,t)} & l = l_0 \end{cases}. \quad (2.27)$$

Entscheidend ist hier der als Lifting-Korrektur bezeichnete Term $\sum_{s \in S(l-1)} Q_{s,t}^l \phi^{(l-1,s)}$. Die Koeffizienten $Q_{s,t}^l$ sind so definiert, daß die neuen primalen Wavelets \bar{M} verschwindende Momente haben:

$$\int_{\mathbb{R}} \bar{\psi}_{(l,t)} x^i dx = 0 \quad \text{für } 0 \leq i < \bar{M}. \quad (2.28)$$

Der Parameter \bar{M} kontrolliert den Stabilisierungseffekt. $\bar{M} = 2$ ist beispielsweise für unsere Zwecke (Vorkonditionieren) eine gute Wahl, wie man bald sehen wird. Neben der Momentenbedingung werden bei der Wahl von Q die folgenden Punkte berücksichtigt:

- Abgesehen von Randmodifikationen sollen die Wavelets $\bar{\psi}_{(l,t)}$ wieder Translate/Dilate einer Mutterfunktion $\bar{\psi}$ sein.
- Die Wavelets $\bar{\psi}_{(l,t)}$ sollen einen möglichst kleinen Träger haben.
- Abgesehen von Randmodifikationen sollen die Wavelets $\bar{\psi}_{(l,t)}$ wieder symmetrisch sein.

Aus diesen drei Forderungen und der Momentenbedingung (2.28) mit $\bar{M} = 2$ ergibt sich für innere Wavelets:

$$\bar{\psi}_{(l,t)} = \psi_{(l,t)} - \frac{1}{4}(\phi^{(l-1,t)} + \phi^{(l-1,t+1)}), \quad N \leq t \leq 2^{l-1} - N.$$

Für randangepaßte Wavelets wird nicht nur die Symmetrie-Bedingung aufgegeben, sondern eventuell auch eine der Moment-Bedingungen: Im Falle homogener Dirichlet-Randwerte brauchen die randangepaßten Wavelets kein verschwindendes 0-tes Moment haben. Für homogene Neumann-Randwerte kann am Rand auf $\int \bar{\psi}_{(l,t)} x dx = 0$ verzichtet werden. Dies beeinträchtigt nicht die Approximationseigenschaften (Jackson-Ungleichung) der primalen und dualen Multiskalenanalyse, so daß auch so die Voraussetzungen für die Normäquivalenzen gegeben sind, vgl. Abschnitt 2.2.

Der Stabilisierungseffekt beim Lifting-Schema beruht darauf, daß durch eine andere Wahl der Komplementräume zwischen V^l und V^{l-1} andere duale Skalierungsfunktionen bzw. Wavelets resultieren. Diese sind nicht mehr Dirac-Funktionale, sondern (falls \bar{M} hinreichend groß ist) Funktionen aus \mathcal{H}^s mit $s > 0$. Der entsprechende Glattheitsparameter, hier $\tilde{\gamma}^*$ genannt, ist also > 0 . Mit der Argumentation zu Theorem 2.2 folgt damit, daß dann die Lifting-Wavelets \mathcal{L}^2 -stabil sind.

Für eine Abschätzung von $\tilde{\gamma}^*$ können wieder die Techniken aus Abschnitt 2.2 benutzt werden. Dafür muß insbesondere das Symbol $\tilde{h}(z)$ der Skalierungskoeffizienten für die neuen dualen Skalierungsfunktionen bekannt sein. Dieses Symbol kann aus den Symbolen $\tilde{h}(z)$, $\tilde{g}(z)$ und dem Symbol $q(z)$ der Matrix Q^l berechnet werden. In Tabelle 2.3 sind einige Schätzwerte der Sobolev-Regularität der neuen dualen Skalierungsfunktionen für Lifting-Interpolets angegeben. Demnach wird für $\bar{M} = 2$ das Ziel $\tilde{\gamma}^* > 0$ erreicht.

\bar{M}	$N = 2$	$N = 4$	$N = 6$
0	-0.500	-0.500	-0.500
1	-0.354	-0.401	-0.431
2	0.171	0.290	0.358

Tabelle 2.3: Sobolev-Regularität $\tilde{\gamma}^*$ der dualen Skalierungsfunktionen für Lifting-Interpolets. ($\bar{M} = 0$: kein Lifting)

Für die Lifting-Wavelettransformation könnte man prinzipiell wieder Algorithmus WT benutzen, wobei dann die Skalierungsmatrizen \tilde{H}^l und \tilde{G}^l durch die Skalierungsmatrizen \tilde{H}^l bzw. \tilde{G}^l für die neuen dualen Funktionen zu ersetzen sind. Aus der Definition der Lifting-Wavelets und den Biorthogonalitätsrelationen (2.9) erhält man

$$(\Phi^{l-1}, \bar{\Psi}_l) = \Phi^l \cdot (H^l, G^l) \cdot \begin{pmatrix} I & Q^l \\ 0 & I \end{pmatrix} \Rightarrow (\tilde{H}^l, \tilde{G}^l) = (\tilde{H}^l, \tilde{G}^l) \cdot \begin{pmatrix} I & 0 \\ Q^{lT} & I \end{pmatrix}.$$

Es ist jedoch günstiger, die faktorisierte Form der Matrizen beizubehalten. In diesem Fall besteht ein Schritt der Wavelettransformation auf Level l aus den Teilschritten

$$\mathbf{u}_l = \tilde{G}^{lT} \cdot \mathbf{u}^l \quad \text{und} \quad \mathbf{u}^{l-1} = \tilde{H}^{lT} \cdot \mathbf{u}^l + Q^l \cdot \mathbf{u}_l .$$

Bei den Interpolets ist die Bandbreite von \tilde{H}^l und auch Q^l sehr klein (1 bzw. 2), wogegen die Bandbreite von \tilde{G}^l deutlich größer sein kann. Somit spart die faktorisierte Form Rechenoperationen, siehe z.B. [DS98].

Kapitel 3

Adaptive Algorithmen in 1D

Für den eindimensionalen Fall werden die adaptiven Algorithmen für die Multiskalenttransformation, die inverse Multiskalenttransformation und die Auswertung von Differentialoperatoren vorgestellt. Wie man in Kapitel 4 und 6 sehen wird, sind die entsprechenden Algorithmen im eigentlich interessanten, multivariaten Fall *vollständig* aus den univariaten Algorithmen aufgebaut. Es bietet sich deshalb an, die Algorithmen für den einfacheren, univariaten Fall zu beschreiben. Dabei werden wir eine exakte, formale Beschreibung geben, da bei einer eher verbalen Beschreibung die Gefahr sehr groß ist, daß wichtige Details einfach übergangen werden, die durchaus den Unterschied zwischen einem praktikablen und einem nichtpraktikablen Verfahren ausmachen können. Um jedoch die Verständlichkeit zu gewährleisten, werden die entscheidenden Ideen zu Beginn jedes Abschnittes skizziert. Zum Nachweis der Korrektheit der Algorithmen bzw. deren Komplexität werden wir einige Aussagen benötigen, deren Beweise recht technisch sind. Für das erste Durchlesen können diese übersprungen werden. Zunächst soll aber auf ein paar allgemeine Dinge im Zusammenhang mit adaptiven Algorithmen eingegangen werden.

Der wesentliche Punkt bei adaptiven Verfahren ist, nicht mehr die vollständigen Vektoren \mathbf{u}_l zu benutzen, sondern nur noch die Komponenten $\{u_{(l,t)}\}_{t \in \mathcal{T}(l)}$ bezüglich einer Indexmenge $\mathcal{T}(l) \subseteq T(l)$, vergleiche (2.25). Diese Indexmengen beschreiben die adaptive Basis. Alle auftretenden Funktionen, wie Lösung, rechte Seite oder Anfangsbedingungen haben somit die Form

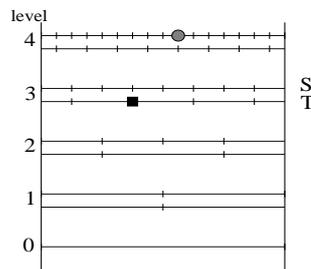
$$u = \sum_{l=l_0}^{\infty} \sum_{t \in \mathcal{T}(l)} u_{(l,t)} \psi_{(l,t)}.$$

Natürlich ist die Bestimmung von adaptiven Basen, die eine möglichst genaue Approximation der Lösung der Differentialgleichung erlauben, ein entscheidender Teil des gesamten Problems. Detaillierte Beschreibungen entsprechender Strategien werden in den Kapiteln 8 und 9 angegeben. In diesem Kapitel wird angenommen, daß die Indexmengen $\mathcal{T}(l)$ im wesentlichen bereits gegeben sind. Neben den Indexmengen $\mathcal{T}(l)$ benötigen wir für adaptive Erzeugendensystemdarstellungen auch Indexmengen $\mathcal{S}(l) \subseteq S(l)$, vergleiche Tabelle 2.2. An einigen Stellen werden wir die auftretenden Indexmengen $\mathcal{T}(l)$ und $\mathcal{S}(l)$ graphisch veranschaulichen. Dazu werden Diagramme analog zu Abbildung 3.1 benutzt. Die Elemente $s \in S(l)$ sind auf den oberen Linien an den Positionen $s2^{-l}$ mittels \bullet dargestellt und Elemente $t \in T(l)$ auf den unteren Linien an den Positionen $(2t+1)2^{-l}$ mittels \blacksquare . Die für die Beschreibung der adaptiven Algorithmen benutzte Notation ist in Tabelle 3.1 zusammengefaßt.

Eine der wichtigsten analytischen Eigenschaften der Skalierungsfunktionen war die polyno-

$\mathcal{T}(l), \mathcal{S}(l)$ $\mathcal{T} := \bigcup_{l=l_0}^{\infty} \{l\} \times \mathcal{T}(l)$ $L := \max\{l : \mathcal{T}(l) \neq \emptyset\}$ $\mathbf{u}_{\mathcal{T}(l)} := \{u_{(l,t)}\}_{t \in \mathcal{T}(l)}, \mathbf{u}^{\mathcal{S}(l)} := \{u^{(l,s)}\}_{s \in \mathcal{S}(l)}$ $\mathbf{u}_{\mathcal{T}} := \{u_{\mathcal{T}(l)}\}_{l=l_0}^L$ $\Psi_{\mathcal{T}(l)} := \{\psi_{(l,t)}\}_{t \in \mathcal{T}(l)}, \Phi^{\mathcal{S}(l)} := \{\phi^{(l,s)}\}_{s \in \mathcal{S}(l)}$ $\Psi_{\mathcal{T}} := \{\psi_{(l,t)}\}_{(l,t) \in \mathcal{T}}$ $\mathcal{S} := \{\mathcal{S}(l_0), \dots, \mathcal{S}(L)\}$ $\mathbf{u}^{\mathcal{S}} = \{u^{\mathcal{S}(l)}\}_{l=l_0}^L$	Teilmengen von $\mathcal{T}(l)$ bzw. $\mathcal{S}(l)$, s. Tab. 2.2 Indexmenge für die adaptive Basis maximales Level der adaptiven Basis Spaltenvektoren von Koeffizienten Zeilenvektoren von Basisfunktionen Indexmengen $\mathcal{S}(l)$ f. Erzeugendensystem Koeffizientenvektor f. Erzeugendensystem
--	---

Tabelle 3.1: Notation für adaptive Algorithmen in 1D.

Abbildung 3.1: Darstellung von Elementen aus $\mathcal{S}(l)$ und $\mathcal{T}(l)$. Hier: $9 \in \mathcal{S}(4)$ und $1 \in \mathcal{T}(3)$.

miale Exaktheit. Auch im adaptiven Fall sollen alle Polynome vom Grad $< N$ von der Basis aufgespannt werden. Wegen $\Psi_{l_0} = \Phi^{l_0}$ reicht es dazu,

$$\mathcal{T}(l_0) = \mathcal{S}(l_0) \quad (3.1)$$

zu fordern, d.h. das größte Level ist immer voll belegt. Da $\mathcal{S}(l_0)$ nur wenige Elemente hat, bleiben die Vorteile der adaptiven Basen aber erhalten.

Für die adaptiven Algorithmen ist es sehr oft notwendig, Abhängigkeiten zwischen zwei Indexmengen zu bestimmen, die durch das Besetzungsmuster von Matrizen induziert werden. Dazu dient der folgende Operator R . Seien Σ, Γ und $\Sigma' \subseteq \Sigma$ Indexmengen sowie A eine $\Sigma \times \Gamma$ -Matrix. Dann ist

$$R(A, \Sigma') := \{t \in \Gamma : A_{\Sigma', t} \neq 0\}.$$

Es werden also die Nicht-Null Einträge von Zeilen von A identifiziert. In Abbildung 3.2 ist $R(A, \Sigma')$ bzw. $R(A^T, \Sigma')$ für den Fall einer Bandmatrix skizziert. Komplementär zu R gibt es den Operator $C(\cdot, \cdot)$, durch den die Komponenten eines Matrix-Vektor-Produktes $A \cdot \mathbf{u}$ bestimmt werden, die exakt berechenbar sind, falls von \mathbf{u} nur die Komponenten $\{u_t\}_{t \in \Gamma'}$, $\Gamma' \subseteq \Gamma$, bekannt sind. Somit ist $C(\cdot, \cdot)$

$$C(A, \Gamma') := \{s \in \Sigma, \text{ so daß für alle } t \text{ mit } a_{s,t} \neq 0 \text{ gilt: } t \in \Gamma'\}.$$

3.1 Multiskalenttransformation

Es werden die Algorithmen für die adaptive Multiskalenttransformation und die Inverse davon angegeben. Diese beiden Operationen werden nicht nur für die Auswertung von nichtlinearen

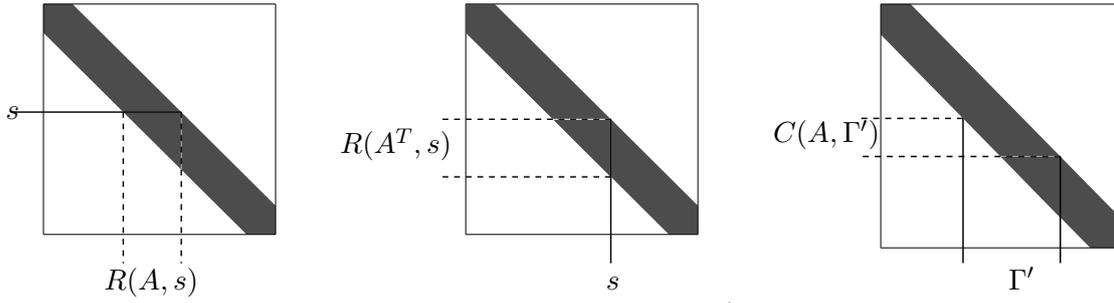


Abbildung 3.2: $R(A, s)$, $R(A^T, s)$ bzw. $C(A, \Gamma')$ für eine Bandmatrix.

Termen, wie bei Pseudo-Spektral Methoden, sondern auch für die Auswertung von Differentialoperatoren benötigt. Die wesentliche Idee für die adaptiven Algorithmen ist, wie in WT bzw. IWT vorzugehen, jedoch die Matrix-Vektor-Multiplikationen auf aktive Indexmengen einzuschränken. Bei der Multiskalenttransformation wollen wir beispielsweise jeweils nur $\mathbf{u}_{\mathcal{T}(l)}$ anstatt des vollständigen Vektors \mathbf{u}_l berechnen:

$$\mathbf{u}_{\mathcal{T}(l)} = \left(\tilde{G}_{\mathcal{S}(l), \mathcal{T}(l)}^l \right)^T \cdot \mathbf{u}^l .$$

Da \tilde{G}^l dünn besetzt ist, ist klar, daß für diese Operation nicht alle Komponenten von \mathbf{u}^l benötigt werden. Gebraucht wird $u^{(l,s)}$ genau dann, wenn $\tilde{G}_{s, \mathcal{T}(l)}^l \neq 0$, also falls $s \in R(\tilde{G}^{lT}, \mathcal{T}(l))$. Es gibt einen weiteren Unterschied zwischen der adaptiven und der nichtadaptiven Transformation in Bezug auf \mathbf{u}^l . In WT war \mathbf{u}^L der Eingabewert und die Vektoren $\{\mathbf{u}^l\}_{l=0}^{L-1}$ wurden sukzessive daraus berechnet. Für den adaptiven Algorithmus kann man sich das nicht leisten, da dann die Anzahl von notwendigen Rechenoperationen im allgemeinen nicht proportional zur Anzahl $|\mathcal{T}|$ von Freiheitsgraden der adaptiven Basis wäre. Aus diesem Grund besteht im adaptiven Fall die Eingabe aus einer adaptiven Erzeugendensystemdarstellung $\{\mathbf{u}^{S(l)}\}_{l=0}^L$, wobei entsprechend unseren einleitenden Betrachtungen die Indexmengen $\mathcal{S}(l)$ alle Indizes s enthalten müssen, für die $u^{(l,s)}$ für die Berechnung eines Koeffizienten $u_{(l,t)}$, $t \in \mathcal{T}(l)$, notwendig ist. Also

$$R(\tilde{G}^{lT}, \mathcal{T}(l)) \subseteq \mathcal{S}(l) . \quad (3.2)$$

In dem ursprünglichen Algorithmus WT sind die Vektoren \mathbf{u}^l jeweils aus \mathbf{u}^{l+1} mittels des Tiefpassfilters \tilde{H}^{l+1} berechnet worden. Speziell für Interpolets entsprach diese Operation einem einfachen Kopieren von Knotenwerten, vgl. Lemma 2.4. Somit ist diese Operation eigentlich überflüssig, falls die Eingabedaten $\{\mathbf{u}^{S(l)}\}$ die Bedingung

$$u^{(l,s)} = u^{(l+1,2s)} \quad \text{falls } s \in \mathcal{S}(l) \text{ und } 2s \in \mathcal{S}(l+1)$$

erfüllen. In Abschnitt 3.6 werden wir jedoch eine Situation kennenlernen, in der diese Bedingung nicht erfüllt ist. Wir nehmen dann an, daß $u^{(l+1,2s)}$ der “richtige” (bzw. genauere) Knotenwert von u ist und $u^{(l,s)}$ der “falsche” (bzw. weniger genauere) Knotenwert, der deshalb überschrieben werden soll. Leider kann im allgemeinen nicht der komplette Vektor $\mathbf{u}^{S(l)}$ aus $\mathbf{u}^{S(l+1)}$ berechnet werden, sondern nur der Teilvektor $\mathbf{u}^{\Delta(l)}$ mit Indizes s aus

$$\Delta(l) := \{s \in \mathcal{S}(l) : 2s \in \mathcal{S}(l+1)\} . \quad (3.3)$$

Der vollständige adaptive Algorithmus lautet dann

```

Funktion  $\mathbf{u}_{\mathcal{T}} = \text{AWT}(\mathbf{u}^{\mathcal{S}}, \mathcal{S}, \Delta)$ 
for  $l = L$  downto  $l_0 + 1$ 
     $\mathbf{u}_{\mathcal{T}(l)} = (\tilde{G}_{\mathcal{S}(l), \mathcal{T}(l)}^l)^T \cdot \mathbf{u}^{\mathcal{S}(l)}$ 
     $\mathbf{u}^{\Delta(l-1)} = (\tilde{H}_{\mathcal{S}(l), \Delta(l-1)}^l)^T \cdot \mathbf{u}^{\mathcal{S}(l)}$ 
end

```

Algorithmus AWT ist über die Indexmengen \mathcal{S} und Δ parametrisiert worden, da wir später AWT für eventuell andere Indexmengen \mathcal{S}' und Δ' benötigen werden. Auf jeden Fall muß aber das zweite Argument \mathcal{S} Bedingung (3.2) erfüllen. Die allgemeine Form von AWT wurde auch in der Implementierung benutzt, wodurch der entsprechende Code oft wiederverwendet werden konnte.

Wir kommen nun zur inversen Multiskalenttransformation, welche aus den Koeffizienten $\mathbf{u}_{\mathcal{T}}$ die Werte $\mathbf{u}^{\mathcal{S}}$ berechnen soll. Der Algorithmus lautet

```

Funktion  $\mathbf{u}^{\mathcal{S}} = \text{AIWT}(\mathbf{u}_{\mathcal{T}}, \mathcal{S})$ 
 $\mathbf{u}^{\mathcal{S}(l_0)} = \mathbf{u}_{\mathcal{T}(l_0)}$ 
for  $l = l_0 + 1$  to  $L$ 
     $\mathbf{u}^{\mathcal{S}(l)} = H_{\mathcal{S}(l), \mathcal{S}(l-1)}^l \cdot \mathbf{u}^{\mathcal{S}(l-1)} + G_{\mathcal{S}(l), \mathcal{T}(l)}^l \cdot \mathbf{u}_{\mathcal{T}(l)}$ 
end

```

Wir fordern, daß die mit AIWT und die mit IWT berechneten Werte $\mathbf{u}^{(l,s)}$ für $s \in \mathcal{S}(l)$ exakt übereinstimmen sollen, falls angenommen wird, daß alle Multiskalenkoeffizienten $u_{(l,t)}$ mit $(l,t) \notin \mathcal{T}$ identisch null sind. Wegen dieser Exaktheitsforderung müssen im Syntheseschritt $\mathbf{u}^{\mathcal{S}(l)} = H_{\mathcal{S}(l), \mathcal{S}(l-1)}^l \cdot \mathbf{u}^{\mathcal{S}(l-1)} + \dots$ alle Koeffizienten $u^{(l-1,s)}$ berücksichtigt werden, die in $\mathbf{u}^{\mathcal{S}(l)}$ eingehen können. D.h. die Indexmenge $\mathcal{S}(l-1)$ muß der Bedingung

$$R(H^l, \mathcal{S}(l)) \subseteq \mathcal{S}(l-1) \quad (3.4)$$

genügen. Fassen wir die Bedingungen (3.2) und (3.4) zusammen, so erhalten wir den folgenden Algorithmus zur rekursiven Bestimmung von $\mathcal{S}(l)$:

```

Funktion  $\mathcal{S} = \text{IndexMengeAWT}(\mathcal{T})$ 
 $\mathcal{S}(L) = R(\tilde{G}^{LT}, \mathcal{T}(L))$ 
for  $l = L - 1$  downto  $l_0$ 
     $\mathcal{S}(l) = R(\tilde{G}^{lT}, \mathcal{T}(l)) \cup R(H^{l+1}, \mathcal{S}(l+1))$ 
end

```

Auf die Frage des Rechenaufwandes für AWT bzw. AIWT gehen wir dezidiert im nächsten Abschnitt ein.

3.2 Bedingungen an \mathcal{T}

Bis auf Formel (3.3) für $\Delta(l)$ wurden im vorangegangenen Abschnitt keine speziellen Eigenschaften der Interpolets berücksichtigt. Diese führen jedoch zu einigen Vereinfachungen. Die wichtigste Eigenschaft in diesem Zusammenhang ist, daß die dualen Skalierungsfunktionen Dirac-Funktionale sind und somit die Koeffizienten $u^{(l,s)}$ Werte von u in den Gitterpunkten $s2^{-l}$ sind. Bei der Multiskalenttransformation entsprechen deshalb die Eingabewerte $\mathbf{u}^{\mathcal{S}}$ den

Werten u auf dem adaptiven Gitter

$$\Omega_{\mathcal{T}} := \bigcup_{l=l_0}^L 2^{-l} \mathcal{S}(l) .$$

Natürlich tauchen in den Vektoren $\mathbf{u}^{\mathcal{S}(l)}$ eine ganze Reihe von Knotenwerten mehrfach auf. So ist z.B. $u^{(l,s)} = u^{(l+1,2s)}$. Es stellt sich die Frage, ob durch solche Redundanzen die Algorithmen AWT und AIWT überhaupt noch einen Rechenaufwand haben, der proportional zu $|\mathcal{T}|$ ist. Im allgemeinen ist das tatsächlich nicht der Fall. Sei z.B. für $\mathcal{T} = (L, 10)$ für ein $L \geq l_0$. Also ist $|\mathcal{T}| = 1$. Für die in AWT bzw. AIWT benötigten Indexmengen $\mathcal{S}(l)$ gilt aber $\sum_{l=l_0}^L |\mathcal{S}(l)| \sim L$, so daß der Rechenaufwand von AWT bzw. AIWT im Verhältnis zu $|\mathcal{T}|$ beliebig schlecht werden kann.

Das zentrale Thema dieses Abschnittes ist eine Bedingung an \mathcal{T} , die solche entarteten Fälle ausschließt und somit unter anderem einen Rechenaufwand $\sim |\mathcal{T}|$ gewährleistet. Wir untersuchen nun die Beziehung zwischen \mathcal{T} und $\Omega_{\mathcal{T}}$ etwas näher. Dazu ist es für Interpolets sinnvoll, jedem Index (l, t) , $t \in T(l)$, einen dyadischen Gitterpunkt zuzuordnen:

$$x(l, t) := \begin{cases} (2t+1)2^{-l} & l > l_0 \\ t2^{-l} & l = l_0 \end{cases} .$$

Wie man sich leicht überlegt, gibt es zu einem beliebigen dyadischen Gitterpunkt $s2^{-k}$, $s \in S(k)$, genau ein (l, t) , $t \in T(l)$, mit $x(l, t) = s2^{-k}$. Man kann also auf der Menge der dyadischen Gitterpunkte die Umkehrabbildung x^{-1} definieren.

Für die Durchführbarkeit von AWT hatte die Indexmenge $\mathcal{S}(l)$ für $t \in T(l)$ alle Indizes $R(\tilde{G}^{lT}, t)$ zu enthalten, siehe (3.2). Die entsprechenden Gitterpunkte $2^{-l}R(\tilde{G}^{lT}, t)$ sind besonderen Multiskalenindizes zugeordnet.

Lemma 3.1 (Hierarchische Väter) *Für $(l, t), t \in T(l), l > l_0$, definieren wir*

$$V(l, t) := x^{-1}\left(2^{-l}R(\tilde{G}^{lT}, t)\right) \setminus (l, t) .$$

Für $(l', t') \in V(l, t)$ gilt dann $l' < l$. Wir bezeichnen deshalb $V(l, t)$ als die Menge der hierarchischen Väter von (l, t) .

Beweis: Wegen Korollar 2.1 (i) sind alle Indizes in $R(\tilde{G}^{lT}, t) \setminus 2t+1$ gerade. Daraus folgt mit der Definition von $x(\cdot)$ die Behauptung. \square

Abbildung 3.3 zeigt einen Index (l, t) (■) und die hierarchischen Väter (■) für Interpolets $N = 2$. Die wesentliche Aussage dieses Abschnittes ist nun in folgendem Theorem zusammengefaßt.

Theorem 3.1 *Die Abbildung x^{-1} beschreibt eine Bijektion zwischen $\Omega_{\mathcal{T}}$ und \mathcal{T} , d.h.*

$$\mathcal{T} = x^{-1}(\Omega_{\mathcal{T}}) , \tag{3.5}$$

genau dann, falls alle hierarchischen Väter eines Index $(l, t) \in \mathcal{T}$ selbst zu \mathcal{T} gehören,

$$\forall (l, t) \in \mathcal{T} : V(l, t) \subseteq \mathcal{T} \quad (\text{Kegel-Bedingung}) . \tag{3.6}$$

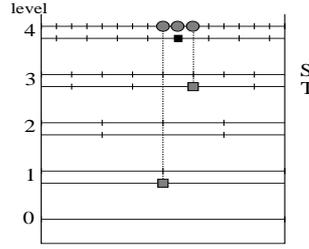


Abbildung 3.3: Die hierarchische Väter von $(4,4)$ bei Interpolets $N = 2$: $(1,0)$ und $(3,2)$. Zusätzlich ist mittels \bullet die Menge $R(\tilde{G}^{4T}, 4)$ dargestellt.

Der Beweis des Theorems ist im Anhang A.3 zu finden. Die Bezeichnung *Kegel-Bedingung* ist an eine explizite Darstellung von $V(l, t)$ für den Fall stückweise linearer Interpolets angelehnt [Gri98]. Analoge Bedingungen findet man auch bei der sogenannten Baum-Approximation [CDDD99]. Wir wollen auf ein paar Folgerungen von Theorem 3.1 eingehen:

Bemerkung 3.1 (Auffüllen) Da die hierarchischen Väter von (l, t) auf größeren Leveln leben, ist es möglich Bedingung (3.6) durch sukzessives Auffüllen der Indexmengen $\mathcal{T}(l)$ in absteigender Level-Reihenfolge $l = L, \dots, l_0$ zu erzwingen.

Bemerkung 3.2 (Interpolationsproblem) Unabhängig davon, ob \mathcal{T} die Kegel-Bedingung erfüllt oder nicht, werden mit AIWT genau die Knotenwerte $u(\Omega_{\mathcal{T}})$ einer Funktion $u = \Psi_{\mathcal{T}} \cdot \mathbf{u}_{\mathcal{T}}$ auf dem Gitter $\Omega_{\mathcal{T}}$ berechnet. AIWT kann man also als eine lineare Abbildung von $\mathbb{R}^{|\mathcal{T}|} \rightarrow \mathbb{R}^{|\Omega_{\mathcal{T}}|}$ interpretieren. In dem gleichen Sinne ist AWT eine Abbildung von $\mathbb{R}^{|\Omega_{\mathcal{T}}|} \rightarrow \mathbb{R}^{|\mathcal{T}|}$. Angewandt auf $u(\Omega_{\mathcal{T}})$ bzw. $\{\mathbf{u}^{S(l)}\}$ liefert diese genau die ursprünglichen Koeffizienten $\mathbf{u}_{\mathcal{T}}$ zurück. D.h. $\text{AWT} \circ \text{AIWT} = \text{id}$. Falls nun aber (3.6) bzw. (3.5) erfüllt ist, gilt $|\mathcal{T}| = |\Omega_{\mathcal{T}}|$. In diesem Fall müssen die Abbildungen AWT bzw. AIWT Bijektionen sein. Insbesondere liefert AWT zu vorgegebenen Knotenwerten $u(\Omega_{\mathcal{T}})$ den eindeutigen Interpolant bzgl. der Basis $\Psi_{\mathcal{T}}$. Dies erlaubt die einfache Auswertung nichtlinearer Funktionale. Für $u \in \text{span} \Psi_{\mathcal{T}}$ sei z.B. u^2 gesucht. u^2 selbst kann in allgemeinen nicht mit der Basis $\Psi_{\mathcal{T}}$ dargestellt werden. Man kann aber mittels AWT und AIWT den Interpolant von u^2 berechnen. Auf diese Art erhält man ein einfaches Schema für die Auswertung von nichtlinearen Funktionalen, die für allgemeine Multiskalendiskretisierungen recht schwierig ist.

Bemerkung 3.3 (Petrov-Galerkin) Die obige Lösung des Interpolationsproblems erlaubt noch eine andere Interpretation. Bedingung (3.5) besagt im wesentlichen, daß

$$\text{span} \tilde{\Psi}_{\mathcal{T}} = \text{span} \{ \delta(\cdot, -z) : z \in \Omega_{\mathcal{T}} \} .$$

Deshalb werden mittels AWT die Koeffizienten $\{ \langle \tilde{\phi}^{(l,s)}, u \rangle \}$ eindeutig in die Koeffizienten $\{ \langle \tilde{\psi}^{(l,t)}, u \rangle \}$ umgerechnet. Bei AIWT genau umgekehrt. D.h. wir können beliebig zwischen zwei äquivalenten Basen von Testfunktionen hin- und herschalten. Angewandt auf das obige Problem der Auswertung von u^2 hat man folgendes Schema

$$\left\{ \langle \tilde{\psi}^{(l,t)}, u \rangle \right\} \xrightarrow{\text{AIWT}} \left\{ \langle \tilde{\phi}^{(l,s)}, u \rangle \right\} \xrightarrow{(\cdot)^2} \left\{ \langle \tilde{\phi}^{(l,s)}, u \rangle^2 = \langle \tilde{\phi}^{(l,s)}, u^2 \rangle \right\} \xrightarrow{\text{AWT}} \left\{ \langle \tilde{\psi}^{(l,t)}, u^2 \rangle \right\}$$

Somit gestatten AWT und AIWT eine exakte Petrov-Galerkin Diskretisierung nichtlinearer Funktionale. Die Testfunktionen sind dabei die dualen Wavelets.

Bemerkung 3.4 (Kompakte Speicherung) Zur Auswertung nichtlinearer Funktionale im multivariaten Fall wird es notwendig sein, AIWT für eine große Zahl von adaptiven eindimensionalen Basen auszuführen und die Ergebnisse zu speichern. Dabei können natürlich

die Redundanzen genutzt werden, so daß letztlich nur die Werte bzgl. adaptiver eindimensionaler Gitter gespeichert werden brauchen. Wegen (3.5) können diese Werte sogar an genau den gleichen Plätzen gespeichert werden, wie die ursprünglichen Multiskalenkoeffizienten. Das führt zu enormen Vereinfachungen in der Struktur der Algorithmen. Erst durch diese Vereinfachungen wird beispielsweise die Multiskalentransformation im multivariaten Fall leicht durchführbar. Die entsprechenden Algorithmen bestehen aus der Hintereinanderanwendung der folgenden Algorithmen für die adaptive (inverse) Multiskalentransformation mit kompakter Speicherung:

Funktion $v_{\mathcal{T}}=AIWTKS(u_{\mathcal{T}})$ $\{\mathcal{S}(l)\}=\text{IndexMengeAWT}(\mathcal{T})$ $\{u^{\mathcal{S}(l)}\}=AIWT(u_{\mathcal{T}}, \{\mathcal{S}(l)\})$ for $l = l_0$ to L für jedes $s \in \mathcal{S}(l)$ setze $v_{x^{-1}(2^{-l}s)} = u^{(l,s)}$ end	Funktion $v_{\mathcal{T}}=AWTKS(u_{\mathcal{T}})$ $\{\mathcal{S}(l)\}=\text{IndexMengeAWT}(\mathcal{T})$ setze $\{\Delta(l)\}$ wie in (3.3) for $l = l_0$ to L für jedes $s \in \mathcal{S}(l)$ setze $v^{(l,s)} = u_{x^{-1}(2^{-l}s)}$ end $u_{\mathcal{T}}=AWT(\{v^{\mathcal{S}(l)}\}, \{\mathcal{S}(l)\})$
--	--

Bemerkung 3.5 (Rechenaufwand) Wir werden in Lemma 3.3 zeigen, daß $|\mathcal{T}| \sim \sum_{l=l_0}^L |\mathcal{S}(l)|$. Somit haben AWT und AIWT einen Rechenaufwand, der ungefähr proportional zu $|\mathcal{T}|$ ist.

Wir haben bereits erwähnt, daß es durch rekursives Auffüllen von fein nach grob möglich ist, aus einer Indexmenge \mathcal{T}^0 , die (3.6) eventuell noch nicht erfüllt, eine Indexmenge \mathcal{T} zu berechnen die (3.6) erfüllt. Dieses Vorgehen hat den Nachteil, daß gegebenenfalls für den Index (l, t) Väter (l', t') auf allen darunterliegenden Leveln $l_0 \leq l' < l$ eingefügt werden müssen. Weiterhin werden viele dieser Indizes (l', t') mehrfach eingefügt. Um diese Probleme zu beheben, werden wir nun eine zur Kegel-Bedingung weitgehend äquivalente Bedingung angeben, die einen effizienten Algorithmus zum Auffüllen erlaubt. Es zeigt sich, daß für $(l, t) \in \mathcal{T}$ jeweils nur die Väter auf Level $l-1$ eingefügt werden müssen. Zum Nachweis der Äquivalenz benötigen wir folgende technische Bedingung an die primalen Skalierungsfunktionen:

Annahme 3.1 Für $l > l_0$ und $t \in T(l)$ gilt bei nichtperiodischen Randbedingungen

$$H_{2t+1,t}^l \neq 0 \quad \text{falls } t \in S(l-1) \quad \text{sowie} \quad H_{2t+1,t+1}^l \neq 0 \quad \text{falls } t+1 \in S(l-1). \quad (3.7)$$

Im Fall periodischer Randbedingungen ist $H_{2t+1,t}^l \neq 0$ und $H_{2t+1, \text{mod}(t+1, 2^{l-1})}^l \neq 0$.

Als Beispiel betrachten wir die Skalierungsfunktionen der Interpolets für den periodischen Fall. Dann ist $H_{2t+1,s}^l = h_{2t+1-2s}$. Damit $H_{2t+1,s}^l \neq 0$ ist, muß also

$$2t+1-2s \in \{-N+1, -N+3, \dots, -1, 0, 1, 3, \dots, N-1\}$$

sein, siehe Theorem 2.4, woraus

$$R(H^l, 2t+1) = t + \left\{ -\frac{N}{2} + 1, -\frac{N}{2} + 2, \dots, \frac{N}{2} \right\}$$

folgt. Da $N \geq 2$ ist, folgt Annahme 3.1. Im Fall von Interpolets auf den Intervall sind für Bedingung (3.7) nur noch die randangepaßten Skalierungskoeffizienten zu überprüfen. Da diese Skalierungskoeffizienten unabhängig vom Level sind, braucht man das nur für z.B. l_0+1 zu tun. Für alle hier konstruierten Interpolets auf dem Intervall ist (3.7) erfüllt. Betrachte dazu auch Abbildung 2.2.

Lemma 3.2 (Reduzierte Kegel-Bedingung) *Für Interpolets erfüllt eine Indexmenge \mathcal{T} genau dann die Kegel-Bedingung (3.6), falls für alle $(l, t) \in \mathcal{T}$, die hierarchischen Väter auf Level $l - 1$ auch zu \mathcal{T} gehören, d.h.*

$$V(l, t) \cap \{l - 1\} \times \mathcal{T}(l - 1) \subseteq \mathcal{T}. \quad (3.8)$$

Der Beweis ist in Anhang A.4 angegeben. Wir erhalten den optimierten Auffüll-Algorithmus:

```

Funktion  $\mathcal{T} = \text{KegelFill}(\mathcal{T}^0)$ 
 $\mathcal{T}(L) = \mathcal{T}^0(L)$ 
for  $l = L$  downto  $l_0 + 2$ 
     $\mathcal{T}(l - 1) = \mathcal{T}^0(l - 1) \cup \{t : 2(2t + 1) \in R(\tilde{G}^{lT}, \mathcal{T}(l))\}$ 
end
 $\mathcal{T}(l_0) = S(l_0)$ 

```

Die letzte noch offene Frage betrifft den Rechenaufwand der Algorithmen AWT und AIWT. Diese wird durch folgendes Lemma beantwortet.

Lemma 3.3 *Falls \mathcal{T} die Kegel-Bedingung (3.6) erfüllt und falls Bedingung (3.7) erfüllt ist, dann gilt für die Indexmengen $\mathcal{S}(l)$ aus Algorithmus \mathcal{S} ,*

$$\mathcal{S}(l) = R(\tilde{G}^{lT}, \mathcal{T}(l)). \quad (3.9)$$

Da die Matrizen \tilde{G}^l Bandmatrizen sind, folgt $|\mathcal{S}(l)| \lesssim |\mathcal{T}(l)|$ und somit $\sum_{l=l_0}^L |\mathcal{S}(l)| \lesssim |\mathcal{T}|$. Die Algorithmen AWT und AIWT haben also einen Rechenaufwand $\sim |\mathcal{T}|$.

Der Beweis ist in Anhang A.5 angegeben.

3.3 Lifting-Wavelettransformation

Die Interpolets haben sehr gute Eigenschaften für Approximation und numerische Algorithmen. Ein großer Nachteil sind jedoch die schlechten Stabilitätseigenschaften in Bezug auf Normäquivalenzen. Diese werden beispielsweise für das Vorkonditionieren von elliptischen linearen Gleichungssystemen benötigt. Ein Basiswechsel von den Interpolets zu den Lifting-Wavelets bzw. umgekehrt kann diesen Nachteil jedoch beheben. Man kann dann die Vorteile beider Basen nutzen.

In diesem Abschnitt wird untersucht, unter welchen Bedingungen im adaptiven Fall der Basiswechsel zwischen Interpolets und Lifting-Wavelets

$$u = \Psi_{\mathcal{T}} \cdot \mathbf{u}_{\mathcal{T}} \longleftrightarrow \bar{\Psi}_{\mathcal{T}} \cdot \bar{\mathbf{u}}_{\mathcal{T}} = u \quad (3.10)$$

möglich ist und wie entsprechende Algorithmen aussehen. Das zentrale Ergebnis dabei ist, daß die Indexmenge \mathcal{T} eine Bedingung ähnlich der Kegel-Bedingung erfüllen muß. Die dafür zusätzlich einzufügenden Indizes beeinträchtigen aber nicht die Komplexität des adaptiven Verfahrens. Die Algorithmen sind Modifikationen von AWT bzw. AIWT.

Angenommen es gilt $\text{span} \Psi_{\mathcal{T}} = \text{span} \bar{\Psi}_{\mathcal{T}}$. Dann gibt es zu $u = \Psi_{\mathcal{T}} \cdot \mathbf{u}_{\mathcal{T}}$ die alternative Darstellung $\bar{\Psi}_{\mathcal{T}} \cdot \bar{\mathbf{u}}_{\mathcal{T}}$. Diese soll nachfolgend bestimmt werden. Mit der Definition der Lifting-Wavelets bedeutet das

$$\sum_{l=l_0}^L \Psi_{\mathcal{T}(l)} \cdot \mathbf{u}_{\mathcal{T}(l)} \stackrel{!}{=} \Phi^{l_0} \cdot \bar{\mathbf{u}}_{\mathcal{S}(l_0)} + \sum_{l=l_0+1}^L \left(\Psi_{\mathcal{T}(l)} + \Phi^{l-1} \cdot Q_{\mathcal{S}(l-1), \mathcal{T}(l)}^l \right) \cdot \bar{\mathbf{u}}_{\mathcal{T}(l)}. \quad (3.11)$$

Testet man diese Gleichung mit $\tilde{\Psi}_{\mathcal{T}(L)}$, folgt $\mathbf{u}_{\mathcal{T}(L)} = \bar{\mathbf{u}}_{\mathcal{T}(L)}$. Wenn man also die Multiskalenkoeffizienten $\mathbf{u}_{\mathcal{T}}$ von u kennt, kann daraus sofort $\bar{\mathbf{u}}_{\mathcal{T}(L)}$ abgelesen werden. Um dieses Schema rekursiv fortführen zu können, muß nun die Teilfunktion $u' := \sum_{l=l_0}^{L-1} \bar{\Psi}_{\mathcal{T}(l)} \cdot \bar{\mathbf{u}}_{\mathcal{T}(l)}$ betrachtet werden. Wegen (3.11) ist

$$u' = \sum_{l=l_0}^{L-1} \Psi_{\mathcal{T}(l)} \cdot \mathbf{u}_{\mathcal{T}(l)} - \Phi^{L-1} \cdot Q_{S(L-1),\mathcal{T}(L)}^L \cdot \mathbf{u}_{\mathcal{T}(L)}. \quad (3.12)$$

Falls der Lifting-Korrektur-Term mittels der Interpolet-Basis bis Level $L - 1$ dargestellt werden kann, d.h.

$$-\Phi^{L-1} \cdot Q_{S(L-1),\mathcal{T}(L)}^L \cdot \mathbf{u}_{\mathcal{T}(L)} \in \text{span}\{\Psi_{\mathcal{T}(l)}\}_{l=l_0}^{L-1}, \quad (3.13)$$

dann ist u' als $u' = \sum_{l=l_0}^{L-1} \Psi_{\mathcal{T}(l)} \cdot \mathbf{u}'_{\mathcal{T}(l)}$ darstellbar. Wie oben könnte man dann $\bar{\mathbf{u}}_{\mathcal{T}(L-1)} = \mathbf{u}'_{\mathcal{T}(L-1)}$ ablesen. Letztlich ist somit (3.13) eine notwendige Voraussetzung für den Basiswechsel. Obige Argumentation kann auf niedrigeren Leveln wiederholt werden, was zur folgenden Aussage führt.

Lemma 3.4 (Basiswechsel und Darstellbarkeit der Lifting-Korrektur mit $\Psi_{\mathcal{T}}$)

Es gilt $\text{span}\Psi_{\mathcal{T}} = \text{span}\bar{\Psi}_{\mathcal{T}}$ genau dann, falls für $l_0 \leq l < L$ gilt

$$\text{span}\Phi^l \cdot Q_{S(l),\mathcal{T}(l+1)}^{l+1} \subseteq \text{span}\{\Psi_{\mathcal{T}(k)}\}_{k=l_0}^l. \quad (3.14)$$

D.h. in diesem Fall ist der Basiswechsel (3.10) für alle $u \in \text{span}\Psi_{\mathcal{T}}$ möglich.

Der Beweis ist in Anhang A.6 angegeben. Bedingung (3.14) ist natürlich noch sehr unhandlich. Die Struktur ist aber klar: aus der Indexmenge $\mathcal{T}(l+1)$ ergeben sich Forderungen an die Indexmengen $\mathcal{T}(k)$ für $k \leq l$. Diese Indexmengen müssen groß genug sein, damit (3.14) gilt. Es wird also wie bei der Kegel-Bedingung möglich sein, durch sukzessives Auffüllen von fein nach grob Bedingung (3.14) zu erzwingen.

Um Bedingung (3.14) zugänglicher zu machen, untersuchen wir den Term $\text{span}\Phi^l \cdot Q_{S(l),\mathcal{T}(l+1)}^{l+1}$ genauer. Dieser Raum wird von Skalierungsfunktionen $\phi^{(l,s)}$ aufgespannt, für die $Q_{s,\mathcal{T}(l+1)}^{l+1} \neq 0$ ist, d.h. s muß in $R(Q^{l+1T}, \mathcal{T}(l+1)) =: W(l)$ sein. Die Matrizen Q^l sind sehr dünn besetzt, so daß $W(l)$ nur wenig mehr Elemente als $\mathcal{T}(l+1)$ enthält, siehe Bild 3.4. Wir versuchen nun,

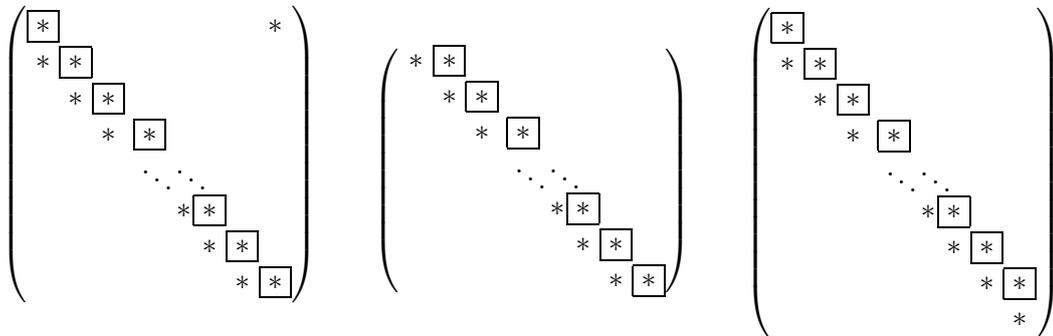


Abbildung 3.4: Besetzungsmuster von Q^l : Links: periodisch, Mitte: D./N.-Randwerte, Rechts: ohne Randbedingungen. Die Diagonaleinträge $Q_{t,t}^l$ sind eingerahmt.

die wichtigen Skalierungsfunktionen $\Phi^{W(l)}$ mittels Ψ_l und Φ^{l-1} darzustellen. Dazu nutzen wir, daß aus (2.7) und den Biorthogonalitätsbeziehungen (2.9) folgt

$$\Phi^l = \Phi^{l-1} \cdot \tilde{H}^{lT} + \Psi_l \cdot \tilde{G}^{lT} .$$

Damit $\Phi^{W(l)}$ mit $\{\Psi_{\mathcal{T}(k)}\}_{k=l_0}^l$ darstellbar ist, muß gelten:

1. Alle Indizes $t \in T(l)$ mit $\tilde{G}_{W(l),t}^l \neq 0$ sind auch in $\mathcal{T}(l)$. D.h. $R(\tilde{G}^l, W(l)) \subseteq \mathcal{T}(l)$.
2. Alle Skalierungsfunktionen $\phi^{(l-1,s)}$ mit $\tilde{H}_{W(l),s}^l \neq 0$ liegen auch in $\text{span} \{\Psi_{\mathcal{T}(k)}\}_{k=l_0}^{l-1}$.

Der erste Punkt führt unmittelbar auf eine rekursive Bedingung an \mathcal{T} :

Definition 3.1 (Lifting-Bedingung) Eine Indexmenge \mathcal{T} erfüllt die Lifting-Bedingung, falls für $l_0 \leq l < L$ gilt

$$R(\tilde{G}^l, W(l)) = R(\tilde{G}^l, R(Q^{l+1T}, \mathcal{T}(l+1))) \subseteq \mathcal{T}(l) . \quad (3.15)$$

Diese Bedingung reicht schon, um auch die zweite Forderung zu erfüllen.

Lemma 3.5 Falls (3.7) gilt und \mathcal{T} die Kegel- und die Lifting-Bedingung erfüllt ist, folgt für $l_0 \leq l < L$ Bedingung (3.14). D.h. der Basiswechsel (3.10) ist für alle $u \in \text{span} \Psi_{\mathcal{T}} = \text{span} \bar{\Psi}_{\mathcal{T}}$ möglich.

Der Beweis ist in Anhang A.7 angegeben. Die Bedingungen aus Lemma 3.5 führen unmittelbar zu folgendem Auffüll-Algorithmus:

```

Funktion  $\mathcal{T} = \text{LiftingFill}(\mathcal{T}^0)$ 
 $\mathcal{T}(L) = \mathcal{T}^0(L)$ 
for  $l = L$  downto  $l_0 + 2$ 
     $\mathcal{T}(l-1) = \mathcal{T}^0(l-1) \cup \{t : 2t+1 \in R(\tilde{G}^{lT}, \mathcal{T}(l))\} \cup R(\tilde{G}^{l-1}, R(Q^{lT}, \mathcal{T}(l)))$ 
end
 $\mathcal{T}(l_0) = S(l_0)$ 

```

In Abbildung 3.5 ist an einem Beispiel der Unterschied zwischen Indexmengen, die nur die Kegel-Bedingung erfüllen und Indexmengen, die auch noch (3.15) erfüllen, dargestellt. Die ursprünglich Indexmenge sei $\mathcal{T}^0 = \{(4, 3)\}$. Die grauen Quadrate stellen die Indizes von \mathcal{T} dar, die für die Kegel-Bedingung notwendig sind und die weißen Quadrate stellen die durch (3.15) zusätzlich induzierten Indizes dar.

Lemma 3.6 Seien \mathcal{T}^1 und \mathcal{T}^2 mittels $\text{KegelFill}((L, t), \mathcal{T}^1)$ bzw. $\text{LiftingFill}((L, t), \mathcal{T}^2)$ berechnet worden, dann gilt

$$|\mathcal{T}^1(l)| \simeq |\mathcal{T}^2(l)| .$$

Beweis: Wir definieren für $i = 1, 2$

$$a^i(l) := \min\{s \in \mathcal{T}^i(l)\}, \quad e^i(l) := \max\{s \in \mathcal{T}^i(l)\} .$$

In beiden Fällen wird $\mathcal{T}^i(l)$ mittels einer Rekursionsformel aus $\mathcal{T}^i(l+1)$ berechnet, so daß es eine Konstante C gibt mit

$$a^i(l) \geq \frac{a^i(l+1)}{2} - C \quad \text{und} \quad e^i(l) \leq \frac{e^i(l+1)}{2} + C .$$

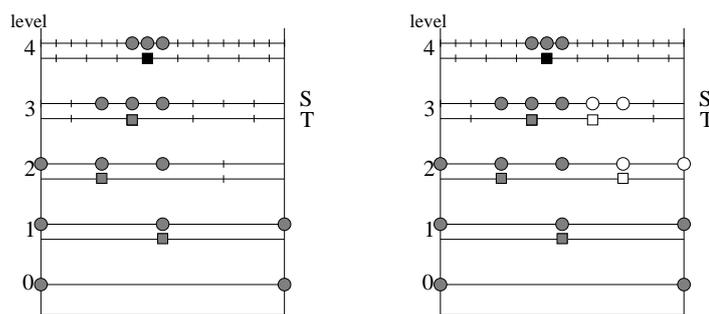


Abbildung 3.5: Unterschied zwischen den von $(4,3)$ (■) induzierten Indexmengen. Links: nur Kegel-Bedingung (graue Quadrate). Rechts: Kegel-Bedingung und (3.15) (graue + weiße Quadrate).

Also

$$a^i(l) \geq 2^{l-L}t - 2C \quad \text{und} \quad e^i(l) \leq 2^{l-L}t + 2C .$$

Daraus folgt die Behauptung. \square

Aus obigem Lemma folgt, daß für eine Indexmenge \mathcal{T}^0 die Algorithmen KegelFill und Lifting-Fill ähnlich große Indexmengen erzeugen. Demnach beeinträchtigt die zusätzliche Bedingung (3.14) nicht die asymptotische Komplexität der adaptiven Verfahren. In der Praxis müssen gegenüber (3.6) erfahrungsgemäß ca. 10%-30% mehr Indizes aufgewendet werden, um auch (3.14) zu erfüllen.

Wir kommen nun zu den eigentlichen Algorithmen für den Basiswechsel. Wir beginnen mit "hierarchisch \rightarrow lifting". Wir hatten mit Hilfe von (3.11) gesehen, daß $\bar{\mathbf{u}}_{\mathcal{T}(L)} = \mathbf{u}_{\mathcal{T}(L)}$. Daraus folgt, daß $\bar{\mathbf{u}}_{\mathcal{T}(L)}$ mittels $\tilde{G}^{LT} \cdot \mathbf{u}^{S(L)}$ aus den Knotenwerten von u berechnet werden kann. Um dieses Schema rekursiv weiterführen zu können, benötigen wir die Knotenwerte von u' aus (3.12).

Lemma 3.7 (Korrektur von Knotenwerten bei Lifting)

Die Knotenwerte von u' aus (3.12) sind durch

$$\mathbf{u}^{S(L-1)} = \mathbf{u}^{S(L-1)} - Q_{S(L-1), \mathcal{T}(L)}^L \cdot \mathbf{u}_{\mathcal{T}(L)}$$

sowie für $l_0 \leq l < L - 1$ durch

$$u'^{(l,s)} = \begin{cases} u^{(l,s)} & \text{falls } 2^{-l}s \notin 2^{-(L-1)}\mathcal{S}(L-1) \\ u'^{(L-1,t)} & \text{falls es ein } t \in \mathcal{S}(L-1) \text{ gibt, mit } 2^{-l}s = 2^{-(L-1)}t \end{cases} \quad (3.16)$$

gegeben.

Beweis: Es ist zu zeigen, daß der Lifting-Korrektur-Term $\Phi^{L-1} \cdot Q_{S(L-1), \mathcal{T}(L)}^L \cdot \mathbf{u}_{\mathcal{T}(L)}$ maximal in den Punkten $2^{-L+1}\mathcal{S}(L-1)$ des Gitters $(0, 2^{-L+1}, \dots, 1)$ ungleich 0 sein kann. Da die Skalierungsfunktionen Φ^{L-1} auf diesem Gitter interpolierend sind, ist also $R(Q^{LT}, \mathcal{T}(L)) \subseteq \mathcal{S}(L-1)$ nachzuweisen. Aus dem Besetzungsmustern von Q^L , siehe Bild 3.4, folgt für $t \in \mathcal{T}(L)$

$$R(Q^{LT}, t) = \mathcal{S}(L-1) \cap \{t, t +^* 1\} .$$

* bei periodischen Randbedingungen $\text{mod}(t+1, 2^{L-1})$. Somit folgt

$$\begin{aligned}
 2R(Q^{LT}, t) &= 2\left(S(L-1) \cap \{t, t+^* 1\}\right) \\
 &\stackrel{(3.7)}{\subseteq} 2R(H^L, 2t+1) \\
 &\stackrel{\text{Korollar 2.1}}{\subseteq} R(\tilde{G}^{LT}, \mathcal{T}(L)) \stackrel{\text{Lemma 3.3}}{=} \mathcal{S}(L)
 \end{aligned}$$

Daraus folgt $R(Q^{LT}, t) \subseteq \mathcal{S}(L-1)$, also die Behauptung. \square

Das Schema aus obigem Lemma kann natürlich rekursiv auf niedrigeren Leveln wiederholt werden. Damit entspricht der Algorithmus für den Basiswechsel weitgehend AWT mit einem zusätzlichen Teilschritt für die Lifting-Korrektur. Die Δ -Korrektur aus AWT realisiert die Kopieroperationen aus (3.16).

```

Funktion   $\bar{\mathbf{u}}_{\mathcal{T}} = \text{AH2L}(\mathbf{u}_{\mathcal{T}})$ 
 $\mathbf{u}^{\mathcal{S}} = \text{AIWT}(\mathbf{u}_{\mathcal{T}}, \mathcal{S})$ 
for  $l = L$  downto  $l_0 + 1$ 
     $\bar{\mathbf{u}}_{\mathcal{T}(l)} = (\tilde{G}_{\mathcal{S}(l), \mathcal{T}(l)}^l)^T \cdot \mathbf{u}^{\mathcal{S}(l)}$ 
     $\mathbf{u}^{\mathcal{S}(l-1)} = \mathbf{u}^{\mathcal{S}(l-1)} - Q_{\mathcal{S}(l-1), \mathcal{T}(l)}^l \cdot \bar{\mathbf{u}}_{\mathcal{T}(l)}$ 
    if  $(l > l_0 + 1)$  then  $\mathbf{u}^{\Delta(l-2)} = (\tilde{H}_{\mathcal{S}(l-1), \Delta(l-2)}^{l-1})^T \cdot \mathbf{u}^{\mathcal{S}(l-1)}$ 
end
 $\bar{\mathbf{u}}_{\mathcal{T}(l_0)} = \mathbf{u}^{\mathcal{S}(l_0)}$ 

```

Die Idee für den inversen Algorithmus "lifting \rightarrow hierarchisch" besteht darin, zunächst wieder die Knotenwerte $\mathbf{u}^{\mathcal{S}(l)}$ zu rekonstruieren, und dann anschließend mittels AWT die Koeffizienten für die hierarchische Basis zu berechnen.

Für die Rekonstruktion gehen wir dabei wie in AIWT vor. Jedoch ist zusätzlich die Lifting-Korrektur entsprechend Lemma 3.7 rückgängig zu machen. Auf Kopieroperationen für die Konsistenzbedingung $u^{(l,s)} = u^{(l+1,2s)}$ kann verzichtet werden, da diese in AWT mit erledigt werden. Der vollständige Algorithmus für den Basiswechsel "lifting \rightarrow hierarchisch" ist dann:

```

Funktion   $\mathbf{u}_{\mathcal{T}} = \text{AL2H}(\bar{\mathbf{u}}_{\mathcal{T}})$ 
 $\mathbf{u}^{\mathcal{S}(l_0)} = \bar{\mathbf{u}}_{\mathcal{T}(l_0)}$ 
for  $l = l_0 + 1$  to  $L$ 
     $\mathbf{u}^{\mathcal{S}(l-1)} = \mathbf{u}^{\mathcal{S}(l-1)} + Q_{\mathcal{S}(l-1), \mathcal{T}(l)}^l \cdot \bar{\mathbf{u}}_{\mathcal{T}(l)}$ 
     $\mathbf{u}^{\mathcal{S}(l)} = G_{\mathcal{S}(l), \mathcal{T}(l)}^l \cdot \bar{\mathbf{u}}_{\mathcal{T}(l)} + H_{\mathcal{S}(l), \mathcal{S}(l-1)}^l \cdot \mathbf{u}^{\mathcal{S}(l-1)}$ 
end
 $\mathbf{u}_{\mathcal{T}} = \text{AWT}(\mathbf{u}^{\mathcal{S}}, \mathcal{S}, \Delta)$ 

```

Zusammen mit diesem Algorithmus sind nun alle Operationen vorgestellt worden, die im weiteren Sinne mit Transformationen der Darstellung von Funktionen zu tun haben. In den folgenden Abschnitten werden Diskretisierungstechniken behandelt.

3.4 Petrov-Galerkin-Diskretisierung

Eine erste einfache Technik für adaptive Multiskalendiskretisierungen stellt die Kollokationsmethode dar, die in diesem Abschnitt beschrieben wird. Wir betrachten das Modellproblem

$$\partial_{xx} u = f \quad \text{in } [0, 1] \quad (3.17)$$

Wie üblich wird die gesuchte Lösung als $u = \Psi_{\mathcal{T}} \cdot \mathbf{u}_{\mathcal{T}}$ dargestellt. Für die Diskretisierung wird nun ein Petrov-Galerkin Verfahren benutzt, wobei die Testfunktionen die dualen Multiskalenfunktionen sind:

$$\langle \tilde{\psi}_{(l,t)}, \partial_{xx} \Psi_{\mathcal{T}} \cdot \mathbf{u}_{\mathcal{T}} \rangle = \langle \tilde{\psi}_{(l,t)}, f \rangle \quad \forall (l,t) \in \mathcal{T}. \quad (3.18)$$

Auf diese Weise erhält man ein lineares Gleichungssystem mit dem die Unbekannten $\mathbf{u}_{\mathcal{T}}$ bestimmt werden können. Analog zu (3.18) werden allgemeinere Differentialoperatoren diskretisiert. Für die erste Ableitung erhält man z.B. $\langle \tilde{\psi}_{(l,t)}, \partial_x \Psi_{\mathcal{T}} \cdot \mathbf{u}_{\mathcal{T}} \rangle$.

Die Wahl der dualen Multiskalenfunktionen als Testfunktionen ist unter anderem durch unsere Beobachtung zur Auswertung nichtlinearer Funktionale in Abschnitt 3.2 motiviert (Bemerkung 3.2). Nachfolgend soll das obige Petrov-Galerkin Schema kurz kommentiert und diskutiert werden:

Bemerkung 3.6 (Glattheitsforderungen) Die dualen Multiskalenfunktionen sind Linearkombinationen von Dirac-Funktionalen. Aus diesem Grund müssen die Multiskalenfunktionen $\Psi_{\mathcal{T}}$ hinreichend glatt sein. Für das Modell-Problem ergibt sich beispielsweise die Forderung $\Psi_{\mathcal{T}} \in C^2$. Entsprechend Tabelle 2.1 sind somit Interpolets der Ordnung $N \geq 6$ notwendig.

Bemerkung 3.7 (Kollokationsmethode) Wegen der Kegelbedingung, siehe Theorem 3.1, können die Koeffizienten $\langle \tilde{\psi}_{(l,t)}, \partial_{xx} u \rangle$ bzw. $\langle \tilde{\psi}_{(l,t)}, f \rangle$ eindeutig in die Knotenwerte $\partial_{xx} u(\Omega_{\mathcal{T}})$ bzw. $f(\Omega_{\mathcal{T}})$ umgerechnet werden, so daß (3.18) äquivalent zu “(3.17) gilt punktweise in $\Omega_{\mathcal{T}}$ “ ist. Deshalb handelt es sich bei obigem Petrov-Galerkin Schema um eine Kollokationsmethode.

Bemerkung 3.8 (Finite-Differenzen) Bezeichne mit T die Matrix der inversen Multiskalentransformation auf Level L . D.h. $\Phi^L \cdot T = (\Psi_{l_0}, \dots, \Psi_L)$. Für den speziellen Fall einer nichtadaptiven Basis, d.h. $\Psi_{\mathcal{T}} = (\Psi_{l_0}, \dots, \Psi_L)$, hat dann die Koeffizientenmatrix in (3.18) die Gestalt

$$T^{-1} \langle \tilde{\Phi}^L, \partial_{xx} \Phi^L \rangle T.$$

Die innere Matrix $\langle \tilde{\Phi}^L, \partial_{xx} \Phi^L \rangle$ kann nun aber als die Darstellungsmatrix eines Finite-Differenzen-Schemas interpretiert werden. D.h. das Anwenden des entsprechenden Operators besteht somit aus den drei Schritten:

1. IWT zur Berechnung von Knotenwerten
2. Anwenden des speziellen Finite-Differenzen-Schemas
3. WT

Es stellt sich die Frage nach der Konsistenzordnung und den Kosten für dieses spezielle Finite-Differenzen-Schema. Bekanntlich ist bei der Kollokationsdiskretisierung mit Ansatzfunktionen, die Polynome $(N - 1)$ ten Grades aufspannen, die Konsistenzordnung für den Operator ∂_{xx} gleich $N - 2$ und für ∂_x gleich $N - 1$. Für den Differenzenstern ergibt sich mit (2.17) und $N = 2n$ bei Interpolets eine Breite von $2N - 1$. D.h. für $N = 6$ erhält man ein Verfahren 4. Ordnung mit einem Stern der 11 nichtverschwindende Einträge hat. Bei den üblichen Finite-Differenzen-Schemen 4. Ordnung sind jedoch nur 5 nichtverschwindende Einträge notwendig. Im Vergleich dazu wird die Auswertung des Operators für die Kollokationsmethode also relativ teuer sein.

Der eigentliche Nachteil der Kollokationsmethode sind aber nicht die Kosten oder die Glattheitsforderungen an die Multiskalenfunktionen sondern die geringe Flexibilität der Diskretisierung. Für hyperbolische Differentialgleichungen ist z.B. eine upwindartige Diskretisierung

des konvektiven Terms notwendig, um unphysikalische Oszillationen zu vermeiden. Wegen der Symmetrie der Funktionen ϕ führt der hier vorgestellte Petrov-Galerkin-Ansatz aber zu einem zentralen Finiten-Differenzen-Schema, das nicht stabil ist.

Ein weiteres Beispiel sind div-grad-Operatoren, die z.B. bei einer Diskretisierung von Sattelpunktproblemen auftreten können. Ein einfaches Hintereinanderanwenden von Petrov-Galerkin-diskretisierten div- bzw. grad-Operatoren würde auch hier zu Lösungen mit unphysikalischen Oszillationen führen. In Kapitel 10 findet sich eine ausführlichere Diskussion dazu. In genau solchen Fällen erweist sich die allgemeine Finite-Differenzen-Technik, die im nächsten Abschnitt beschrieben wird, als deutlich flexibler.

Trotzdem sollen hier der Vollständigkeit halber die entsprechenden Algorithmen für die schnelle Auswertung der Petrov-Galerkin Operatoren angegeben werden. Für das Modell-Problem sind beispielsweise die Terme $\langle \tilde{\psi}_{(l,t)}, \partial_{xx} \sum_k \Psi_{\mathcal{T}(k)} \cdot \mathbf{u}_{\mathcal{T}(k)} \rangle$, $l_0 \leq l \leq L$ gesucht. Das Hauptproblem dabei sind die für Multiskalensbasen typischen, unangenehmen Kopplungen $\langle \tilde{\Psi}_{\mathcal{T}(l)}, \partial_{xx} \sum_k \Psi_{\mathcal{T}(k)} \rangle$ für $l \neq k$, welche zu einer Fingerstruktur der Operator-Matrix führen, siehe Bild 3.6. Für andere Multiskalensbasen oder Operatoren ergibt sich ein ähnli-

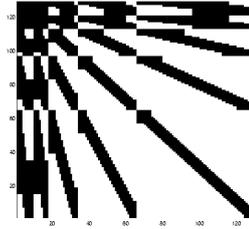


Abbildung 3.6: *Fingerstruktur eines Operators in Multiskalendarstellung $\langle \tilde{\Psi}_{\mathcal{T}}, \partial_{xx} \Psi_{\mathcal{T}} \rangle$.*

ches Besetzungsmuster. Die wesentliche Idee ist nun, diese Kopplungen zu umgehen. Für ein beliebiges l zwischen l_0 und L betrachten wir dazu die Zerlegung $u = a_{l-1} + b_l$, wobei

$$a_{l-1} := \sum_{k \leq l-1} \Psi_{\mathcal{T}(k)} \cdot \mathbf{u}_{\mathcal{T}(k)} \quad \text{und} \quad b_l := \sum_{k \geq l} \Psi_{\mathcal{T}(k)} \cdot \mathbf{u}_{\mathcal{T}(k)} ,$$

so daß

$$\langle \tilde{\Psi}_{\mathcal{T}(l)}, \partial_{xx} u \rangle = \langle \tilde{\Psi}_{\mathcal{T}(l)}, \partial_{xx} a_{l-1} \rangle + \langle \tilde{\Psi}_{\mathcal{T}(l)}, \partial_{xx} b_l \rangle . \quad (3.19)$$

Wir betrachten nun die Terme a_{l-1} . Ähnlich wie bei der Multiskalentransformation wird anstatt des kompletten Terms a_{l-1} nur eine unvollständige Erzeugendensystemdarstellung

$$\Phi^{\mathcal{S}^a(l-1)} \cdot \mathbf{a}^{\mathcal{S}^a(l-1)}$$

berechnet bzw. benötigt. Dabei muß $\mathcal{S}^a(l-1)$ natürlich groß genug sein, so daß die für (3.19) notwendigen Werte berechnet werden können

$$\langle \tilde{\Psi}_{\mathcal{T}(l)}, \partial_{xx} a_{l-1} \rangle \stackrel{!}{=} \langle \tilde{\Psi}_{\mathcal{T}(l)}, \partial_{xx} \Phi^{\mathcal{S}^a(l-1)} \cdot \mathbf{a}^{\mathcal{S}^a(l-1)} \rangle . \quad (3.20)$$

Definiert man $O^{al} := \langle \tilde{\Psi}_l, \partial_{xx} \Phi^{l-1} \rangle$, dann resultiert daraus die Bedingung

$$R(O^{al}, \mathcal{T}(l)) \subseteq \mathcal{S}^a(l-1) .$$

Andererseits muß $\mathcal{S}^a(l-1)$ auch die Bedingung (3.4) erfüllen, damit die Koeffizienten $\mathbf{a}^{\mathcal{S}^a(l-1)}$ einfach mittels $\mathbf{a}^{\mathcal{S}^a} = \text{AIWT}(\mathbf{u}_{\mathcal{T}}, \mathcal{S}^a)$ berechnet werden können. Der vollständige Algorithmus zur Bestimmung von $\mathcal{S}^a(l)$ ist somit:

```

Funktion   $\mathcal{S}^a = \text{IndexMengeAPGa}(\mathcal{T})$ 
 $\mathcal{S}^a(L) = \emptyset$ 
for  $l = L$  downto  $l_0 + 1$ 
     $\mathcal{S}^a(l-1) = R(O^{al}, \mathcal{T}(l)) \cup R(H^l, \mathcal{S}^a(l))$ 
end

```

Für die Auswertung von $R(O^{al}, \mathcal{T}(l))$ muß zumindest das Besetzungsmuster von O^{al} bekannt sein. Da die inneren Skalierungsfunktionen Dilate und Translate von ϕ sind, ist O^{al} im wesentlichen eine Bandmatrix, deren Bandbreite a priori aus der Größe des Trägers von ϕ berechnet werden kann. Für periodische Randbedingungen ist O^{al} zirkulant; für alle anderen Randbedingungen hat O^{al} in der linken, oberen und der rechten, unteren Ecke nichtverschwindende Blöcke, deren Größe ebenfalls a priori aus dem Träger von ϕ berechnet werden können. Die eigentlichen Koeffizienten von O^{al} können wie in [DM93, MP98] bestimmt werden. Gespeichert werden brauchen dabei nur die Rand-Blöcke und eine Zeile von Koeffizienten für nicht randangepaßte Skalierungsfunktionen. Die eigentlich für (3.19) gesuchten Werte $\mathbf{a}_{\mathcal{T}(l)} := \langle \tilde{\Psi}_{\mathcal{T}(l)}, \partial_{xx} a_{l-1} \rangle$ erhält man mit folgendem Algorithmus:

```

Funktion   $\mathbf{a}_{\mathcal{T}} = \text{APGa}(\mathbf{u}_{\mathcal{T}})$ 
 $\mathcal{S}^a = \text{IndexMengeAPGa}(\mathcal{T})$ 
 $\mathbf{a}^{\mathcal{S}^a} = \text{AIWT}(\mathbf{u}_{\mathcal{T}}, \mathcal{S}^a)$ 
for  $l = l_0 + 1$  to  $L$ 
     $\mathbf{a}_{\mathcal{T}(l)} = O_{\mathcal{T}(l), \mathcal{S}^a(l-1)}^{al} \cdot \mathbf{a}^{\mathcal{S}^a(l-1)}$ 
end
 $\mathbf{a}_{\mathcal{T}(l_0)} = 0$ 

```

Ähnliche Tricks wie für a_l sind auch für b_l möglich. Mit kleiner werdendem Level l wächst sowohl die Anzahl der Terme in b_l als auch der Träger von b_l . Andererseits wird aber eine Darstellung von $\partial_{xx} b_l$ auf immer größeren Leveln benötigt, was ausgenutzt werden kann. Die Idee ist hier, die gesuchten Werte mittels

$$\langle \tilde{\Psi}_{\mathcal{T}(l)}, \partial_{xx} b_l \rangle = (\tilde{G}_{\mathcal{S}^b(l), \mathcal{T}(l)}^l)^T \cdot \mathbf{c}^{\mathcal{S}^b(l)} \quad (3.21)$$

aus Hilfskoeffizienten $\mathbf{c}^{\mathcal{S}^b(l)} := \langle \tilde{\Phi}^{\mathcal{S}^b(l)}, \partial_{xx} b_l \rangle$ zu berechnen. Wegen der Definition von b_l erfüllen diese Werte die Rekursionsformel

$$\begin{aligned} \mathbf{c}^{\mathcal{S}^b(l)} &= \langle \tilde{\Phi}^{\mathcal{S}^b(l)}, \partial_{xx} b_{l+1} \rangle + \langle \tilde{\Phi}^{\mathcal{S}^b(l)}, \partial_{xx} \Psi_{\mathcal{T}(l)} \rangle \cdot \mathbf{u}_{\mathcal{T}(l)} \\ &= (\tilde{H}_{\mathcal{S}^b(l+1), \mathcal{S}^b(l)}^{l+1})^T \cdot \mathbf{c}^{\mathcal{S}^b(l+1)} + O_{\mathcal{S}^b(l), \mathcal{T}(l)}^{bl} \cdot \mathbf{u}_{\mathcal{T}(l)} \quad \text{wobei } O^{bl} := \langle \tilde{\Phi}^l, \partial_{xx} \Psi_l \rangle. \end{aligned}$$

Daraus resultieren zwei Forderungen an die noch nicht weiter spezifizierte Indexmenge $\mathcal{S}^b(l)$. Einerseits muß $\mathcal{S}^b(l)$ alle Indizes s enthalten, für die $O_{s, \mathcal{T}(l)}^{bl} \cdot \mathbf{u}_{\mathcal{T}(l)}$ ungleich Null sein kann, d.h.

$$R(O^{blT}, \mathcal{T}(l)) \subseteq \mathcal{S}^b(l). \quad (3.22)$$

Weiterhin muß $\mathcal{S}^b(l)$ alle Indizes s enthalten, für die $(\tilde{H}_{\mathcal{S}^b(l+1), s}^{l+1})^T \cdot \mathbf{c}^{\mathcal{S}^b(l+1)}$ ungleich null sein kann, d.h.

$$R(\tilde{H}^{l+1}, \mathcal{S}^b(l+1)) \subset \mathcal{S}^b(l). \quad (3.23)$$

Daraus ergibt sich die Rekursionsformel $\mathcal{S}^b(l) = R(O^{blT}, \mathcal{T}(l)) \cup R(\tilde{H}^{l+1}, \mathcal{S}^b(l+1))$, die zu folgendem Algorithmus für \mathcal{S}^b führt.

Funktion $\mathcal{S}^b = \text{IndexMengeAPGb}(\mathcal{T})$
 $\mathcal{S}^b(L) = R(O^{bLT}, \mathcal{T}(L))$
for $l = L - 1$ **downto** l_0
 $\mathcal{S}^b(l) = R(O^{blT}, \mathcal{T}(l)) \cup R(\tilde{H}^{l+1}, \mathcal{S}^b(l+1))$
end

Die Koeffizienten $\mathbf{b}_{\mathcal{T}(l)} := \langle \tilde{\Psi}_{\mathcal{T}(l)}, \partial_{xx} b_l \rangle$ für (3.19) können dann folgendermaßen berechnet werden:

Funktion $\mathbf{b}_{\mathcal{T}} = \text{APGb}(\mathbf{u}_{\mathcal{T}})$
 $\mathbf{c}^{\mathcal{S}^b(L)} = O_{\mathcal{S}^b(L), \mathcal{T}(L)}^{bL} \cdot \mathbf{u}_{\mathcal{T}(L)}$
 $\mathbf{b}_{\mathcal{T}(L)} = (\tilde{G}_{\mathcal{S}^b(L), \mathcal{T}(L)}^L)^T \cdot \mathbf{c}^{\mathcal{S}^b(L)}$
for $l = L - 1$ **downto** l_0
 $\mathbf{c}^{\mathcal{S}^b(l)} = (\tilde{H}_{\mathcal{S}^b(l+1), \mathcal{S}^b(l)}^{l+1})^T \cdot \mathbf{c}^{\mathcal{S}^b(l+1)} + O_{\mathcal{S}^b(l), \mathcal{T}(l)}^{bl} \cdot \mathbf{u}_{\mathcal{T}(l)}$
 $\mathbf{b}_{\mathcal{T}(l)} = (\tilde{G}_{\mathcal{S}^b(l), \mathcal{T}(l)}^l)^T \cdot \mathbf{c}^{\mathcal{S}^b(l)}$
end

Schließlich hat man für das Endergebnis

Funktion $\mathbf{v}_{\mathcal{T}} = \text{APG}(\mathbf{u}_{\mathcal{T}})$
 $\mathbf{v}_{\mathcal{T}} = \text{APGa}(\mathbf{u}_{\mathcal{T}}) + \text{APGb}(\mathbf{u}_{\mathcal{T}})$

Bemerkung 3.9 (Nicht-Standardform) Der hier vorgestellte Algorithmus ist Algorithmus, die auf der sogenannten Nicht-Standardform für Multiskalendarstellungen von Operatoren aufbauen, sehr ähnlich [BK97]. Dabei wird, ähnlich wie hier, ausschließlich mit Zweiskalen-Matrizen, wie $\langle \tilde{\Psi}_l, \partial_{xx} \Psi_l \rangle$, $\langle \tilde{\Phi}^{l-1}, \partial_{xx} \Psi_l \rangle$ oder $\langle \tilde{\Psi}_l, \partial_{xx} \Phi^{l-1} \rangle$, gearbeitet. Der technische Aufwand ist im Großen und Ganzen genauso hoch wie bei den hier vorgestellten Algorithmen.

3.5 Randbedingungen beim Petrov-Galerkin–Verfahren

Noch offen ist die Frage, wie beim Petrov-Galerkin–Verfahren Randbedingungen diskretisiert werden können. Eine Möglichkeit ist hier, die Differentialgleichungen so umzuformulieren, daß man homogene Dirichlet-, Neumann- oder periodische Randbedingungen erhält. Sei beispielsweise das ursprüngliche Modell-Problem durch

$$\partial_{xx} u = f \quad \text{in }]0, 1[\quad \text{sowie} \quad u(0) = g_0 \quad \text{und} \quad \partial_x u(1) = g_1$$

gegeben. Falls man eine glatte Funktion w kennt, die die Randbedingungen annimmt, d.h. $w(0) = g_0$ und $\partial_x w(1) = g_1$, dann kann das Modellproblem in

$$\partial_{xx} \hat{u} = (f - \partial_{xx} w) \quad \text{in }]0, 1[\quad \text{sowie} \quad \hat{u}(0) = 0 \quad \text{und} \quad \partial_x \hat{u}(1) = 0 \quad (3.24)$$

überführt werden. Die Lösung des ursprünglichen Problems ist dann $u = \hat{u} + w$. Für das homogene Randwertproblem können wir als Ansatzfunktionen die in Abschnitt 2.4 konstruierten Interpolets mit den entsprechenden Randbedingungen in 0 bzw. 1 benutzen und als

Testfunktionen die entsprechenden dualen Multiskalenfunktionen. Die Operatorauswertung kann dann mit Algorithmus APG realisiert werden. Bei der Herleitung von APG hatten die Randbedingungen ja schließlich keine Rolle gespielt.

Zur Konstruktion von w : Wir nehmen an, daß $\Psi_{\mathcal{T}}$ die Multiskalenfunktionen mit den homogenen Randwerten sind und ${}^+\Psi_{\mathcal{T}^+}$ die Multiskalenfunktionen auf $[0, 1]$ ohne Randbedingungen. Dabei ist $\mathcal{T}^+ := \mathcal{T} \cup (l_0, 0) \cup (l_0, 2^{l_0})$ die leicht vergrößerte adaptive Indexmenge. w kann nun als

$$w = g_0 \cdot {}^+\psi_{(l_0, 0)} + \frac{g_1}{\partial_x {}^+\psi_{(l_0, 2^{l_0})}(1)} \cdot {}^+\psi_{(l_0, 2^{l_0})} =: {}^+\Psi_{\mathcal{T}^+} \cdot \mathbf{w}_{\mathcal{T}^+}$$

definiert werden. Da der Träger von ${}^+\psi_{(l_0, 0)}$ in $[0, 1[$ liegt und der Träger von ${}^+\psi_{(l_0, 2^{l_0})}$ in $]0, 1]$, folgt, daß w genau die gewünschten Randbedingungen erfüllt.

Für die rechte Seite des homogenen Randwert-Problems (3.24) ist $\langle \tilde{\Psi}_{\mathcal{T}}, \partial_{xx} {}^+\Psi_{\mathcal{T}^+} \rangle \cdot \mathbf{w}_{\mathcal{T}^+}$ auszuwerten. Dazu kann auch wieder Algorithmus APG benutzt werden: Bei der Herleitung von APG hatten wir der Einfachheit halber angenommen, daß die Indexmengen und die Randbedingungen für die Ansatz- und Testfunktionen identisch sind. Die ist jedoch überhaupt nicht notwendig. Der Term $\langle \tilde{\Psi}_{\mathcal{T}}, \partial_{xx} {}^+\Psi_{\mathcal{T}^+} \rangle \cdot \mathbf{w}_{\mathcal{T}^+}$ kann mit APG ausgewertet werden, falls dort die folgenden Modifikationen vorgenommen werden:

1. O^{al} ist als $\langle \tilde{\Psi}^l, \partial_{xx} {}^+\Phi^{l-1} \rangle$ und O^{bl} als $\langle \tilde{\Phi}^l, \partial_{xx} {}^+\Psi_l \rangle$ zu setzen.
2. $\mathbf{u}_{\mathcal{T}}$ ist durch $\mathbf{w}_{\mathcal{T}^+}$ zu ersetzen. Insbesondere ist in AIWT (für APGa) die inverse Transformation bezüglich Multiskalenfunktionen ohne Randbedingungen auszuführen.

Für die Berechnung der eigentlichen Lösung $u = \hat{u} + w$ ist nun noch ein weiteres Problem zu lösen. w ist in der Basis ${}^+\Psi_{\mathcal{T}^+}$ gegeben und \hat{u} in der Basis $\Psi_{\mathcal{T}}$. Da diese beiden Basen leicht unterschiedlich sind, sind die Multiskalenkoeffizienten von u nicht einfach die Summe der Multiskalenkoeffizienten von \hat{u} und w . Die Idee ist nun, \hat{u} ebenfalls mittels ${}^+\Psi_{\mathcal{T}^+}$ darzustellen. Dies ist jedoch im allgemeinen nur möglich, falls $\text{span } \Psi_{\mathcal{T}} \subset \text{span } {}^+\Psi_{\mathcal{T}^+}$. Diese Bedingung soll nun näher untersucht werden.

Lemma 3.8 (Vollständigkeitsbedingung) Falls \mathcal{T} die Kegel-Bedingung erfüllt und weiterhin für $l > l_0$ auch noch gilt:

1. Gibt es $t \in \mathcal{T}(l)$, mit $t \leq \frac{N}{2} - 1$, dann muß auch $\{0, \dots, \frac{N}{2} - 1\} \subset \mathcal{T}(l)$ sein.
2. Gibt es $t \in \mathcal{T}(l)$, mit $2^{l-1} - \frac{N}{2} \leq t$, dann muß auch $\{2^{l-1} - \frac{N}{2}, \dots, 2^{l-1} - 1\} \subset \mathcal{T}(l)$ sein.

Dann gilt $\text{span } \Psi_{\mathcal{T}} \subset \text{span } {}^+\Psi_{\mathcal{T}^+}$.

Die Indizes $t < N/2$ und $2^{l-1} - N/2 \leq t$ entsprechen genau den randangepaßten Interpolets auf Level l . Das obige Lemma sagt also aus, daß wenn am linken oder rechten Rand eines von diesen Interpolets in $\Psi_{\mathcal{T}}$ ist, dann müssen alle randangepaßten Interpolets auch in $\Psi_{\mathcal{T}}$ sein. Der Grund für diese Bedingung wird im Beweis für Lemma 3.8 klar, der in Anhang A.8 zu finden ist.

Die Idee für den Algorithmus für den Basiswechsel " $\Psi_{\mathcal{T}} \rightarrow {}^+\Psi_{\mathcal{T}^+}$ " ist nun die folgende: ${}^+\Psi_{\mathcal{T}^+} \cdot \hat{\mathbf{u}}_{\mathcal{T}^+}$ ist der eindeutige Interpolant von \hat{u} auf dem Gitter $\Omega_{\mathcal{T}^+} = \Omega_{\mathcal{T}} \cup \{0, 1\}$. Mittels AWT können wir also $\hat{\mathbf{u}}_{\mathcal{T}^+}$ aus den Knotenwerten $\hat{u}(\Omega_{\mathcal{T}^+})$ berechnen. Mittels AIWT können wiederum die Knotenwerte $\hat{u}(\Omega_{\mathcal{T}})$, genauer $\hat{\mathbf{u}}^{S(l)}$, ausgerechnet werden. Die noch zusätzlich notwendigen Knotenwerte in 0 und 1 ergeben sich daraus nach folgendem Schema: Sei L_0

das maximale Level in Punkt 0, d.h. es gibt kein Interpolet mit Level $> L_0$ in der adaptiven Basis, das nicht in 0 verschwindet. Dann gilt

$$\hat{u}(0) = \sum_{s=1}^{N-1} \hat{u}^{(L_0,s)} \phi^{(L_0,s)}(0) .$$

Der Grund dafür ist, daß in $\hat{\mathbf{u}}^{S(L_0)}$ alle Interpolets von Level $\leq L_0$ eingehen. Andererseits verschwinden alle Interpolets von höheren Leveln in 0. Falls die Multiskalenfunktionen $\Psi_{\mathcal{T}}$ homogene Dirichlet-Randbedingungen haben, ist natürlich trivialerweise $\hat{u}(0) = 0$. Der komplette Algorithmus für den Basiswechsel “ $\Psi_{\mathcal{T}} \rightarrow {}^+\Psi_{\mathcal{T}^+}$ ” ist somit

```

Funktion  $\mathbf{u}_{\mathcal{T}^+} = \mathbf{AUP}(\mathbf{u}_{\mathcal{T}})$ 
 $\mathbf{u}^{\mathcal{S}} = \mathbf{AIWT}(\mathbf{u}_{\mathcal{T}})$  (bezüglich Multiskalenfunktionen mit Randbedingungen)
for  $l = l_0$  to  $L$ 
   $\mathcal{S}^+(l) = \mathcal{S}(l)$ 
  if  $1 \in \mathcal{S}(l)$  then
     $\mathcal{S}^+(l) = \mathcal{S}(l) \cup 0$  und  $u^{(l,0)} = \sum_{s=1}^{N-1} \hat{u}^{(l,s)} \phi^{(l,s)}(0)$ 
  endif
  if  $2^l - 1 \in \mathcal{S}(l)$  then
     $\mathcal{S}^+(l) = \mathcal{S}^+(l) \cup 2^l$  und  $u^{(l,2^l)} = \sum_{s=1}^{N-1} \hat{u}^{(l,2^l-s)} \phi^{(l,2^l-s)}(1)$ 
  endif
end
 $\mathbf{u}_{\mathcal{T}^+} = \mathbf{AWT}(\mathbf{u}^{\mathcal{S}^+}, \mathcal{S}^+, \Delta^+)$  (bezüglich Multiskalenfunktionen ohne Randbedingungen)

```

Der umgekehrte Algorithmus, mit dem die Projektion von $\text{span} {}^+\Psi_{\mathcal{T}^+}$ auf $\text{span} \Psi_{\mathcal{T}}$ realisiert werden kann, ist noch einfacher. Mit $\mathbf{AIWT}(\mathbf{u}_{\mathcal{T}^+})$ können die Knotenwerte auf $\Omega_{\mathcal{T}^+} \supset \Omega_{\mathcal{T}}$ berechnet werden und mit $\mathbf{AIWT}(\mathbf{u}^{\mathcal{S}})$ die gesuchten Koeffizienten $\mathbf{u}_{\mathcal{T}}$.

```

Funktion  $\mathbf{u}_{\mathcal{T}} = \mathbf{ADOWN}(\mathbf{u}_{\mathcal{T}^+})$ 
 $\mathbf{u}^{\mathcal{S}^+} = \mathbf{AIWT}(\mathbf{u}_{\mathcal{T}^+})$  (bezüglich Multiskalenfunktionen ohne Randbedingungen)
 $\mathbf{u}_{\mathcal{T}} = \mathbf{AWT}(\mathbf{u}^{\mathcal{S}^+}, \mathcal{S}, \Delta)$  (bezüglich Multiskalenfunktionen mit Randbedingungen)

```

Natürlich gilt hier nicht $\text{span} {}^+\Psi_{\mathcal{T}^+} \subset \text{span} \Psi_{\mathcal{T}}$, so daß die DOWN-Operation in gewissem Sinne mit Informationsverlust verbunden ist.

3.6 Finite-Differenzen

Neben der Kollokationsmethode stellt die Finite-Differenzen-Technik einen zweiten Weg zur Diskretisierung von Differentialoperatoren auf adaptiven Gittern – später auch adaptiven dünnen Gittern – dar. Die wesentlichen Vorteile gegenüber der Kollokationsmethode sind höhere Effizienz, keine Glattheitsforderungen an ψ und die Möglichkeit spezieller stabilisierter Diskretisierungen für z.B. Konvektions- oder Sattelpunktprobleme.

Finite Differenzen werden schon seit längerem als Diskretisierung auf adaptiven strukturierten Gittern benutzt. Zu nennen sind hier die Arbeiten [Bra72, BC89, HB97]. Bei diesen Verfahren wird das Gebiet im wesentlichen mit disjunkten, rechteckigen Teilgebieten überdeckt, in denen jeweils eine bestimmte Verfeinerung gewählt wird. Adaptive Multiskalenverfahren mit finiten Differenzen zur Auswertung von Operatoren wurden unter anderem von [Har95, CKMP00, Hol99] sowie [Sch98a, Gri98] vorgestellt. Während in den ersten Arbeiten

auf isotrope multivariate Multiskalenfunktionen aufgesetzt wird, sind in [Sch98a, Gri98] anisotrope Tensorprodukt-Multiskalenfunktionen benutzt worden. Wie bei der Kollokationsmethode können dabei die Algorithmen für den multivariaten Fall vollständig auf Algorithmen für den eindimensionalen Fall zurückgeführt werden. Wir werden deshalb in diesem Abschnitt entsprechende Algorithmen für den eindimensionalen Fall angeben.

Eine erste Idee für einen solchen Algorithmus ist, zunächst Knotenwerte von u auf dem Gitter $\Omega_{\mathcal{T}}$ auszurechnen, dann auf diese Werte das gewünschte Finite-Differenzen-Schema anzuwenden und das Ergebnis in eine Multiskalendarstellung zu transformieren. Dieser Weg wurde im wesentlichen in [Sch98a] eingeschlagen. Ein Problem dabei ist, daß im adaptiven Fall $\Omega_{\mathcal{T}}$ kein äquidistantes Gitter ist. Speziell bei der hierarchischen Basis (Interpolets mit $N = 2$) können sogar beliebig große Sprünge in der Maschenweite auftreten. Ein Beispiel dafür ist in Bild 3.7 angegeben. Links sind die Indexmenge \mathcal{T} , die die Kegel-Bedingung erfüllt, bzw. die Mengen $\mathcal{S}(l)$ zu sehen und rechts das resultierende adaptive Gitter $\Omega_{\mathcal{T}}$. Die linksseitige Maschenweite zum Punkt $1/2$ ist gleich $1/2$, wogegen die rechtsseitige Maschenweite gleich $1/16$ ist.

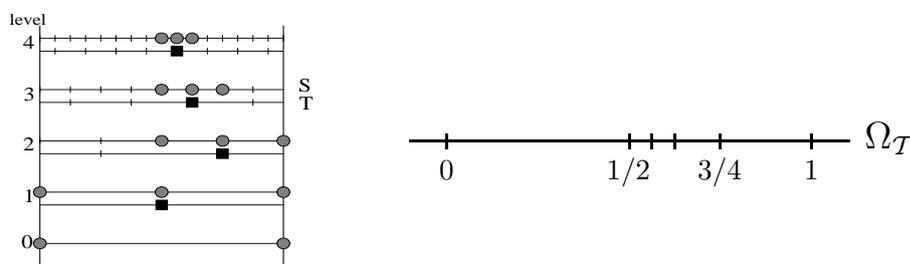


Abbildung 3.7: Indexmengen \mathcal{T} und $\mathcal{S}(l)$ sowie $\Omega_{\mathcal{T}}$.

Bei Interpolets höherer Ordnung erzwingt die Kegel-Bedingung jedoch automatisch glattere Gitter, bei denen die Maschenweite nicht allzusehr springen kann.

Lemma 3.9 *Sei \mathcal{T} eine Indexmenge, die für Interpolets der Ordnung $N \geq 4$ die Kegel-Bedingung erfüllt. Dann unterscheiden sich die links- bzw. rechtsseitige Maschenweite für einen Punkt $x \in \Omega_{\mathcal{T}}$ maximal um den Faktor 4.*

Der Beweis ist im Anhang A.9 skizziert. Ein sehr nützlicher Nebeneffekt von Lemma 3.9 ist, daß es für Punkte $x \in \Omega_{\mathcal{T}}$ nur einige bestimmte Konfigurationen von Nachbarn geben kann, auf denen lokale Finite-Differenzen-Sterne aufsetzen können. In der Regel gibt es dabei mehrere mögliche Sterne, die sich auch bei gleicher Konsistenzordnung in der Qualität deutlich voneinander unterscheiden können. So kann der Konsistenzfehler für ein zentrales Schema vierter Ordnung zur Berechnung der zweiten Ableitung auf einem Gitter der Maschenweite h durch $\frac{2}{180}h^4|D^6u|_{\infty}$ abgeschätzt werden. Bei einem komplett rechtsseitigen Stern erhält man jedoch nur die Abschätzung $\frac{137}{180}h^4|D^6u|_{\infty}$. D.h. der Fehler dürfte deutlich größer sein.

Eine mögliche Strategie für adaptive Finite-Differenzen-Verfahren ist nun, a priori für alle bzw. bestimmte mögliche Nachbarschaftskonfigurationen die Differenzen-Sterne aufzustellen, Konsistenzabschätzungen wie oben zu berechnen und die Sterne entsprechend der gefundenen Konstanten zu sortieren. Bei der eigentlichen Auswertung des Operators kann man dann entsprechend dieser Reihenfolge alle Sterne durchgehen und testen, ob in der Umgebung des aktuellen Gitterpunktes die für den Stern notwendigen Nachbarn tatsächlich vorhanden sind. Falls ja, kann die gesuchte Ableitung berechnet werden. Auf diese Weise wird jeweils der Stern mit der bestmöglichen a priori Abschätzung benutzt.

Funktion $\mathbf{v}_{\mathcal{T}} = \mathbf{AFD}(\mathbf{u}_{\mathcal{T}})$
 Berechne mit AIWT($\mathbf{u}_{\mathcal{T}}$) die Knotenwerte $u(\Omega_{\mathcal{T}})$
for $x \in \Omega_{\mathcal{T}}$
 Bestimme den bestmöglichen Finite-Differenzen-Stern
 vorgegebener Konsistenzordnung und berechne damit $v(x)$
end
 Berechne $\mathbf{v}_{\mathcal{T}}$ mittels AWT aus $v(\Omega_{\mathcal{T}})$

Etwas nachteilig an dem obigen Vorgehen ist, daß für jeden Gitterpunkt die doch etwas aufwendigere Nachbarsuche zu durchlaufen ist und außerdem auch mit Differenzen-Sternen für nichtäquidistante Gitter umgegangen werden muß. Manche Diskretisierungen wie ENO/WENO-Schemen können aber gar nicht so einfach auf nichtäquidistante Gitter übertragen werden, wie die Diskussion in z.B. [Shu99] zeigt. Hier bietet sich eine zweite adaptive Finite-Differenzen-Diskretisierung an.

Die Idee ist im wesentlichen, durch Einfügen von zusätzlichen Hilfsgitterpunkten dafür zu sorgen, daß jeweils alle Aufsetzpunkte für den äquidistanten Differenzen-Stern gegeben sind. Die Knotenwerte in den Hilfspunkten werden durch Interpolation berechnet. Dies kann sehr einfach mittels AIWT geschehen. Der zweite Teil der Idee ist, die Differenzen-Sterne unmittelbar auf die mit AIWT berechnete Erzeugendensystemdarstellung von u anzuwenden. Dadurch vereinfacht sich die Suchen von Nachbarn enorm. In groben Zügen hat der Algorithmus somit die Form:

1. Berechne mittels AIWT Knotenwerte von u auf Teilgittern $2^{-l}\mathcal{S}^f(l)$.
2. Anwenden des Finite-Differenzen-Schemas auf diesen (äquidistanten) Teilgittern. Man erhält Werte $\{v^{(l,s)}\}$.
3. AWT für $\{v^{(l,s)}\}$.

In der obigen Beschreibung des Algorithmus sind die Indexmengen $\mathcal{S}^f(l)$ für die adaptiven äquidistanten Teilgitter noch nicht näher spezifiziert worden. Eine notwendige Voraussetzung an \mathcal{S}^f für die Durchführung von AWT ist analog zu Abschnitt 3.1

$$\mathcal{S}^f(l) \supseteq R(\tilde{G}^{lT}, \mathcal{T}(l)) \stackrel{(3.9)}{=} \mathcal{S}(l) .$$

Das allein reicht aber noch nicht, da für den Differenzen-Stern zur Berechnung von z.B. $v^{(l,s)}$, $s \in \mathcal{S}(l)$, auch noch die Werte von Nachbarpunkten benötigt werden. Welche das sind, ergibt sich aus dem Stern bzw. – formal ausgedrückt – aus dem Besetzungsmuster der Darstellungsmatrix O^{fl} des Finite-Differenzen-Schemas auf Level l . Für ein zentrales Schema zweiter Ordnung wäre O^{fl} eine $\mathcal{S}(l) \times \mathcal{S}(l)$ -Bandmatrix, mit Bandbreite 3. In Bild 3.8 ist für ein Beispiel skizziert, welche Indizes s in diesem Fall neben $\mathcal{S}(l)$ (markiert mit \bullet) auch noch in $\mathcal{S}^f(l)$ sein müssen (markiert mit \circ).

Im allgemeinen Fall muß \mathcal{S}^f also zusätzlich

$$R(O^{fl}, \mathcal{S}(l)) \subseteq \mathcal{S}^f(l)$$

erfüllen. Als dritte Bedingung an $\mathcal{S}^f(l)$ erfordert AIWT für die Berechnung der Werte $\mathbf{u}^{\mathcal{S}^f(l)}$:

$$R(H^l, \mathcal{S}^f(l)) \subseteq \mathcal{S}^f(l-1) .$$

Mittels des folgenden Algorithmus werden neben \mathcal{S}^f auch Indexmengen \mathcal{S}^e und Δ^e ermittelt. $\mathcal{S}^e(l)$ ist jeweils die Indexmenge der Punkte, für die das Finite-Differenzen-Schema exakt

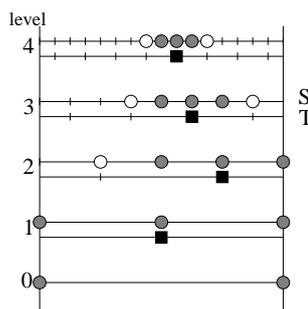


Abbildung 3.8: Indexmengen für Auswertung von Finite-Differenzen-Sternen. Grau: Indexmengen $\mathcal{S}(l)$ auf denen das Finite-Differenzen-Schema angewendet werden muß. Weiß: Die dafür zusätzlich notwendigen Gitterpunkte.

angewandt werden kann, da alle für den Stern notwendigen Werte $u^{(l,s)}$ gegeben sind. $\Delta^e(l)$ bezeichnet die sich aus \mathcal{S}^e ergebende Menge für die Δ -Korrektur in AWT.

```

Funktion  { $\mathcal{S}^f, \mathcal{S}^e, \Delta^e$ }=IndexMengeAFD( $\mathcal{T}$ )
 $\mathcal{S}^e(L) = R(\tilde{G}^{LT}, \mathcal{T}(L))$ 
 $\mathcal{S}^f(L) = R(O^{fL}, \mathcal{S}^e(L))$ 
for  $l = L - 1$  downto  $l_0$ 
     $\mathcal{S}^f(l) = R(O^{fl}, R(\tilde{G}^{lT}, \mathcal{T}(l))) \cup R(H^{l+1}, \mathcal{S}^f(l+1))$ 
     $\mathcal{S}^e(l) = C(O^{fl}, \mathcal{S}^f(l))$ 
     $\Delta^e(l) = \{s : 2s \in \mathcal{S}^e(l+1)\}$ 
end

```

Im adaptiven Finite-Differenzen-Algorithmus kommt der Δ -Korrektur in AWT eine besondere Bedeutung zu: Im allgemeinen gilt auch für $s \in \Delta^e(l)$ nicht $v^{(l,s)} = v^{(l+1,2s)}$. Vielmehr ist $v^{(l,s)}$ auf einem Gitter der Maschenweite 2^{-l} berechnet worden und $v^{(l+1,2s)}$ auf einem Gitter der Maschenweite $2^{-(l+1)}$. Dieser Wert wird also genauer als $v^{(l,s)}$ sein. Der Trick ist nun, daß durch die Δ -Korrektur in AWT der weniger genaue Wert überschrieben wird. Somit stellt die Δ -Korrektur den Mechanismus dar, durch den genauere Werte wann immer möglich, ungenauere Werte ersetzen. Eine Analyse dieses Schemas ist in Kapitel 6 angegeben.

```

Funktion   $\mathbf{v}_{\mathcal{T}} = \text{AFD2}(\mathbf{u}_{\mathcal{T}})$ 
{ $\mathcal{S}^f, \mathcal{S}^e, \Delta^e$ }=IndexMengeAFD( $\mathcal{T}$ ).
 $\mathbf{u}^{\mathcal{S}^f} = \text{AIWT}(\mathbf{u}_{\mathcal{T}}, \mathcal{S}^f)$ 
for  $l = l_0$  to  $L$ 
    berechne  $\mathbf{v}^{\mathcal{S}^e(l)}$  aus  $\mathbf{u}^{\mathcal{S}^f(l)}$  mittels des Finite-Differenzen-Schemas
end
 $\mathbf{v}_{\mathcal{T}} = \text{AWT}(\mathbf{v}^{\mathcal{S}^e}, \mathcal{S}^e, \Delta^e)$ 

```

In Bezug auf die Konsistenzordnung des Verfahrens gibt es eine Besonderheit. Wir nehmen an, daß $\Psi_{\mathcal{T}} \cdot \mathbf{u}_{\mathcal{T}}$ der Interpolant einer Funktion u bezüglich $\Omega_{\mathcal{T}}$ ist. Für das Anwenden des Finite-Differenzen-Schemas waren jedoch Knotenwerte in $2^{-l}\mathcal{S}^f(l)$ notwendig und nicht nur in $2^{-l}\mathcal{S}(l) \subseteq \Omega_{\mathcal{T}}$. Die Knotenwerte in $2^{-l}(\mathcal{S}^f(l) \setminus \mathcal{S}(l))$ werden somit durch Interpolation (Auswertung des Interpolanten $\Psi_{\mathcal{T}} \cdot \mathbf{u}_{\mathcal{T}}$) berechnet. Die Approximationsordnung dieses Interpolationsschemas ist N . Werden nun die interpolierten Knotenwerte für das Finite-Differenzen-Schema verwendet, verliert man natürlich an Konsistenzordnung. Hat das nichta-

daptive Finite-Differenzen-Verfahren für die Approximation der r -ten Ableitung die Konsistenzordnung K , dann hat das adaptive Verfahren in diesem Sinne die Konsistenzordnung $\min(N - r, K)$. Im Beispiel aus Abbildung 3.8 tritt diese Situation in den grauen Knotenwerten auf, die unmittelbar neben einem weißen Knoten liegen.

Für die Diskretisierung von Randbedingungen gibt es beim Finite-Differenzen-Verfahren mehrere Möglichkeiten. Zum einen können die Randbedingungen wie in Abschnitt 3.5 homogen gemacht werden. Mit AFD/AFD2 muß dann die Projektion der Ableitung auf den Ansatzraum berechnet werden. Dazu sind ganz einfach z.B. in $\text{AWT}(\mathbf{v}^{\mathcal{S}^e}, \mathcal{S}^e, \Delta^e)$ die Multiskalenfunktionen mit den homogenen Randbedingungen zu benutzen. Auf diese Weise wird der eigentlich notwendige Aufruf von ADOWN, vergleiche Abschnitt 3.5, gleich miterledigt. Bevor das Finite-Differenzen-Schema angewandt werden kann, müssen außerdem noch wie in AUP die Werte $u^{(l,0)}$ bzw. $u^{(l,2^l)}$ berechnet werden.

Eine andere Möglichkeit ist, für die Berechnung von $v^{(l,s)}$ Finite-Differenzen-Sterne zu benutzen, die die gewünschten Randbedingungen beinhalten, siehe z.B. [Sch98a]. Falls inhomogene Randbedingungen gestellt sind, müssen diese Werte dabei auch berücksichtigt werden. Bei Neumann-Randwerten wird dann die Gültigkeit der Differentialgleichung auch auf dem Rand gefordert.

(Zusammenfassung) In diesem Kapitel sind die adaptiven Algorithmen für die Multiskalentransformationen, die Basiswechsel sowie die Diskretisierung von Differentialoperatoren vorgestellt worden, die in den folgenden Kapiteln für die Konstruktion des Finite-Differenzen-Verfahrens auf adaptiven dünnen Gittern benötigt werden. Es ist gezeigt worden, daß die univariaten Basen zwei Bedingungen (Kegel-B., Lifting-B.) erfüllen müssen, damit die Algorithmen einfach realisierbar sind. Die Algorithmen selbst basieren auf den alternativen Multiskalen- bzw. Erzeugendensystemdarstellungen von Funktionen und können auf diese Weise die Vorzüge beider Betrachtungsweisen kombinieren.

Kapitel 4

Multivariate Multiskalenfunktionen

In diesem Kapitel werden zwei Tensorprodukt-artige Konstruktionen für multivariate Multiskalenfunktionen vorgestellt. Es wird gezeigt, wie die Vorarbeiten aus den Kapiteln zwei und drei, Normäquivalenzen und Algorithmen für die Basistransformationen, auf den mehrdimensionalen Fall verallgemeinert werden können.

In den letzten Jahren haben sich für Anwendungen im wesentlichen zwei Typen von multivariaten Multiskalenbasen durchgesetzt:

- Multiskalenfunktionen auf Dreiecks- bzw. Tetraeder Gittern [Yse86, SS95, Ste96, VW97, Sch98b, FQ99, Ger01].
- Multiskalenfunktionen, die mittels Tensorprodukten aus univariaten Wavelets gebildet werden [LM86, Mal89b, Zen91].

Daneben gibt es noch weitere Ansätze, wie nicht separable Wavelets auf kartesischen Gittern [KV92], die hierarchischen Zerlegungen der eins mit reproduzierendem Kern auf unstrukturierten Punktwolken [LCUC96, LL99] oder die sogenannten Ridgelets [Don99], auf die hier aber nicht weiter eingegangen werden soll. Der große Vorteil von Multiskalenfunktionen auf Dreiecks- bzw. Tetraeder-Triangulierungen ist die hohe geometrische Flexibilität, die Konstruktionen auch auf kompliziert geformten Gebieten erlaubt. Allerdings ist es dabei recht aufwendig, einen hohen Grad an polynomialer Exaktheit zu realisieren.

Im Gegensatz dazu ist es sehr einfach, mittels der in diesem Kapitel vorgestellten Tensorprodukt-Konstruktionen einen beliebig hohen Grad an polynomialer Exaktheit zu erzielen. Eine der Varianten führt dabei zu Basisfunktionen, die in etwa einen isotropen Träger haben, wogegen bei der anderen Variante sowohl isotrope als auch stark anisotrope Basisfunktionen gebildet werden. Beim Vergleich dieser beiden Varianten ist neben algorithmischen Fragen von besonderem Interesse, wie gut sich Funktionen aus einer bestimmten Klasse, z.B. Sobolev- oder Besov-Räume, mit den entsprechenden Basen approximieren lassen. Diese Frage wird in Kapitel 5 ausführlich behandelt. Es zeigt sich dabei, daß in vielen Fällen die anisotrope Konstruktion deutliche Vorteile aufweist. Ein Beispiel, wo dies unmittelbar einsichtig ist, sind (turbulente) Strömungen in der Nähe von festen Körpern, siehe z.B. Bild 4.1, wo die entscheidenden Strukturen langgestreckte hairpin-Wirbel sind. Wegen der besseren Approximationseigenschaften basiert die in dieser Arbeit beschriebene Lösungsmethode für partielle Differentialgleichungen auf der anisotropen Konstruktion. Entsprechend wird diese hier ausführlicher behandelt als die isotrope Konstruktion.

Ein Problem bei allen Tensorprodukt-Konstruktionen besteht darin, daß zunächst nur das Gebiet $[0, 1]^d$, d ist die Raumdimension, diskretisiert werden kann. Komplizierter geformte

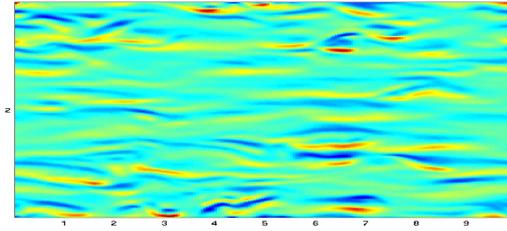


Abbildung 4.1: *Strömung im Kanal* $[0, 10] \times [0, 2] \times [0, 4]$. Farbdarstellung der y -Komponente von $\nabla \times \mathbf{u}$ für die Ebene $[0, 10] \times \{0.001\} \times [0, 4]$. Daten von H.J. Kaltenbach (TU-Berlin).

Gebiete können aber durch Aneinanderkleben von parametrischen Bildern von $[0, 1]^d$ realisiert werden [TW85]. Im Zusammenhang mit Multiskalendiskretisierungen ist diese Technik z.B. in [Hem95, DP96, BD98, Sch98a, CTU99] angewendet worden. Der Vorteil davon ist, daß die Ansatzfunktionen in gewisser Weise entsprechend der Form des Gebietes bzw. entlang des Randes ausgerichtet sind. Da aber andererseits oft auch die Singularitäten/Quasisingularitäten der betrachteten Funktion entlang des Randes ausgerichtet sind, können diese effizient und genau dargestellt werden – insbesondere mit den anisotropen Multiskalenfunktionen. Dieser Vorteil würde verloren gehen, falls z.B. für die Lösung von partiellen Differentialgleichungen Techniken wie die immersed boundary Methode [LeV94, LeV97, FKBS99] benutzt würden, bei denen einfach ein kartesisches Gitter über das Gebiet gelegt wird bzw. einfach die Tensorprodukt-Basisfunktionen benutzt werden. Die Randbedingungen werden darin mittels (quasi-)singulärer Terme auf der rechten Seite diskretisiert. In diesem Fall wären die Basisfunktionen nicht entlang des Gebietsrandes ausgerichtet und man muß zusätzlich die künstlichen Singularitäten auf der rechten Seite berücksichtigen. In dieser Arbeit wird allerdings nicht weiter auf die Frage der Geometrieapproximation eingegangen, dafür sei auf die oben erwähnten Originalarbeiten verwiesen.

4.1 Isotrope Verfeinerung

Es handelt sich hierbei um die klassische Tensorprodukt-Konstruktion für mehrdimensionale Multiskalenfunktionen [LM86, Mal89b], da wie im eindimensionalen Fall linear angeordnete Approximationsräume entstehen. Für $\mathbf{x} \in [0, 1]^d$ und $\mathbf{s} \in S(l) \times \dots \times S(l)$ werden die Skalierungsfunktionen als

$$\phi^{(l,t_1)}(x_1) \cdot \dots \cdot \phi^{(l,t_d)}(x_d) \quad (4.1)$$

definiert. Die Approximationsräume sind somit Tensorprodukträume $V^l \otimes \dots \otimes V^l$, die ineinander geschachtelt sind:

$$V^l \otimes \dots \otimes V^l \subset V^{l+1} \otimes \dots \otimes V^{l+1} .$$

Für die Konstruktion der Komplementräume wird üblicherweise von der Zerlegung $V^l = V^{l-1} \oplus W_l$ ausgegangen. Im zweidimensionalen Fall erhält man dann beispielsweise

$$\begin{aligned} V^l \otimes V^l &= (V^{l-1} \oplus W_l) \otimes (V^{l-1} \oplus W_l) \\ &= V^{l-1} \otimes V^{l-1} \\ &\quad \oplus W_l \otimes V^{l-1} \oplus V^{l-1} \otimes W_l \oplus W_l \otimes W_l . \end{aligned}$$

Der Raum $V^{l-1} \otimes V^{l-1}$ stellt den grobskaligen Approximationsraum dar und die drei anderen Tensorprodukträume bilden zusammen den Komplementraum. Im d -dimensionalen Fall erhält man auf diese Art $2^d - 1$ verschiedene Typen von Multiskalenfunktionen der Form

$$\psi_{(l,t,e)}(\mathbf{x}) := \prod_{i=1}^d \psi_{(l,t_i)}^{e_i}(x_i), \quad \mathbf{e} \in \{0, 1\}^d \setminus \mathbf{0} \quad (4.2)$$

wobei

$$\psi_{(l,t)}^0 := \phi^{(l-1,t)} \quad \text{und} \quad \psi_{(l,t)}^1 := \psi_{(l,t)}.$$

Der Träger dieser Funktionen hat in jeder der Koordinatenrichtungen ungefähr die gleiche Ausdehnung von $O(2^{-l})$, was die Bezeichnung 'isotrope Verfeinerung' motiviert. Die dualen Skalierungs- bzw. Multiskalenfunktionen werden analog zu (4.1) und (4.2) aus den dualen univariaten Pendanten gebildet.

Mit der isotropen Konstruktion ist es besonders einfach, Aussagen zu Normäquivalenzen auf den multivariaten Fall zu übertragen. Der Beweis von Theorem 2.3 beruhte ja lediglich auf der linearen Anordnung der Approximationsräume, Jackson- und Bernstein-Ungleichungen für die primalen und dualen Skalierungsfunktionen und der \mathcal{L}^2 -Stabilität der primalen Multiskalenfunktionen eines Levels mit Level-unabhängigen Stabilitätskonstanten.

Die erste Voraussetzung ist hier natürlich trivialerweise erfüllt. Der Nachweis von Jackson- bzw. Bernstein-Ungleichungen ist üblicherweise auch recht einfach, allerdings ist dabei zu beachten, daß für verallgemeinerte hierarchische Basen, wie die Interpolets, die dualen Skalierungsfunktionen (Dirac-Funktionale) im d -dimensionalen Fall nur in \mathcal{H}^s , für $s < -d/2$, sind, was zu einer recht unangenehmen Einschränkung der Skala von Sobolev-Räumen führt, für die die Normäquivalenzen gelten. Die \mathcal{L}^2 -Stabilität folgt mit der \mathcal{L}^2 -Stabilität der Skalierungsfunktionen eines Levels. Auf einen Beweis soll hier verzichtet werden. Insgesamt erhält man [Ber96, Sch98b]:

Theorem 4.1 (Normäquivalenzen für isotrope hierarchische Basen)

Sei $u = \sum_{(l,t,e)} u_{(l,t,e)} \psi_{(l,t,e)}$, dann gilt für $0 < s < \gamma^*$

$$\|u\|_{\mathcal{H}^s}^2 \lesssim \sum_{(l,t,e)} |u_{(l,t,e)}|^2 2^{2sl} \cdot 2^{-dl}.$$

Die umgekehrte Relation

$$\sum_{(l,t,e)} |u_{(l,t,e)}|^2 2^{2sl} \cdot 2^{-dl} \lesssim \|u\|_{\mathcal{H}^s}^2.$$

gilt für $N > s > d/2$, falls $u \in \mathcal{H}^s$. Die eigentlichen Normäquivalenzen gelten also für $s \in]d/2, \gamma^*[$. Dieser Bereich kann leer sein.

Der in den obigen Abschätzungen auftretende Faktor 2^{-dl} resultiert aus der Normierung $\|\psi_{(l,t,e)}\|_{\mathcal{L}^2}^2 \sim 2^{-dl}$. Der Beweis für das obige Theorem geht natürlich auch für isotrope Tensorprodukte von Lifting-Wavelets durch. In diesem Fall sind die dualen Skalierungsfunktionen aus \mathcal{H}^s mit $0 \leq s < \tilde{\gamma}^*$, so daß dann die Normäquivalenzen für $s \in]-\tilde{\gamma}^*, \gamma^*[$ gelten.

Da die isotropen Tensorprodukt-Multiskalenfunktionen nicht den eigentlichen Schwerpunkt dieser Arbeit bilden, gehen wir nur kurz auf die algorithmischen Aspekte ein. Wegen des Tensorproduktansatzes wird beispielsweise die Multiskalentransformation so realisiert, daß

zunächst ein Schritt (2.11) der univariaten Transformation für alle Zeilen bzgl. der ersten Koordinatenrichtung ausgeführt wird und dann analog bzgl. aller weiteren Koordinatenrichtungen. Die Matrix für diesen einen Schritt der Transformation auf Level l ist somit

$$\begin{pmatrix} \tilde{H}^{lT} \\ \tilde{G}^{lT} \end{pmatrix} \otimes \dots \otimes \begin{pmatrix} \tilde{H}^{lT} \\ \tilde{G}^{lT} \end{pmatrix}. \quad (4.3)$$

Das Symbol \otimes bezeichnet hier das Kronecker-Produkt zweier Matrizen [Zha99]. Da wie im eindimensionalen Fall die Approximationsräume linear angeordnet sind, können die in den vorangegangenen Kapiteln vorgestellten Algorithmen direkt übernommen werden. Natürlich sind dabei die Transformations- bzw. Operatormatrizen durch ihre multivariaten Gegenstücke, wie z.B. (4.3), zu ersetzen.

4.2 Anisotrope Verfeinerung

In diesem Abschnitt beschreiben wir die Multiskalenfunktionen, die die Grundlage der in dieser Arbeit vorgestellten Lösungsverfahren für partielle Differentialgleichungen bilden. Wir zeigen, wie für diese Basisfunktionen die Normabschätzungen aus Kapitel zwei sowie die adaptiven Algorithmen für Basistransformationen aus Kapitel drei auf den multivariaten Fall übertragen werden. Die Normabschätzungen aber auch z.B. die multivariate Form der Kegel-Bedingung spielen eine wichtige Rolle für unsere Untersuchungen zur Approximation von Funktionen im nächsten Kapitel. Auf die eigentlichen Diskretisierungen von Differentialoperatoren gehen wir erst in Kapitel 6 ein.

Gegeben sei eine univariate Multiskalenbasis $\{\psi_{(l,t)}\}$. Dann sind die multivariaten Basisfunktionen als

$$\psi_{(\mathbf{l},\mathbf{t})}(\mathbf{x}) := \prod_{i=1}^d \psi_{(l_i,t_i)}(x_i) \quad (4.4)$$

definiert. Dieses einfache Konstruktionsschema ist z.B. auch von den multivariaten Polynombasen bekannt, die für Spektralverfahren benutzt werden [CHQZ88]. Die Träger der so definierten Multiskalenfunktionen sind d -dimensionale Quader mit Kantenlängen $O(2^{-l_i})$ bzgl. der i -ten Koordinatenrichtung. Demnach sind isotrope, aber auch anisotrope Träger möglich, was die Bezeichnung "anisotrope Verfeinerung" motiviert. Der Levelindex \mathbf{l} ist ein d -dimensionaler Multiindex $\mathbf{l} = (l_1, \dots, l_d)$ und die von den Basisfunktionen eines Levels aufgespannten Räume sind $W_{\mathbf{l}} := W_{l_1} \otimes \dots \otimes W_{l_d}$. Im Gegensatz zur isotropen Verfeinerung gibt es keine kanonische lineare Anordnung der zugrundeliegenden Approximationsräume $V^{\mathbf{l}} := V^{l_1} \otimes \dots \otimes V^{l_d}$. Anstatt dessen kann man $\{V^{\mathbf{l}}\}$ bzw. $\{W_{\mathbf{l}}\}$ zu einem d -dimensionalen Array anordnen, wie dies z.B. in [Osw94, HP97] vorgeschlagen wurde.

Für den Nachweis von Normabschätzungen ist in [GO95c, Kna00] eine Beweistechnik angegeben worden, mittels der die Normabschätzungen aus Kapitel 2 sehr einfach auf die hier betrachteten Tensorprodukt-Multiskalenfunktionen übertragen werden können. Das Vorgehen basiert dabei auf der Darstellung von Sobolev-Räumen $\mathcal{H}^s([0, 1]^d)$, $s \geq 0$, in der Form

$$\mathcal{H}^s([0, 1]^d) = \bigcap_{i=1}^d \underbrace{\mathcal{L}^2([0, 1]) \otimes \dots \otimes \mathcal{L}^2([0, 1])}_{(i-1)\text{-mal}} \otimes \mathcal{H}^s([0, 1]) \otimes \underbrace{\mathcal{L}^2([0, 1]) \otimes \dots \otimes \mathcal{L}^2([0, 1])}_{(d-i)\text{-mal}}.$$

Für die einzelnen einfach zusammengesetzten Tensorprodukträume erhält man Normabschätzungen, falls die univariaten Multiskalenfunktionen $\psi_{(l,t)}$ Normabschätzungen für \mathcal{H}^s

bzw. \mathcal{L}^2 erfüllen. Aus diesem Grund erfüllen die Interpolets *nicht* die Voraussetzungen des nachfolgenden Lemmas.

Lemma 4.1 (Normabschätzung für anisotrope Tensorprodukt-Multiskalenbasen)

Die univariaten Multiskalenfunktionen $\{\psi_{(1,t)}\}$ seien \mathcal{L}^2 -stabil. Die Zahl der verschwindenden Momente sei $M > 0$ und die maximale Sobolev-Regularität sei $\gamma^* > 0$. Dann gilt für $-M < s < \gamma^*$ die einseitige Normabschätzung

$$\left\| \sum_{(1,t)} u_{(1,t)} \psi_{(1,t)} \right\|_{\mathcal{H}^s} \lesssim \sum_{(1,t)} 4^{s|l|_\infty} |u_{(1,t)}|^2 \|\psi_{(1,t)}\|_{\mathcal{L}^2}^2 \quad \forall \{u_{(1,t)}\} \quad (4.5)$$

Beweis: Für $0 \leq s \leq \gamma^*$ ist die Aussage schon in z.B. [Kna00] bewiesen worden. Sei nun $-M < s < 0$. Für die dualen Wavelets folgt aus der Argumentation in [Kna00] (Abschnitt 2.3) auch, daß für $0 \leq -s < M = \tilde{N}$ gilt

$$\left\| \sum_{(1,t)} \tilde{u}_{(1,t)} \tilde{\psi}_{(1,t)} \right\|_{\mathcal{H}^{-s}} \gtrsim \sum_{(1,t)} 4^{-s|l|_\infty} |\tilde{u}_{(1,t)}|^2 \|\tilde{\psi}_{(1,t)}\|_{\mathcal{L}^2}^2 \quad \forall \{\tilde{u}_{(1,t)}\} .$$

Mit einem Dualitätsargument folgt daraus die Abschätzung (4.5). \square

Die komplette Normäquivalenz folgt aus Lemma 4.1 angewendet auf die primalen bzw. die dualen Tensorprodukt-Multiskalenfunktionen.

Theorem 4.2 (Normäquivalenz für anisotrope Tensorprodukt-Multiskalenbasen)

Die univariate Multiskalenbasis erfüllt für $s \in]-\tilde{\gamma}^*, \gamma^*[$ mit $\gamma^*, \tilde{\gamma}^* > 0$ die Normäquivalenz aus Theorem 2.3. Dann erfüllen die multivariaten Basisfunktionen für $s \in]-\tilde{\gamma}^*, \gamma^*[$ die Normäquivalenz

$$\left\| \sum_{(1,t)} u_{(1,t)} \psi_{(1,t)} \right\|_{\mathcal{H}^s}^2 \sim \sum_{(1,t)} 4^{s|l|_\infty} |u_{(1,t)}|^2 2^{-|l|_1} .$$

In den obigen Aussagen wurde jeweils die \mathcal{L}^2 -Stabilität der univariaten Multiskalenfunktionen gefordert, die für die Interpolets nicht gilt. Trotzdem gelten für diese Basen aber zumindest einseitige Normabschätzungen in einem bestimmten Bereich. Wir benötigen für unsere weiteren Untersuchungen aber nur Abschätzungen für die \mathcal{L}^2 -Norm. Aus Theorem 2.3 folgt mittels Abschätzungen der Eigenwerte der univariaten Massematrizen und der Tensorproduktstruktur der Massematrix im multivariaten Fall:

Lemma 4.2 (Einseitige Abschätzung für hierarchische Basen) Die Multiskalenfunktionen seien anisotrope Tensorprodukte von verallgemeinerten hierarchischen Basen, dann gilt

$$\left\| \sum_{(1,t): |l|_\infty \leq L} u_{(1,t)} \psi_{(1,t)} \right\|_{\mathcal{L}^2}^2 \lesssim L^d \sum_{(1,t)} |u_{(1,t)}|^2 2^{-|l|_1} .$$

Es ist zu vermuten, daß die einseitigen Abschätzungen aus Lemma 4.1 in analoger Weise auch auf $s = -M$ erweitert werden können, falls auf der rechten Seite von (4.5) ähnliche logarithmische Vorfaktoren berücksichtigt werden.

Im restlichen Teil dieses Abschnittes gehen wir auf die adaptiven Algorithmen für die Multiskalentransformationen bzw. den Basiswechsel von/zu den Lifting-Wavelets ein. Wir beschränken uns dabei auf Interpolets. Die benutzte Notation ist in Tabelle 4.1 aufgeführt.

$T(\mathbf{l}) := T(l_1) \times \dots \times T(l_d)$	
$\mathcal{T}(\mathbf{l}) \subseteq T(\mathbf{l})$	Teil der adaptiven Indexmenge für Level \mathbf{l}
$\mathcal{T} := \bigcup_{\mathbf{l}} \mathbf{l} \times \mathcal{T}(\mathbf{l})$	Indexmenge für die adaptive Basis
$L := \max\{ \mathbf{l} _\infty : \mathcal{T}(\mathbf{l}) \neq \emptyset\}$	maximales Level der adaptiven Basis
$\mathbf{u}_{\mathcal{T}(\mathbf{l})} := \{u_{(\mathbf{l}, \mathbf{t})}\}_{\mathbf{t} \in \mathcal{T}(\mathbf{l})}$	Zeilenvektor von Koeffizienten
$\mathbf{u}_{\mathcal{T}} := \{u_{(\mathbf{l}, \mathbf{t})}\}_{(\mathbf{l}, \mathbf{t}) \in \mathcal{T}}$	
$\Psi_{\mathcal{T}(\mathbf{l})} := \{\psi_{(\mathbf{l}, \mathbf{t})}\}_{\mathbf{t} \in \mathcal{T}(\mathbf{l})}$	Spaltenvektor von Basisfunktionen
$\Psi_{\mathcal{T}} := \{\psi_{(\mathbf{l}, \mathbf{t})}\}_{(\mathbf{l}, \mathbf{t}) \in \mathcal{T}}$	
$\Omega_{\mathcal{T}} := \bigcup_{(\mathbf{l}, \mathbf{t}) \in \mathcal{T}} \text{supp} \tilde{\psi}_{(\mathbf{l}, \mathbf{t})}$	adaptives dünnes Gitter

Tabelle 4.1: Notation für adaptive Algorithmen in d Dimensionen.

Im wesentlichen werden dabei die multivariaten Pendanten der Größen aus Tabelle 3.1 einfach nur fettgedruckt. Im nichtadaptiven multivariaten Fall wird die Multiskalentransformation durch sukzessives Anwenden der univariaten Wavelettransformation WT bzw. IWT bzgl. aller Koordinatenrichtungen ausgeführt. Analog wie auch bei der multivariaten diskreten Fourier-Transformation. Das gleiche Schema soll auch im adaptiven Fall benutzt werden. Damit dabei die adaptive (inverse) Transformation wieder einen Isomorphismus zwischen den Multiskalenkoeffizienten und den Knotenwerten des *adaptiven dünnen Gitters* $\Omega_{\mathcal{T}}$ darstellt, vgl. Theorem 3.1, muß die Indexmenge \mathcal{T} bzgl. jeder Koordinatenrichtung die Kegel-Bedingung (3.6) erfüllen:

Definition 4.1 (Multivariate Kegel-/Lifting-/Vollständigkeitsbedingung)

Eine Indexmenge \mathcal{T} erfüllt genau dann die multivariate Kegel-Bedingung, falls für alle $(\mathbf{l}, \mathbf{t}) \in \mathcal{T}$ und $1 \leq i \leq d$ die univariate Indexmenge

$$\mathcal{T} := \{ (k, s) : ((l_1, \dots, l_{i-1}, k, l_{i+1}, \dots, l_d), (t_1, \dots, t_{i-1}, s, t_{i+1}, \dots, t_d)) \in \mathcal{T} \}$$

die Kegel-Bedingung (3.6) erfüllt. Analog sind die multivariaten Versionen der Lifting-Bedingung und der Vollständigkeitsbedingung definiert.

Lemma 4.3 (Auffüllen für multivariate Kegel-/Lifting-Bedingung) Sei \mathcal{T} durch sukzessives Anwenden von Algorithmus *Fill* bzw. *LiftingFill* bzgl. der Koordinatenrichtungen $i = 1, \dots, d$ aus einer beliebigen endlichen Indexmenge \mathcal{T}^0 berechnet worden. Dann erfüllt \mathcal{T} die multivariate Kegel- bzw. Lifting-Bedingung.

Der Beweis ist in Anhang A.10 angegeben. Die Behandlung der Vollständigkeitsbedingung kann durch eine Modifikation von z.B. Algorithmus *Fill* realisiert werden.

Lemma 4.4 (Adaptive multivariate Interpolet-Transformation) Die Indexmenge \mathcal{T} erfüllt die multivariate Kegel-Bedingung. Für gegebene Knotenwerte $u(\Omega_{\mathcal{T}})$ einer Funktion u können mittels sukzessivem Hintereinanderanwenden von *AWTKS* bzgl. aller Koordinatenrichtungen die Koeffizienten $\mathbf{u}_{\mathcal{T}}$ des eindeutigen Interpolanten $\Psi_{\mathcal{T}} \cdot \mathbf{u}_{\mathcal{T}}$ von u auf $\Omega_{\mathcal{T}}$ berechnet werden. Umgekehrt kann man durch Hintereinanderanwenden von *AIWTKS* eine Funktion $\Psi_{\mathcal{T}} \cdot \mathbf{u}_{\mathcal{T}}$ auf $\Omega_{\mathcal{T}}$ auswerten.

Beweisskizze: Für Interpolets sind die dualen Skalierungsfunktionen $\tilde{\phi}^{(l, s)}$ Dirac-Funktionale, so daß $\tilde{\phi}^{(l, s)} = \prod_{i=1}^d \delta(\cdot - s_i 2^{-l_i})$. Somit können die Knotenwerte $u(\Omega_{\mathcal{T}})$ als $\{\langle \tilde{\phi}^{(l, s)}, u \rangle\}$ interpretiert werden, falls der Punkt $(s_1 2^{-l_1}, \dots, s_d 2^{-l_d})$ in $\Omega_{\mathcal{T}}$ liegt. Entsprechend Bemerkung

3.2 werden die inneren Produkte $\langle \tilde{\phi}^{(l,s)}, u \rangle$ beim Anwenden von AWTKS bzgl. der ersten Koordinatenrichtung in die inneren Produkte von u mit Funktionen der Form $\tilde{\psi}_{(l_1, t_1)}(x_1) \cdot \tilde{\phi}^{(l_2, s_2)}(x_2) \cdot \dots \cdot \tilde{\phi}^{(l_d, s_d)}(x_d)$ umgerechnet. Hier geht ein, daß \mathcal{T} die multivariate Kegel-Bedingung erfüllt. Durch das dimensionsrekursive Schema erhalten wir somit am Ende genau die Werte $\langle \tilde{\psi}_{(\mathbf{l}, \mathbf{t})}, u \rangle$. Die Umkehrung zeigt man analog.

Auf ganz ähnliche Weise kann man durch Hintereinanderanwenden von AH2L bzw. AL2H bzgl. aller Koordinatenrichtungen eine Funktion $\Psi_{\mathcal{T}} \cdot \mathbf{u}_{\mathcal{T}}$ in die Lifting-Interpolet-Basis umrechnen bzw. umgekehrt. Allerdings muß dabei \mathcal{T} die multivariate Lifting-Bedingung erfüllen. Nicht nur die adaptiven Basistransformationen, sondern alle multivariaten Algorithmen, die in dieser Arbeit beschrieben werden, bestehen aus dem Hintereinanderanwenden univariater adaptiver Algorithmen bezüglich einer oder mehrerer Koordinatenrichtungen. Es bietet sich deshalb an, eine einheitliche Schablone zu definieren, mit der solche abgeleiteten multivariaten Algorithmen bezeichnet werden können. Diese ist in Bild 4.2 dargestellt. Mit AX wird dabei einer der Algorithmen AWTKS, AIWTKS, AFD u.s.w. bezeichnet, die einen Vektor $\mathbf{u}_{\mathcal{T}}$ auf einen anderen Vektor $\mathbf{w}_{\mathcal{T}}$ abbilden. Die Koordinatenrichtung, bezüglich derer AX angewandt werden soll, ist $i \in \{1, \dots, d\}$.

Funktion $\mathbf{v}_{\mathcal{T}} = \text{MV}(\text{AX}, \mathbf{i}, \mathbf{u}_{\mathcal{T}})$
für alle Indizes
 $\underline{\mathbf{l}} = (l_1, \dots, l_{i-1}), \bar{\mathbf{l}} = (l_{i+1}, \dots, l_d)$ und $\underline{\mathbf{t}} = (t_1, \dots, t_{i-1}), \bar{\mathbf{t}} = (t_{i+1}, \dots, t_d)$
für die $\mathcal{T} := \{(l, t) : ((\underline{\mathbf{l}}, \bar{\mathbf{l}}), (\underline{\mathbf{t}}, \bar{\mathbf{t}})) \in \mathcal{T}\} \neq \emptyset$ do
 $\mathbf{u}_{\mathcal{T}} := \{u_{((\underline{\mathbf{l}}, \bar{\mathbf{l}}), (\underline{\mathbf{t}}, \bar{\mathbf{t}}))}^{(l, t)}\}_{(l, t) \in \mathcal{T}}$
 $\mathbf{w}_{\mathcal{T}} := \text{AX}(\mathbf{u}_{\mathcal{T}})$
 $\{v_{((\underline{\mathbf{l}}, \bar{\mathbf{l}}), (\underline{\mathbf{t}}, \bar{\mathbf{t}}))}^{(l, t)}\}_{(l, t) \in \mathcal{T}} := \mathbf{w}_{\mathcal{T}}$
end

Abbildung 4.2: Allgemeine Schablone für die Anwendung eines univariaten Algorithmus AX bzgl. der i -ten Koordinatenrichtung auf $\mathbf{u}_{\mathcal{T}}$.

In dieser Notation ist beispielsweise die zweidimensionale inverse Multiskalenttransformation des Vektors $\mathbf{u}_{\mathcal{T}}$ gleich $\mathbf{v}_{\mathcal{T}} = \text{MV}(\text{AIWTKS}, 2, \text{MV}(\text{AIWTKS}, 1, \mathbf{u}_{\mathcal{T}}))$. Die Knotenwerte der entsprechenden Funktion sind dann in der multivariaten Form der kompakten Speicherung aus Abschnitt 3.2 in $\mathbf{v}_{\mathcal{T}}$ abgelegt.

Die in $\text{MV}(\dots)$ enthaltene Struktur der multivariaten Algorithmen bietet sich auch ideal für eine Implementierung an. Der multivariate Algorithmus besteht dann nur aus Lade- und Speicheroperationen und dem Anwenden von AX, was eine Reihe von Vorteilen hat:

Bemerkung 4.1 (Einfache Programmierung) Ähnlich wie bei der Notation für $\text{MV}(\dots)$ kann man eine allgemeine Funktion implementieren, die den Zugriff auf die multivariaten Daten realisiert und die als weiteres Argument einen Zeiger auf die entsprechende univariate Funktion, wie AWTKS, AFD u.s.w., erhält. Der Zugriff auf die multivariate Daten braucht somit nur einmal implementiert werden, wodurch der Code sehr kompakt und einfach zu pflegen wird. Die eigentlichen Rechen- und Indexoperationen, wie z.B. Matrix-Vektor-Multiplikationen oder die Auswertung der Indexoperationen $R(\dots)$, brauchen nur für den univariaten Fall implementiert zu werden und können separat getestet werden.

Bemerkung 4.2 (Datenlokalität und Effizienz des Codes) Für die Verwaltung der multivariaten Daten müssen Speicher-effiziente Datenstrukturen verwendet werden, wogegen

für die univariaten Algorithmen Datenstrukturen verwendet werden können, die bezüglich Rechengeschwindigkeit optimiert sind (z.B. Arrays, Vektoren). Durch das damit verbundene Umkopieren in eine eindimensionale Datenstruktur entsteht eine hohe Datenlokalität. Dies ist um so wichtiger, da in den univariaten Algorithmen bei den Matrix-Vektor-Multiplikationen mitunter recht oft auf ein und denselben Koeffizienten zugegriffen werden muß. In dem für diese Arbeit implementierten Code werden im univariaten Teil Vektoren benutzt. Die Adaptivität entsteht durch die zusätzliche Verwaltung der Indextmengen $\mathcal{T}(l)$ bzw. $\mathcal{S}(l)$, die den tatsächlich aktiven Bereich der Vektoren bestimmen. Auf diese Weise ist es unter anderem sehr leicht möglich, nur temporär notwendige Werte, wie $\mathbf{u}^{\mathcal{S}(l)}$ u.s.w., schnell einzufügen oder zu löschen. Angesichts der großen Häufigkeit dieser Operationen würde dies selbst bei den schnellsten Speicher-effizienten Datenstrukturen, wie Hash-Tabellen oder Bäume, den Flaschenhals bilden.

Kapitel 5

Approximation von Funktionen

Es wird untersucht, wie gut sich Funktionen aus bestimmten Sobolev- bzw. Besov-Räumen mittels der im vorangegangenen Kapitel vorgestellten isotropen bzw. anisotropen Tensorprodukt-Multiskalenbasen approximieren lassen. Konkret soll für einen gegebenen Funktionenraum \mathcal{F} die maximal mögliche Konvergenzordnung $\alpha > 0$ der Best- n -Term-Approximation [DeV98] bestimmt werden. Oder anders ausgedrückt:

'Wenn man nur weiß, daß $u \in \mathcal{F}$ ist, welche Approximationsordnung kann man für die beiden unterschiedlichen Basen erwarten?'

Die Approximationsordnung ist dabei wie folgt definiert:

$$\alpha(\mathcal{F}) := \sup\{\alpha : \sigma_{n,\mathcal{L}^2}(u) \lesssim n^{-\alpha}\|u\|_{\mathcal{F}}, \forall u \in \mathcal{F}\}, \quad (5.1)$$

wobei

$$\sigma_{n,\mathcal{X}}(u) := \inf_{\Lambda: |\Lambda|=n} \inf_{\{u_\lambda\}_{\lambda \in \Lambda}} \|u - \sum_{\lambda \in \Lambda} u_\lambda \psi_\lambda\|_{\mathcal{X}} \quad (\text{Best-}n\text{-Term-Approx. in } \mathcal{X}\text{-Norm}).$$

Dabei bezeichnet ψ_λ isotrope bzw. anisotrope Tensorprodukt-Wavelets. Diese generische Notation, bei der λ für einen Index der Form $\lambda = (l, \mathbf{t}, \mathbf{e})$ oder $\lambda = (\mathbf{l}, \mathbf{t})$ steht, wird auch in den folgenden Teilen dieses Kapitels benutzt. Welche der Bedeutungen gerade gemeint ist, geht aus dem jeweiligen Kontext hervor. Außerdem nehmen wir an, daß beide Waveletbasen auf den gleichen univariaten Wavelets basieren, so daß wirklich nur das multivariate Konstruktionschema den Unterschied ausmacht. Zur weiteren Vereinfachung der Notation definieren wir $\alpha^{iso}(\mathcal{F})$ bzw. $\sigma_{n,\mathcal{X}}^{iso}(u)$ für eine Approximation mit isotropen Multiskalenfunktionen, und $\alpha^{aniso}(\mathcal{F})$ bzw. $\sigma_{n,\mathcal{X}}^{aniso}(u)$ für eine Approximation mit anisotropen Multiskalenfunktionen. Nimmt man an, daß der Rechenaufwand, der notwendig ist, um den Bestapproximanten bzw. eine hinreichend genaue Näherung davon zu bestimmen, proportional zu n ist, dann bestimmt (5.1) im wesentlichen die Effizienz der Approximation, d.h. Aufwand zu Genauigkeit. Es zeigt sich nun bei den hier untersuchten Funktionenräumen, daß $\alpha^{aniso} \geq \alpha^{iso}$ ist. Für einige Räume ist aber sogar $\alpha^{aniso} \gg \alpha^{iso}$. Die anisotropen Multiskalenfunktionen sind somit die potentiell effizienteren Ansatzfunktionen.

Für die Beweise der Abschätzungen analog zu (5.1) wird üblicherweise der Fehler zwischen einer Funktion u und der biorthogonalen Projektion von u auf den entsprechenden Ansatzraum herangezogen. In diesem Fall sind die Koeffizienten u_λ gleich $\langle \tilde{\psi}_\lambda, u \rangle$. Damit diese eindeutig bestimmt sind, müssen die dualen Multiskalenfunktionen mindestens in \mathcal{L}^2 sein. Deshalb nehmen wir in diesem Kapitel an, daß die multivariaten Ansatzfunktionen Tensorprodukte von

\mathcal{L}^2 -stabilen Multiskalenfunktionen, also echten Wavelets, wie z.B. den Lifting-Wavelets, sind. Jeweils am Ende der beiden Abschnitte wird skizziert, wie sich die Aussagen auch auf die instabilen Interpolets übertragen lassen. Die Idee dabei ist, daß man ohne große Schwierigkeiten zwischen der stabilen Lifting- und der instabilen Interpolet-Basis wechseln kann.

5.1 Sobolev-Räume und lineare Approximation

Für Sobolev-Räume sind die Approximationseigenschaften besonders gut untersucht und entsprechende Ergebnisse Standard. Das wesentliche Merkmal von Sobolev-Räumen ist, daß sie die globale (nicht-lokale) Glattheit von Funktionen charakterisieren. Besonders einsichtig ist dies, wenn man die Definition

$$\|u\|_{\mathcal{H}^s} := \left(\sum_{\mathbf{k}} |\hat{u}_{\mathbf{k}}|^2 (1 + |\mathbf{k}|_{\infty})^{2s} \right)^{1/2} < \infty, \quad u(\mathbf{x}) = \sum_{\mathbf{k}} \hat{u}_{\mathbf{k}} e^{i\mathbf{k} \cdot \mathbf{x}} \quad (5.2)$$

betrachtet, bei der die Sobolev-Räume durch das Abfall-Verhalten der Koeffizienten von u bzgl. der globalen Fourier-Basis charakterisiert werden.

Übertragen auf die Problemstellung von (5.1) führt dies dazu, daß die Rate α am stärksten durch solche Funktionen beschränkt wird, deren lokale Glattheit überall in etwa gleich ist, so daß das bestmögliche Λ in (5.1) den üblichen Voll-Gitter-Räumen entspricht. D.h. Λ hat die Form $\Lambda^L = \{(\mathbf{l}, \mathbf{t}) : |\mathbf{l}|_{\infty} \leq L\}$ bzw. $\Lambda^L = \{(l, \mathbf{t}, \mathbf{e}) : l \leq L\}$. Somit brauchen wir bei den Sobolev-Räumen nur den Approximationsfehler bzgl. der fest vorgegebenen Indexmengenfamilien $\{\Lambda^L\}$, $L \in \mathbb{N}$, zu untersuchen. Diesen Typ der Approximation nennt man auch *lineare Approximation*, da der Best-Approximant die orthogonale Projektion von u auf einen von u unabhängigen Ansatzraum ist. Der Best-Approximant ist also linear von u abhängig.

Theorem 5.1 *Sei $u \in H^s$ und sei $0 \leq t < s \leq N$, $t < \theta^*$. Dann gilt für die isotropen bzw. die anisotropen Tensorprodukt-Wavelets*

$$\inf_{\{u_{\lambda}\}_{\lambda \in \Lambda^L}} \|u - \sum_{\lambda \in \Lambda^L} u_{\lambda} \psi_{\lambda}\|_{\mathcal{H}^t} \lesssim |\Lambda^L|^{-(s-t)/d} \|u\|_{\mathcal{H}^s}.$$

Dabei bezeichnet Λ^L die Indexmenge

$$\Lambda^L = \{(l, \mathbf{t}, \mathbf{e}) : l \leq L\} \quad \text{bzw.} \quad \Lambda^L = \{(\mathbf{l}, \mathbf{t}) : |\mathbf{l}|_{\infty} \leq L\}.$$

Analoge Aussagen wie oben gelten auch für die Sobolev-Räume \mathcal{H}_{mix}^s [Kna00],

$$\mathcal{H}_{mix}^s([0, 1]^d) := \mathcal{H}^s([0, 1]) \otimes \dots \otimes \mathcal{H}^s([0, 1]) \quad (5.3)$$

die die globale Glattheit von gemischten Ableitungen charakterisieren. Die zentrale Aussage ist hier im folgenden Theorem zusammengefaßt.

Theorem 5.2 ([Kna00]) *Sei $u \in \mathcal{H}_{mix}^s$ und sei $0 \leq t < s \leq N$, $t < \theta^*$. Dann gilt für die anisotropen Tensorprodukt-Wavelets*

$$\inf_{\{u_{(\mathbf{l}, \mathbf{t})}\}} \|u - \sum_{(\mathbf{l}, \mathbf{t}) \in \mathcal{T}_L} u_{(\mathbf{l}, \mathbf{t})} \psi_{(\mathbf{l}, \mathbf{t})}\|_{\mathcal{H}^t} \lesssim |\mathcal{T}_L|^{-(s-t)} \|u\|_{\mathcal{H}_{mix}^s} \begin{cases} (\log |\mathcal{T}_L|)^{(d-1)(s-t)} & 0 \leq t/s \leq 1/L \\ 1 & 1/L < t/s \end{cases}.$$

Dabei bezeichnet \mathcal{T}_L die Indexmenge

$$\mathcal{T}_L = \{(\mathbf{l}, \mathbf{t}) : \mathbf{t} \in T(\mathbf{l}), \|\mathbf{l} - t/s\|_\infty \leq d - 1 + L(1 - t/s)\} .$$

Die isotrop verfeinerten Wavelets können die zusätzliche Glattheit von \mathcal{H}_{mix}^s -Räumen gegenüber \mathcal{H}^s -Räumen nicht ausnutzen. Mit der Abschätzung $\mathcal{H}^s \subset \mathcal{H}_{mix}^s$, welche im Sinne von $\mathcal{H}^{s+\epsilon} \not\subset \mathcal{H}_{mix}^s \forall \epsilon > 0$ scharf ist, folgen somit nur die Aussagen von Theorem 5.1.

Für \mathcal{H}_{mix} -Räume erlauben die anisotropen Wavelets bei hoher Dimension d also eine dramatisch effizientere Approximation als die isotropen Wavelets:

$$\alpha^{aniso}(\mathcal{H}_{mix}^s) = s > \frac{s}{d} = \alpha^{iso}(\mathcal{H}_{mix}^s) .$$

Die Indexmengen \mathcal{T}_L für die anisotropen Wavelets sind nicht willkürlich gewählt, sondern wurden im Sinne von (5.1) optimal gewählt, siehe [Kna00]. Die resultierenden Approximationsräume sind dann die sogenannten *Dünn-Gitter*- bzw. *hyperbolic cross* Räume [Tem93, Zen91, Kna00]. Die dabei auftretenden Indexmengen sind dadurch gekennzeichnet, daß für ein Level \mathbf{l} entweder alle oder gar kein Index \mathbf{t} in $\mathcal{T}(\mathbf{l})$ ist. Solche Indexmengen bekommen eine eigene Bezeichnung:

Definition 5.1 (Level-adaptive Indexmenge) Eine Indexmenge \mathcal{T} heißt *Level-adaptiv*, falls mit $(\mathbf{l}, \mathbf{t}) \in \mathcal{T}$ auch alle weiteren Indizes $(\mathbf{l}, \mathbf{s}), \mathbf{s} \in T(\mathbf{l})$, auch in \mathcal{T} sind. Außerdem muß \mathcal{T} die *multivariate Kegel-Bedingung* erfüllen.

Die obigen Aussagen sind z.B. in [Kna00] mit Hilfe von Normäquivalenzen gezeigt worden, die für anisotrope Tensorprodukte von Interpolets nicht gelten. Für die Lifting-Wavelets $\bar{\psi}_{(\mathbf{l}, \mathbf{t})}$ gehen die Beweise aber durch. Der Punkt ist nun, daß für die hier definierten Level-adaptiven Indexmengen $\text{span} \bar{\Psi}_{\mathcal{T}_L} = \text{span} \Psi_{\mathcal{T}_L}$ gilt. Damit folgen die Abschätzungen für die lineare Approximation von \mathcal{H}^s - oder \mathcal{H}_{mix}^s -regulären Funktionen auch für die nicht stabilen Interpolets.

Bemerkung 5.1 In [Bun92] sind für die anisotrope hierarchische Basis Abschätzungen für den *Interpolationsfehler* in der \mathcal{L}^2 -Norm angegeben worden. In diesem Fall scheinen nur Aussagen möglich zu sein, die um einen logarithmischen Faktor schlechter sind, als die Abschätzungen in Theorem 5.2. Dort wurde ja der Best-Approximationsfehler betrachtet. Sei $I_{\mathcal{T}_L} u$ der Interpolant von u bzgl. der durch \mathcal{T}_L definierten Ansatzräume, dann erhält man für den Interpolationsfehler nur $\|u - I_{\mathcal{T}_L} u\|_{\mathcal{L}^2} \lesssim 2^{-2L} L^{d-1} \|u\|_{\mathcal{H}_{mix}^2}$, wogegen der Approximationsfehler für diese Ansatzräume $\lesssim 2^{-2L} \|u\|_{\mathcal{H}_{mix}^2}$ ist. Für den Interpolationsfehler in der \mathcal{H}^1 -Semi-Norm erhält man wie in Theorem 5.2 die Abschätzung $\|u - I_{\mathcal{T}_L} u\|_{\mathcal{H}^1} \lesssim |\mathcal{T}_L|^{-1} \|u\|_{\mathcal{H}_{mix}^2}$.

5.2 Besov-Räume und nichtlineare Approximation

Der große Nachteil von Sobolev-Räumen ist, daß durch sie nur die globale Glattheit charakterisiert wird. In vielen Anwendungen, wie die Lösung von partiellen Differentialgleichungen, die Bearbeitung von Bildern u.s.w., treten jedoch Funktionen auf, die zwar fast überall sehr glatt sind, jedoch in einzelnen Punkten oder Kanten Singularitäten bzw. Unstetigkeitsstellen aufweisen. Die globale Glattheit ist dann gering. Entsprechend klein ist die Konvergenzrate bei der naiven linearen Approximation, vgl. Theoreme 5.1 und 5.2. Es ist aber offensichtlich, daß für den Fall lokalisierter Singularitäten eine echt adaptive Approximation in Sinne von

(5.1), bei der sowohl über die Indexmenge Λ als auch über die Koeffizienten $\{u_\lambda\}$ minimiert wird, deutlich effizienter ist. Besov-Räume stellen nun das Werkzeug dar, mit dem bestimmt werden kann, *wievieleffizienter* eine solche adaptive bzw. nichtlineare Approximation ist. Der Begriff 'nichtlinear' bezieht sich hierbei darauf, daß die Best- n -Term-Approximation in (5.1) keine lineare Abbildung darstellt, da beispielsweise im allgemeinen $2n$ Ansatzfunktionen notwendig sind, um die Summe der Best- n -Term-Approximanten zweier Funktionen u und v darzustellen.

Besov-Räume sind bereits Ende der 50er, Anfang der 60er Jahre im Zusammenhang mit Interpolationsmethoden zwischen Sobolev-Räumen entwickelt worden, [Tri78]. Eine der Definitionen war

$$\mathcal{B}_{p,q}^s := \left\{ u : \|u\|_{\mathcal{B}_{p,q}^s} := \|u\|_{\mathcal{L}^p} + \left(\int_{\mathbb{R}^d} |\mathbf{h}|^{-(qs+d)} \|\Delta_{\mathbf{h}}^m u\|_{\mathcal{L}^p}^q d\mathbf{h} \right)^{1/q} < \infty \right\}, \quad s < m \in \mathbb{N}. \quad (5.4)$$

Dabei bezeichnet $\Delta_{\mathbf{h}}^m$ den Differenz-Operator der Ordnung m in Richtung $\mathbf{h} \in \mathbb{R}^d$. Die obige Definition der Norm hängt vom Parameter m ab, jedoch sind die Normen für alle $m > s$ äquivalent [Tri78]. Vergleicht man diese Definition mit einer der alternativen Definitionen für Sobolev-Normen

$$\|u\|_{\mathcal{H}^s} \sim \|f\|_{\mathcal{L}^2} + \left(\int_{\mathbb{R}^d} |\mathbf{h}|^{-(2s+d)} \|\Delta_{\mathbf{h}}^m f\|_{\mathcal{L}^2}^2 d\mathbf{h} \right)^{1/2}, \quad (5.5)$$

dann zeigt sich, daß im wesentlichen die in (5.5) auftretenden \mathcal{L}^2 -Normen durch (schwächere) \mathcal{L}^p - bzw. \mathcal{L}^q -Normen ersetzt wurden. In diesem Sinne sind also Besov-Räume Verallgemeinerungen von Sobolev-Räumen mit $\mathcal{B}_{2,2}^s = \mathcal{H}^s$. Eine der entscheidenden Entdeckungen war nun, daß Besov-Räume sehr eng verknüpft sind mit der nichtlinearen Approximation von Funktionen durch Wavelets [FJ90]. Dies führt auch zu etwas handlicheren Charakterisierungen von Besov-Normen. Entsprechende Aussagen sind in Theorem 5.3 zusammengefaßt.

Wird bei der Approximation der Fehler in der \mathcal{L}^w -Norm gemessen, dann ist es sinnvoll, für den Raum \mathcal{F} in (5.1) die Einbettung $\mathcal{F} \subset \mathcal{L}^w$ zu fordern. Für Besov-Räume ergibt sich dabei folgende Situation [DeV98]:

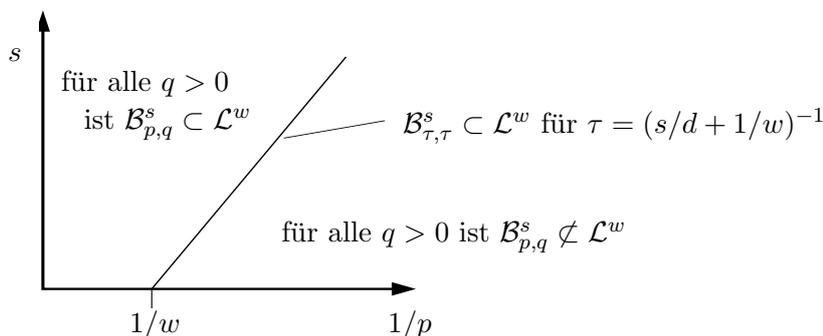
$$\begin{aligned} \mathcal{B}_{p,q}^s &\subset \mathcal{L}^w && \text{falls } p > \tau = (s/d + 1/w)^{-1} \text{ oder } p = q = \tau \\ \mathcal{B}_{p,q}^s &\not\subset \mathcal{L}^w && \text{falls } p < \tau \end{aligned}$$

Zur Darstellung der entsprechenden Bereiche benutzt man oft Diagramme wie in Bild 5.1. Analoge Aussagen gelten für die Einbettung von $\mathcal{B}_{p,q}^s$ in \mathcal{H}^t . Hier ist der kritische Parameter $\tau = ((s-t)/d + 1/2)^{-1}$. In diesem Abschnitt betrachten wir die Approximation in \mathcal{H}^t -Räumen. Dafür benutzen wir die etwas intuitivere Notation

$$\mathcal{B}_{\mathcal{H}^t}^s = \mathcal{B}_{\tau,\tau}^s, \quad \tau = ((s-t)/d + 1/2)^{-1}$$

für die in obigem Sinne maximalen Besov-Räume, die in den jeweiligen Meß-Raum \mathcal{H}^t eingebettet sind.

Theorem 5.3 ([Dah99b, CDH97, DeV98, DDD97]) *Seien $\{\psi_\lambda\}$ isotrop verfeinerte Wavelets mit $\|\psi_\lambda\|_{\mathcal{L}^2} \sim 1$, die für $r \in]-\tilde{\gamma}^*, \gamma^*[$ \mathcal{H}^r -stabil sind. Die dualen Wavelets haben N verschwindende Momente. Für die Parameter s und t gelte $t \in]-\tilde{\gamma}^*, \gamma^*[$, $s \geq t$ und $s \leq N$. Weiterhin sei $\psi_\lambda \in \mathcal{B}_{r,\tau}^s$ für ein $r > s$. Dann sind die folgenden Aussagen äquivalent*

Abbildung 5.1: Einbettung von Besov-Räumen in \mathcal{L}^w .

$$(i) \quad u \in \mathcal{B}_{\mathcal{H}^t}^s$$

$$(ii) \quad \sum_{(l,t,e)} \left(2^{lt} |\langle \tilde{\psi}_{(l,t,e)}, u \rangle| \right)^\tau < \infty \quad (5.6)$$

$$(iii) \quad \sum_{n=1}^{\infty} \left(2^{n(s-t)/d} \sigma_{2^n, \mathcal{H}^t}^{iso}(u) \right)^\tau < \infty \quad (5.7)$$

Die Terme auf der linken Seite von (5.6) bzw. (5.7) sind äquivalent zu $\left(\|u\|_{\mathcal{B}_{\mathcal{H}^t}^s} \right)^\tau$.

Aussage (ii) zeigt, daß Besov-Räume genau durch das Abfallverhalten der Koeffizienten bzgl. einer Waveletbasis charakterisiert sind. Im Gegensatz zu Sobolev-Räumen, vgl. (5.5), sind diese Basisfunktionen in Ort und Level aber lokal. Mittels des Kriteriums (5.6) ist es möglich, die Besov-Regularität für gegebene Funktionen nachzurechnen.

Aussage (iii) entspricht schon fast der gesuchten Beziehung zwischen dem Fehler der Best- n -Term-Approximation $\sigma_{n, \mathcal{H}^t}(u)$ und der Besov-Regularität von u , denn aus (5.7) folgt sofort

$$\sigma_{n, \mathcal{H}^t}(u) \lesssim n^{-(s-t)/d} \|u\|_{\mathcal{B}_{\mathcal{H}^t}^s}. \quad (5.8)$$

An einem kleinen Beispiel soll nun der enorme Unterschied zwischen der linearen Approximation/Sobolev-Regularität und der nichtlinearen Approximation/Besov-Regularität dargestellt werden.

Beispiel 5.1 (Punkt-Singularität) Für $\mathbf{x} \in]0, 1]^d$ und $\beta \in \mathbb{R}$ sei $u(\mathbf{x}) = |\mathbf{x}|^\beta$. Mit einer Nebenrechnung (Anhang A.11) folgt daraus, daß $u \in \mathcal{H}^{d/2+\beta-\epsilon}$ für beliebig kleines $\epsilon > 0$. Für eine Approximation in \mathcal{L}^2 muß u selbst in \mathcal{L}^2 sein, woraus die Bedingung $\beta > -d/2$ resultiert. Falls $\beta > -d/2$, erhält man für die lineare Approximation mit isotropen Wavelets die \mathcal{L}^2 -Fehler-Abschätzung $\lesssim |\Lambda|^{-\min(N, \beta+d/2)/d}$. Mit einer analogen Nebenrechnung (Anhang A.11) folgt aber auch, daß für $\beta > -d/2$, u in *jedem* Besov-Raum $\mathcal{B}_{\mathcal{H}^0}^s$, $s > 0$, liegt. Somit wird die Konvergenzordnung der nichtlinearen Approximation in der \mathcal{L}^2 -Norm nur durch die verwendeten Wavelets beschränkt: $\sigma_{n, \mathcal{L}^2}^{iso}(u) \lesssim n^{N/d} \|u\|_{\mathcal{B}_{\mathcal{H}^0}^N}$.

Die obigen Ausführungen zeigen, daß die nichtlineare Approximation mit isotropen Wavelets sehr gut verstanden ist. Es stellt sich die Frage, welche Aussagen für die Approximation mit anisotropen Wavelets möglich sind. Eine erste einfache Antwort gibt das folgende Lemma. Die

Idee dabei ist, daß sich jede univariate Skalierungsfunktion $\phi^{(l,s)}$ mit maximal $O(l)$ Wavelets $\psi_{(k,t)}$, $k \leq l$, darstellen läßt, so daß für die Darstellung von $\psi_{(l,t,e)}$ maximal $O(l^{d-1})$ anisotrope Wavelets genügen.

Lemma 5.1 *Sei $u \in \mathcal{B}_{\mathcal{H}^t}^s$ und sei $\{\Lambda_k\}_{k \in \mathbb{N}}$ eine aufsteigende Folge von Indermengen für isotrope Wavelets mit $\inf_{\{u_\lambda\}} \|u - \sum_{\lambda \in \Lambda_k} u_\lambda \psi_\lambda\|_{\mathcal{H}^t} \lesssim |\Lambda_k|^{-(r-t)/d}$, $r = \min\{s, N\}$. L_k sei das maximale Level von Λ_k .*

(i) *Falls $|\Lambda_k|$ exponentiell mit L_k anwächst, d.h. es gibt ein $a > 1$ mit*

$$|\Lambda_k| \sim a^{L_k} , \quad (5.9)$$

dann gilt für die Best- n -Term-Approximation mit anisotropen Wavelets $\sigma_{n,\mathcal{H}^t}^{aniso}(u) \lesssim (n/(\log n)^{d-1})^{-(r-t)/d}$.

(ii) *Falls die Anzahlen von Indizes auf Level l in Λ_k ab einem bestimmten Level exponentiell abnehmen, d.h. es gibt $a > 1$ und l_k , so daß*

$$|\{(l, \mathbf{t}, \mathbf{e}) \in \Lambda_k\}| \lesssim \begin{cases} 2^{ld} & l_0 \leq l \leq l_k \quad (\text{voll besetzte Level}) \\ |\Lambda_k| a^{-l} & l_k < l \leq L_k \end{cases} , \quad (5.10)$$

dann gilt $\sigma_{n,\mathcal{H}^t}^{aniso}(u) \lesssim n^{-(r-t)/d}$.

Beweis: Zu Λ_k definiert man \mathcal{T}_k als die kleinste Menge von Indizes (\mathbf{l}, \mathbf{t}) mit $\text{span}\{\psi_{(l,\mathbf{t},\mathbf{e})}\}_{(l,\mathbf{t},\mathbf{e}) \in \Lambda_k} \subset \text{span}\{\psi_{(\mathbf{l},\mathbf{t})}\}_{(\mathbf{l},\mathbf{t}) \in \mathcal{T}_k}$. Dann gilt nach Voraussetzung

$$\inf_{\{u_{(\mathbf{l},\mathbf{t})}\}} \|u - \sum_{(\mathbf{l},\mathbf{t}) \in \mathcal{T}_k} u_{(\mathbf{l},\mathbf{t})} \psi_{(\mathbf{l},\mathbf{t})}\|_{\mathcal{H}^t} \lesssim |\Lambda_k|^{-(r-t)/d} .$$

Man muß also nur noch $|\mathcal{T}_k|$ durch $|\Lambda_k|$ abschätzen. Für (i) gilt $|\mathcal{T}_k| \lesssim |\Lambda_k| (\log_2 |\Lambda_k|)^{d-1}$. Für (ii) gilt $|\mathcal{T}_k| \lesssim 2^{l_k d} + |\Lambda_k| \sum_{l=l_k}^{L_k} a^{-l} l^{d-1} \lesssim |\Lambda_k|$. Daraus folgt die Behauptung. \square

Die Aussagen von Lemma 5.1 lassen vermuten, daß man Funktionen aus Besov-Räumen mit anisotropen Wavelets mindestens ähnlich gut approximieren kann, wie mit den isotropen Wavelets, falls beide Basen auf den gleichen univariaten Wavelets basieren. Dies ist tatsächlich der Fall, wie Untersuchungen in [Hoc99] gezeigt haben. Es gilt $\alpha^{aniso}(\mathcal{B}_{\mathcal{H}^t}^s) \geq \alpha^{iso}(\mathcal{B}_{\mathcal{H}^t}^s)$. Die entsprechenden Ergebnisse sollen nachfolgend für den Fall der Approximation in \mathcal{L}^2 wiedergegeben werden. Man beachte jedoch, daß in der Definition der Konvergenzrate α logarithmische Terme wie in $(n/(\log n)^{d-1})^{-(r-t)/d}$ nicht erfaßt werden. Solche logarithmischen Terme spielen bei festem d in der Asymptotik nur eine untergeordnete Rolle. In praktischen Anwendungen können sie aber bedeuten, daß man absolut gesehen sehr viel mehr anisotrope Ansatzfunktionen benötigt, um den gleichen Fehler zu erreichen, wie mit isotropen Wavelets. Dieses Problem ist für hohe Raumdimensionen d besonders akut.

Für den Nachweis von $\alpha^{aniso} \geq \alpha^{iso}$ definiert man zunächst in Analogie zu Theorem 5.3 (iii) Räume, die genau die Approximierbarkeit mit anisotropen Wavelets charakterisieren.

Definition 5.2 *Für $0 < s$ und $\tau = (s/d + 1/2)^{-1}$ sei*

$$\mathcal{A}_{\mathcal{H}^t}^s := \left\{ u : \left(\sum_{n=1}^{\infty} \left(2^{ns/d} \sigma_{2^n, \mathcal{H}^t}^{aniso}(u) \right)^\tau \right)^{1/\tau} < \infty \right\} .$$

Alternativ können diese Räume auch durch das Abfallverhalten der Waveletkoeffizienten beschrieben werden.

Theorem 5.4 ([Hoc99] Theorem 4.2.2.4)

$$u \in \mathcal{A}_{\mathcal{H}^t}^s \Leftrightarrow \left(\sum_{(\mathbf{1}, \mathbf{t})} |2^{|\mathbf{1}| \infty t} \langle \tilde{\psi}_{(\mathbf{1}, \mathbf{t})}, u \rangle |^\tau \right)^{1/\tau} < \infty .$$

Es ist natürlich kein Zufall, daß die obige Aussage vollkommen analog zum isotropen Fall ist. Für die Abschätzung des \mathcal{L}^2 -Fehlers bei der Best- n -Term-Approximation benötigt man lediglich die \mathcal{L}^2 -Stabilität der Wavelets. Die Bauart (isotrop/anisotrop) spielt dabei keine Rolle.

Im Hinblick auf die Approximierbarkeit durch anisotropen Wavelets ist es nun wichtig, die Einbettungen von $\mathcal{B}_{p,q}^s$ in $\mathcal{A}_{\mathcal{H}^0}^s$ zu untersuchen. Man betrachtet dazu sogenannte anisotrope Besov-Räume, die wie folgt definiert sind:

$$\mathcal{B}_{p,q}^s := \{u : \|u\|_{\mathcal{L}^p} + \sum_{i=1}^d \left(\int_{\mathbb{R}} |h|^{-(s_i q + 1)} \|\Delta_{h\mathbf{e}_i}^{m_i}\|_{\mathcal{L}^p}^q dh \right)^{1/q} < \infty\} , \quad m_i > s_i . \quad (5.11)$$

In [Hoc99] wird zuerst eine etwas andere Definition angegeben, die aber äquivalent zu (5.11) ist (Theorem 4.1.2.1). Diese Räume sind Verallgemeinerungen der isotropen Besov-Räume.

Lemma 5.2 ([Hoc99]) *Es gilt*

$$\mathcal{B}_{p,q}^{s\mathbf{1}} = \mathcal{B}_{p,q}^s .$$

Die anisotropen Besov-Räume stehen zwischen den isotropen Besov-Räumen und den Approximationsräumen $\mathcal{A}_{\mathcal{H}^0}^s$. Einerseits umfassen sie die $\mathcal{B}_{p,q}^s$, andererseits umfassen sie auch Funktionen mit anisotroper Glattheit, welche vermutlich mit anisotropen Wavelets besser approximiert werden können als mit isotropen.

Als weitere Vorbereitung benötigt man wieder Aussagen bzgl. der Einbettung von $\mathcal{B}_{p,q}^s$ in den Mess-Raum \mathcal{L}^2 :

Lemma 5.3 ([Hoc99]) *Sei $\mathbf{s} > \mathbf{0}$ und $\tau = (s/d + 1/2)^{-1}$ wobei $s = d / \sum_i 1/s_i$. Dann gilt*

$$\mathcal{B}_{\tau,\tau}^s \subset \mathcal{L}^2 .$$

Der Einfachheit halber kann man also analog zu $\mathcal{A}_{\mathcal{H}^0}^s$ die Räume

$$\mathcal{B}_{\mathcal{H}^0}^s := \mathcal{B}_{\tau,\tau}^s$$

definieren. Der Parameter s ist eine Art gemittelte Glattheit. Es gilt für alle $\mathbf{s} > \mathbf{0}$, daß $s \geq \min(s_1, \dots, s_d)$, wobei $s = \min(s_1, \dots, s_d)$ genau dann gilt, wenn $s_1 = \dots = s_d$. Dieser Glattheitsparameter bestimmt auch die Konvergenzrate der Best- n -Term-Approximation für $\mathcal{B}_{\mathcal{H}^0}^s$, wie das folgende zentrale Theorem zeigt.

Theorem 5.5 *Sei $\mathbf{0} < \mathbf{s} \leq N\mathbf{1}$ und $s = d / \sum_i 1/s_i$, dann gilt*

$$\alpha^{iso}(\mathcal{B}_{\mathcal{H}^0}^s) = \min(s_1, \dots, s_d)/d \quad \text{bzw.} \quad \alpha^{aniso}(\mathcal{B}_{\mathcal{H}^0}^s) = s/d .$$

Beweis: Die Aussage bzgl. α^{iso} folgt aus der Einbettung $\mathcal{B}_{\mathcal{H}^0}^s \subset \mathcal{B}_{\mathcal{H}^0}^{\min(s_1, \dots, s_d)}$, die im Sinne von $\mathcal{B}_{\mathcal{H}^0}^s \not\subset \mathcal{B}_{\mathcal{H}^0}^{\min(s_1, \dots, s_d)+\epsilon}$ scharf ist, und (5.8). Die Aussage bzgl. α^{aniso} wird in [Hoc99] Theorem 4.2.3.3 bewiesen. \square

Eine wichtige Folgerung aus diesem Theorem ist, daß für $\mathcal{B}_{\mathcal{H}^0}^{s_1} = \mathcal{B}_{\mathcal{H}^0}^s$ mit anisotropen Wavelets bis auf eventuelle logarithmische Faktoren die gleiche Approximationsordnung erzielt wird wie mit den isotropen Wavelets. Man verliert also nichts. Dies verallgemeinert die Aussage von Lemma 5.1. Hat die zu approximierende Funktion aber eine anisotrope Glattheit, d.h. $s_i \neq s_j$ für ein i, j , dann sind die anisotropen Wavelets asymptotisch besser, da dann $s > \min(s_1, \dots, s_d)$. Aus diesem Grund sind die anisotropen Wavelets die aussichtsreicheren Kandidaten für effiziente adaptive Lösungsverfahren für partielle Differentialgleichungen.

Wir wollen für drei charakteristische Beispiele untersuchen, ob mit den anisotropen Wavelets eine ungefähr gleich gute oder bessere Approximation möglich ist als mit isotropen Wavelets. Dazu werden für die isotropen Wavelets jeweils Indexmengen Λ angegeben, mit denen man die optimale \mathcal{L}^2 -Approximationsrate erreicht, so daß die Indexmengen der Best- n -Term-Approximation sehr ähnlich aussehen müssen. Es wird dann untersucht, ob die Λ s eine der Bedingungen aus Lemma 5.1 erfüllen.

Beispiel 5.2 (Punktsingularität) Wir betrachten wieder die Funktion $u(\mathbf{x}) = |\mathbf{x}|^\beta$, allerdings nur für den zweidimensionalen Fall $\mathbf{x} \in]0, 1]^2$. Aus den Untersuchungen in Anhang A.11 erhalten wir für die Waveletkoeffizienten bezüglich isotroper Daubechies-Wavelets mit $N > \beta + 1$ verschwindenden Momenten

$$|\langle \psi_{(l, \mathbf{t}, \mathbf{e})}, u \rangle| \lesssim 2^{-l(1+\beta)} \begin{cases} 1 & \text{falls } l = l_0 \text{ oder } |\mathbf{t}| \leq 2N \\ |\mathbf{t}|^{\beta-N} & \text{sonst} \end{cases} .$$

Wir benutzen nun die Menge aller Indizes, für die die obige Abschätzung größer als ein vorgegebener Schwellwert δ ist:

$$\Lambda_\delta := \left\{ (l, \mathbf{t}, \mathbf{e}) : \mathbf{t} \in T(l) \times T(l) \text{ für } l \leq \frac{1}{N+1} \log_2 \delta^{-1} =: l_\delta \right. \\ \left. \text{oder } |\mathbf{t}| \leq (\delta^{-1} 2^{-l(1+\beta)})^{1/(N-\beta)} =: R(l) \text{ für } l_\delta < l \leq \frac{1}{1+\beta} \log_2 \delta^{-1} =: L_\delta \right\} .$$

D.h. alle Level bis l_δ sind vollbesetzt und das maximale Level ist L_δ . Für diese Indexmengen kann man nachweisen (siehe Anhang A.12), daß

$$|\Lambda_\delta| \sim \delta^{-2/(N+1)} \quad \text{sowie} \quad \|u - \sum_{\lambda \in \Lambda_\delta} u_\lambda \psi_\lambda\|_{\mathcal{L}^2} \lesssim \delta^{\frac{N}{N+1}} \sim |\Lambda_\delta|^{-N/2} .$$

Man erhält also mit dem Schwellwert-Ansatz genau die optimale Approximationsordnung, vergleiche (5.8). Da für $l_\delta < l \leq L_\delta$

$$|\{(l, \mathbf{t}, \mathbf{e}) \in \Lambda_\delta\}|/|\Lambda_\delta| \sim R(l)^2 \delta^{2/(N+1)} \lesssim 2^{-2l(1+\beta)/(N-\beta)} ,$$

ist offensichtlich die Bedingung (5.10) erfüllt, so daß die Best- n -Term-Approximation mit anisotropen Tensorprodukt-Wavelets mindestens genauso schnell konvergiert wie mit isotropen Wavelets.

Beispiel 5.3 (Diagonaler Sprung) In diesem Fall ist

$$u(x, y) := \begin{cases} 1 & x > y \\ 0 & \text{sonst} \end{cases} , \quad (x, y) \in [0, 1]^2 .$$

Für die Approximation betrachten wir wieder orthogonale Wavelets der Ordnung N . Es ist klar, daß für $l > l_0$ alle Waveletkoeffizienten $\langle u, \psi_{(l, \mathbf{t}, \mathbf{e})} \rangle$ verschwinden, falls der Träger von $\psi_{(l, \mathbf{t}, \mathbf{e})}$ die Diagonale $\{(x, x) : 0 \leq x \leq 1\}$ nicht schneidet. Ansonsten erhält man lediglich die Abschätzungen

$$|\langle \psi_{(l, \mathbf{t}, \mathbf{e})}, u \rangle| \lesssim \int_{x \geq y} |\psi_{(l, \mathbf{t}, \mathbf{e})}| d\mathbf{x} \lesssim 2^{-l},$$

so daß bei einem Abschneiden mit Schwellwert δ die Indexmengen

$$\Lambda_\delta := \{(\mathbf{l}, \mathbf{t}) : |t_1 - t_2| \lesssim 1, l_0 \leq l \leq |\log_2 \delta|\}$$

entstehen. Also $|\Lambda_\delta| \sim \delta^{-1}$. Der Abschneidefehler selbst ist $\sim \delta^{1/2}$. Also hat man $\sigma_{n, \mathcal{L}^2}^{iso}(u) \lesssim n^{-1/2}$. Die Indexmengen erfüllen Bedingung (5.9), so daß bei der Approximation mit anisotropen Wavelets die etwas schwächere Abschätzung $\sigma_{n, \mathcal{L}^2}^{aniso}(u) \lesssim (n/\log n)^{-1/2}$ gilt.

Beispiel 5.4 (Kantensingularität) Die Funktion u ist durch

$$u(x, y) = x^\beta, \quad (x, y) \in [0, 1]^2, \quad \beta > -1/2 \quad (5.12)$$

gegeben. Wir betrachten wieder die Approximation in der \mathcal{L}^2 -Norm, zunächst für isotrope Wavelets. Analog wie bei der Punktsingularität können die Waveletkoeffizienten durch

$$|\langle \psi_{(l, \mathbf{t}, \mathbf{e})}, u \rangle| \lesssim 2^{-l(1+\beta)} \begin{cases} 1 & l = l_0 \text{ oder } t_1 < 2N \\ t_1^{\beta-N} & \text{sonst} \end{cases}$$

abgeschätzt werden. Damit $u \in \mathcal{B}_{\mathcal{H}^0}^s$ ist, muß demnach entsprechend Theorem 5.3 (ii)

$$\sum_{l=l_0}^{\infty} 2^{-\tau l(1+\beta)} \left[2N + \sum_{t_1 \geq 2N}^{2^l} t_1^{-\tau(N-\beta)} \right] \left(\sum_{t_2 \in T(l)} 1 \right) < \infty$$

sein. O.B.d.A. können wir annehmen, daß N so groß ist, daß $1 - \tau(N - \beta) < 0$. Dann ist der obige Ausdruck wegen $|T(l)| \sim 2^l$ proportional zu $\sum_l 2^{l(1-\tau(1+\beta))}$. Daraus ergibt sich die Bedingung $1 - \tau(1 + \beta) < 0$, d.h. $s < 2(1/2 + \beta)$. Obwohl die Funktion u lediglich die triviale Fortsetzung der 1D Punkt-Singularität ist, ist die Besov-Regularität erheblich kleiner als im 1D Fall: $u \in \mathcal{B}_{\mathcal{H}^0}^{1+2\beta-\epsilon}$. Die Sobolev-Regularität ist wie im 1D Fall: $u \in \mathcal{H}^{\frac{1}{2}(1+2\beta-\epsilon)}$. Somit ist die Konvergenzrate für die Best- n -Term-Approximation mit isotropen Wavelets nur noch doppelt so hoch wie mit linearer Approximation.

Bei der Approximation mit anisotropen Wavelets kann dagegen der 1D Charakter von u optimal ausgenutzt werden. Es gibt hier weder die Beschränkung durch die (isotrope) Besov-Regularität noch eine Abhängigkeit von der Dimension d , so daß wie in 1D $\sigma_{n, \mathcal{L}^2}^{aniso} \lesssim n^{-N}$ folgt. Untersuchungen der Funktion u für $\beta = -1/2$ und Approximation in $\mathcal{H}^{-1/2}$ mit Haar-Wavelets findet man in [Osw00]. Dort wird gezeigt, daß $\sigma_{n, \mathcal{H}^{-1/2}}^{iso}(u) \lesssim n^{-1/2}$, wogegen $\sigma_{n, \mathcal{H}^{-1/2}}^{aniso}(u) \lesssim n^{-5/4}$.

Diese drei Beispiele zeigen, daß der Nachweis der Besov-Regularität mit recht elementaren Mitteln erfolgt. Die dabei notwendigen Rechnungen entsprechen weitgehend denen, die für das direkte Abschätzen der Konvergenzrate der Best- n -Term-Approximation notwendig sind. Somit erscheint das Besov-Raum-Kalkül als keine wirklich neue Hilfe, da letztlich doch wieder elementar das Approximationsverhalten der für die Anwendung interessanten Singularitätenfunktionen untersucht werden muß. Nach diesem Schema werden auch die in der Literatur

[DD97, Dah99a] bekannten Beweise für die Besov-Regularität der Lösungen von elliptischen bzw. den Stokes-Gleichungen geführt.

Zum Abschluß noch eine Bemerkung zu nicht \mathcal{L}^2 -stabilen Wavelets wie den Interpolets:

Bemerkung 5.2 (Approximation mit Interpolets) In gewissem Sinne kann man hier wie am Ende von Abschnitt 5.1 argumentieren. Für die Tensorprodukte von Lifting-Wavelets gelten alle in diesem Abschnitt gemachten Aussagen. Mindestens falls Bedingung (5.10) erfüllt ist, ist es aber möglich die Indextmengen der Best- n -Term-Approximation so zu modifizieren, daß man die Lifting-Wavelets in Interpolets umrechnen kann – ohne die Zahl der Freiheitsgrade signifikant zu erhöhen, vgl. Abschnitt 3.2. Dann folgt, daß auch für Interpolets der Approximationsfehler wie z.b. $n^{-r/d}\|u\|_{\mathcal{B}_{\mathcal{L}^2}^r}$ abfällt.

An dieser Stelle sei auch noch auf die Arbeit [Don92] verwiesen, in der der Interpolationsfehler mit adaptiven Interpolet-Basen für Besov-Räume untersucht wird, die in den Raum der stetigen Funktionen eingebettet sind.

Kapitel 6

Adaptive Diskretisierungen

In diesem Kapitel werden adaptive Finite-Differenzen- und Kollokationsdiskretisierungen vorgestellt, die auf anisotropen Multiskalenfunktionen basieren. Für diese Diskretisierungen geben wir in den Abschnitten 6.1 bis 6.3 Algorithmen an, mit denen einfache bzw. gemischte partielle Ableitungen, $D^{r\mathbf{e}_i}u$ bzw. $D^{r_i\mathbf{e}_i+r_j\mathbf{e}_j}u$, sowie nichtlineare Funktionale $g(u)$ ausgewertet werden können. Die Multiplikation von u mit einer Funktion v ist ein spezielles nichtlineares Funktional. Damit sind wir in der Lage, allgemeine Differentialoperatoren, z.B.

$$A = \sum_{i,j=1}^d a_{ij}(\mathbf{x})D^{\mathbf{e}_i+\mathbf{e}_j} + \sum_{i=1}^d b_i(\mathbf{x})D^{\mathbf{e}_i} + c(\mathbf{x}) \quad (6.1)$$

zu diskretisieren. Die Algorithmen für die Teilprobleme bestehen vollständig aus dem Anwenden der in Kapitel 3 vorgestellten Algorithmen AFD,APG,AWTKS und AIWTKS bezüglich einer oder mehrerer Koordinatenrichtungen entsprechend dem Schema MV aus Abschnitt 4.2. Der anisotrope Tensorproduktansatz führt also nicht nur zur effizienten Approximation von Funktionen, sondern auch zu besonders einfachen Algorithmen.

In Abschnitt 6.4 vergleichen wir die Finite-Differenzen- und die Kollokationsmethode mit dem Galerkin-Verfahren in Bezug auf Anwendbarkeit und Konvergenztheorie am Beispiel des Poisson-Problems. Bei allen Verfahren basieren die Konvergenzaussagen auf der Stabilität der diskreten Operatoren und den Approximationseigenschaften der exakten Lösung u . Beim Galerkin-Verfahren führt dabei das Lemma von Céa und der Nitsche-Trick unmittelbar auf das Problem der Best- n -Term-Approximation von u . Für die Kollokations- bzw. die Finite-Differenzen-Technik wird ein etwas anderer Ansatz entwickelt, bei dem der Fehler zwischen u und der numerischen Lösung in einer \mathcal{H}^s -Norm durch den Interpolationsfehler in der \mathcal{H}^s -Norm und den Konsistenzfehler in der \mathcal{H}^{s-2} -Norm abgeschätzt wird. Im Unterschied zu früheren ähnlichen Ansätzen [Hac96] werden dabei keine über Gitterpunkte definierte, diskrete Sobolev-Normen verwendet. Unser Weg ist somit allgemein genug für beliebige adaptive Petrov-Galerkin- oder Finite-Differenzen-Verfahren. Im letzten Abschnitt dieses Kapitels werden Abschätzungen des Konsistenzfehlers für die Maximum-Norm bzgl. des adaptiven Gitters $\Omega_{\mathcal{T}}$ und für die \mathcal{H}^0 -Norm angegeben. Das Fernziel wäre hier, Abschätzungen in negativen Sobolev-Normen, z.B. $\|\cdot\|_{\mathcal{H}^{-2}}$, herleiten zu können. Dies gelingt jedoch noch nicht, da wir nur eine Darstellung des Konsistenzfehlers mittels der Interpolets kennen, die jedoch für negative Sobolev-Normen keine einseitigen Abschätzungen erlauben, siehe Lemma 4.2. Keine Fortschritte konnten in Bezug auf die für die Konvergenztheorie notwendigen Stabilitätsaussagen erzielt werden, so daß hier lediglich auf die in der Literatur bekannten Resultate

verwiesen werden kann.

6.1 Kollokationsmethode

Für einen Differentialoperator A der Form (6.1) und eine vorgegebene rechte Seite f betrachten wir das Modellproblem

$$Au = f \text{ in }]0, 1[^d \quad \text{und} \quad u = 0 \text{ auf } \partial]0, 1[^d. \quad (6.2)$$

Der Einfachheit halber nehmen wir an, daß eventuelle inhomogene Randwerte auf die rechte Seite f abgewälzt wurden. Im nächsten Abschnitt gehen wir detailliert auf die Behandlung von Randbedingungen ein. Bei der Kollokationsmethode wird die numerische Lösung $u_{\mathcal{T}}$ als Linearkombination der Ansatzfunktionen $\{\psi_{(\mathbf{l}, \mathbf{t})}\}_{(\mathbf{l}, \mathbf{t}) \in \mathcal{T}}$ dargestellt. Dabei ist \mathcal{T} eine vorgegebene endliche Indexmenge, die die multivariate Kegel-Bedingung erfüllt. Als Testfunktionen für (6.2) werden die dualen Multiskalenfunktionen benutzt, d.h. $u_{\mathcal{T}}$ ist durch $\langle \tilde{\psi}_{(\mathbf{l}, \mathbf{t})}, Au_{\mathcal{T}} \rangle = \langle \tilde{\psi}_{(\mathbf{l}, \mathbf{t})}, f \rangle \quad \forall (\mathbf{l}, \mathbf{t}) \in \mathcal{T}$ bestimmt. Dieses Gleichungssystem kann in Operatorform als

$$A_{\mathcal{T}} u_{\mathcal{T}} = I_{\mathcal{T}} f \quad , \quad A_{\mathcal{T}} := I_{\mathcal{T}} A \quad (6.3)$$

geschrieben werden, wobei $I_{\mathcal{T}} v := \langle \tilde{\Psi}_{\mathcal{T}}, v \rangle \Psi_{\mathcal{T}}$ der biorthogonale Projektor auf den Ansatzraum ist und $A_{\mathcal{T}}$ der diskrete Differentialoperator. Für die hier betrachteten Interpolets ist $I_{\mathcal{T}}$ der Interpolationsoperator bzgl. des Gitters $\Omega_{\mathcal{T}}$.

Die Kollokationsmethode ist algorithmisch besonders einfach und entsprechend effizient. Für die Berechnung der Koeffizienten $\langle \tilde{\psi}_{(\mathbf{l}, \mathbf{t})}, f \rangle$ der rechten Seite braucht man wegen Lemma 4.4 nur $MV(\text{AWTKS}, \dots)$ nacheinander bezüglich aller Koordinatenrichtungen auf die Knotenwerte $f(\Omega_{\mathcal{T}})$ anzuwenden. Die Auswertung von partiellen Ableitungen der Form $D^{r \mathbf{e}_i} u$ ist ebenfalls recht einfach.

Lemma 6.1 *Beim Petrov-Galerkin-Verfahren mit biorthogonalen anisotropen Multiskalenfunktionen als Ansatz- und Testfunktionen kann das Produkt $\mathbf{v}_{\mathcal{T}} := \langle \tilde{\Psi}_{\mathcal{T}}, D^{r \mathbf{e}_i} \Psi_{\mathcal{T}} \rangle \cdot \mathbf{u}_{\mathcal{T}}$ mittels $MV(\text{APG}, i, \mathbf{u}_{\mathcal{T}})$ berechnet werden.*

Beweis: Sei $(\mathbf{l}, \mathbf{t}) \in \mathcal{T}$, dann ist

$$\begin{aligned} v_{(\mathbf{l}, \mathbf{t})} &= \langle \tilde{\psi}_{(\mathbf{l}, \mathbf{t})}, D^{r \mathbf{e}_i} \sum_{(\mathbf{k}, \mathbf{s}) \in \mathcal{T}} u_{(\mathbf{k}, \mathbf{s})} \psi_{(\mathbf{k}, \mathbf{s})} \rangle = \sum_{(\mathbf{k}, \mathbf{s}) \in \mathcal{T}} u_{(\mathbf{k}, \mathbf{s})} \langle \tilde{\psi}_{(\mathbf{l}, \mathbf{t})}, D^r \psi_{(\mathbf{k}, \mathbf{s})} \rangle \prod_{j \neq i} \underbrace{\langle \tilde{\psi}_{(\mathbf{l}_j, \mathbf{t}_j)}, \psi_{(\mathbf{k}_j, \mathbf{s}_j)} \rangle}_{=\delta_{\mathbf{l}_j, \mathbf{k}_j} \delta_{\mathbf{t}_j, \mathbf{s}_j}} \\ &= \langle \tilde{\psi}_{(\mathbf{l}, \mathbf{t})}, D^r \sum_{k_i, s_i} u_{((\mathbf{l}, k_i, \bar{\mathbf{l}}), (\mathbf{t}, s_i, \bar{\mathbf{t}}))} \psi_{(k_i, s_i)} \rangle. \end{aligned}$$

Auf der rechten Seite steht am Ende der Ausdruck, der definitionsgemäß mit $MV(\text{APG}, i, \mathbf{u}_{\mathcal{T}})$ berechnet wird. Daraus folgt die Behauptung. \square

Etwas komplizierter ist die Auswertung von gemischten Ableitungen, wie $D^{r_i \mathbf{e}_i + r_j \mathbf{e}_j}$ mit $i \neq j$.

Lemma 6.2 *Beim Petrov-Galerkin-Verfahren mit biorthogonalen anisotropen Multiskalenfunktionen als Ansatz- und Testfunktionen kann speziell für Level-adaptive Indexmengen, vergleiche Abschnitt 5.1, das Matrix-Vektor-Produkt $\mathbf{v}_{\mathcal{T}} := \langle \tilde{\Psi}_{\mathcal{T}}, D^{r_i \mathbf{e}_i + r_j \mathbf{e}_j} \Psi_{\mathcal{T}} \rangle \cdot \mathbf{u}_{\mathcal{T}}$ wie folgt berechnet werden:*

$$\mathbf{v}_{\mathcal{T}} = MV(\text{APG}\mathbf{a}, j, MV(\text{APG}, i, \mathbf{u}_{\mathcal{T}})) + MV(\text{APG}, i, MV(\text{APG}\mathbf{b}, j, \mathbf{u}_{\mathcal{T}})).$$

Das heißt, man muß zunächst APG bezüglich der i -ten Koordinatenrichtung anwenden. Auf das Ergebnis wendet man den Teilalgorithmus APGa bezüglich der j -ten Koordinatenrichtung an. Dazu wird das Ergebnis des analogen zweiten Terms addiert. Der Beweis ist etwas länger und deshalb im Anhang A.13 aufgeführt. Möglicherweise funktioniert das Schema aus Lemma 6.2 auch für beliebige Indexmengen \mathcal{T} , die die multivariate Kegel-Bedingung erfüllen. Entsprechende Untersuchungen sind noch nicht abgeschlossen.

Mit den Algorithmen aus Lemma 6.1 und 6.2 können Differentialoperatoren mit konstanten Koeffizienten ausgewertet werden. Für Differentialoperatoren mit variablen Koeffizienten, hier mit $a(\mathbf{x})$ bezeichnet, sind aber beispielsweise Ausdrücke wie $\langle \tilde{\Psi}_{\mathcal{T}}, aD^{r_{e_i}}u \rangle$ auszuwerten. Dieser Term stellt die Multiskalenkoeffizienten des Interpolanten von $aD^{r_{e_i}}u$ bezüglich $\Omega_{\mathcal{T}}$ dar, welche wegen Lemma 4.4 mittels MV(AWTKS,.) aus den Knotenwerten $(aD^{r_{e_i}}u)(\Omega_{\mathcal{T}})$ berechnet werden können. Die Knotenwerte $D^{r_{e_i}}u(\Omega_{\mathcal{T}})$ können wegen Lemma 4.4 mittels MV(AIWTKS,.) aus $\langle \tilde{\Psi}_{\mathcal{T}}, D^{r_{e_i}}u \rangle$ berechnet werden. Für a ist üblicherweise eine geschlossene Formel bekannt, so daß auch die Berechnung von $a(\Omega_{\mathcal{T}})$ keine Probleme bereitet. Somit sind die Knotenwerte $(aD^{r_{e_i}}u)(\Omega_{\mathcal{T}})$ und damit $\langle \tilde{\Psi}_{\mathcal{T}}, aD^{r_{e_i}}u \rangle$ berechenbar. Das soeben beschriebene Vorgehen ist nur ein Spezialfall für die Auswertung allgemeiner nichtlinearer Funktionale:

Lemma 6.3 *Sei u eine beliebige Funktion, deren Multiskalenkoeffizienten $\mathbf{u}_{\mathcal{T}} = \langle \tilde{\Psi}_{\mathcal{T}}, u \rangle$ bekannt sind und $g : \mathbb{R} \rightarrow \mathbb{R}$ eine stetige Funktion. Dann können die Multiskalenkoeffizienten $\mathbf{v}_{\mathcal{T}} := \langle \tilde{\Psi}_{\mathcal{T}}, g \circ u \rangle$ mittels des folgenden Schemas berechnet werden:*

1. Berechne $u(\Omega_{\mathcal{T}})$ mittels Hintereinanderanwenden von MV(AIWTKS,...)
2. Berechne $g \circ u(\Omega_{\mathcal{T}})$
3. Berechne aus diesen Knotenwerten die gesuchten Koeffizienten $\mathbf{v}_{\mathcal{T}}$ mittels Hintereinanderanwenden von MV(AWTKS,...)

Beweis: Die Aussage ist eine unmittelbare Folgerung aus Lemma 4.4. \square

Auf analoge Weise kann man auch die Multiskalenkoeffizienten $\langle \tilde{\Psi}_{\mathcal{T}}, uv \rangle$ des Produktes zweier Funktionen u und v berechnen, falls die Koeffizienten $\mathbf{u}_{\mathcal{T}}$ und $\mathbf{v}_{\mathcal{T}}$ bekannt sind. Damit sind alle wesentlichen Algorithmen vorgestellt worden, die für das Kollokationsverfahren auf adaptiven dünnen Gittern notwendig sind.

6.2 Randbedingungen bei der Kollokationsmethode

Randbedingungen können bei der Kollokationsmethode ähnlich wie im univariaten Fall diskretisiert werden. Angenommen wir fordern $Au = f$ in $]0, 1[^d$ und $u = g_D$ auf Γ_D sowie $\partial_{\mathbf{n}}u = g_N$ auf Γ_N mit $\Gamma_D, \Gamma_N \subseteq \partial]0, 1[^d$. Mit Hilfe einer Funktion w , die die geforderten inhomogenen Randwerte annimmt, können wir dieses Randwertproblem in ein Problem mit homogenen Randwerten

$$A\hat{u} = f - Aw \text{ in }]0, 1[^d \quad \text{sowie} \quad \hat{u} = 0 \text{ auf } \Gamma_D, \quad \partial_{\mathbf{n}}\hat{u} = 0 \text{ auf } \Gamma_N$$

überführen. Für dieses Problem können Ansatzfunktionen benutzt werden, die die homogenen Randwerte bereits beinhalten. Bei diesem Ansatz besteht die Schwierigkeit, daß die anisotropen Tensorprodukt-Basisfunktionen die Randbedingungen immer auf kompletten Seitenflächen von $[0, 1]^d$ erfüllen. Es ist nicht ohne weiteres möglich, daß die Randbedingungen

innerhalb einer Seitenfläche wechseln, z.B. von $u = 0$ auf $\partial_{\mathbf{n}}u = 0$. In Bild 6.1 (rechts) ist so eine verbotene Situation skizziert. Solche Fälle können mit Gebietszerlegungsverfahren

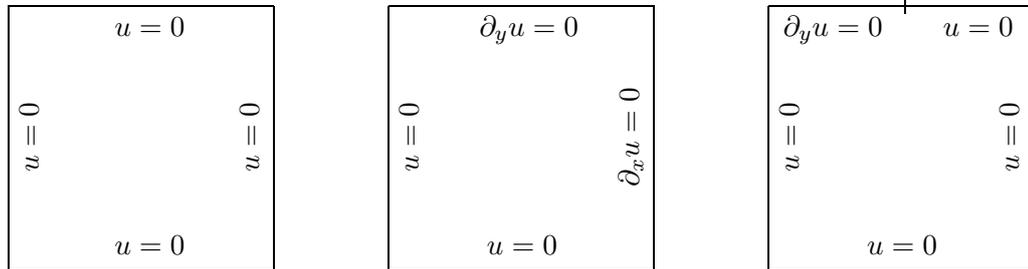


Abbildung 6.1: Links und Mitte: Beispiele für mögliche homogene Randbedingungen mit anisotropen Tensorprodukt-Wavelets. Randbedingungen wie im Bild rechts können nicht realisiert werden.

behandelt werden, bei denen das Gebiet zunächst in zwei Teilgebiete mit jeweils konstanten Randbedingungen auf den Seitenflächen zerlegt wird, siehe Abbildung 6.2 bzw. vgl. [Sch98a]. Allerdings wird dann eine äußere Iteration notwendig, durch die die Lösungen auf den beiden

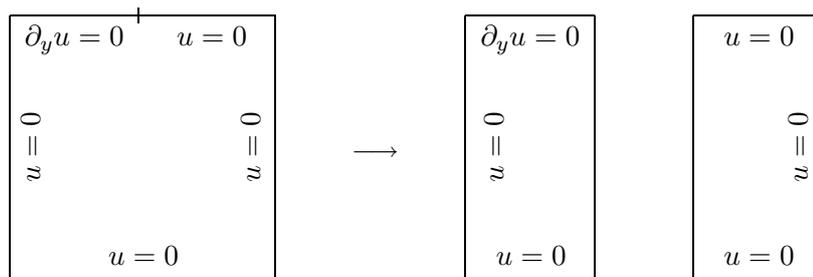


Abbildung 6.2: Zerlegung des Gebietes in zwei Teilgebiete mit konstantem Typ von Randbedingungen an den Seitenflächen.

Teilgebieten am inneren, künstlichen Rand miteinander gekoppelt werden [GSB96].

Deutlich komplizierter als im univariaten Fall ist die Konstruktion der Funktion w , die die eigentlich vorgegebenen inhomogenen Randwerte annimmt. Eines der dabei entstehenden Probleme ist, daß die Randwerte bestimmte Kompatibilitätsbedingungen erfüllen müssen. Dazu betrachten wir die folgenden drei Fälle, bei denen die Randwerte g_{i0} und g_{i1} , $i = 0, \dots, d$, vorgegeben sind.

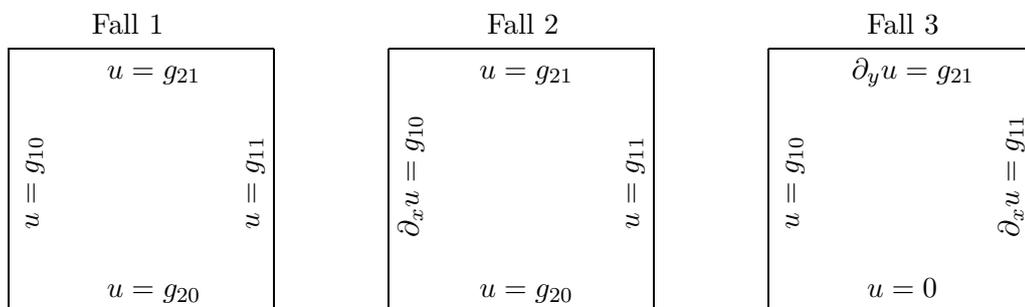


Abbildung 6.3: Unterschiedliche Fälle für inhomogene Randwerte für die Funktion w .

Da die zu konstruierende Funktion $w \in \text{span}\Psi_{\mathcal{T}}$ mindestens stetig differenzierbar ist, ergeben sich z.B. im Fall 1 die Kompatibilitätsbedingungen $g_{21}(1) = g_{11}(1)$, $g_{21}(0) = g_{10}(1)$, $g_{20}(1) = g_{11}(0)$ und $g_{10}(0) = g_{20}(0)$. Aber auch wenn solche Bedingungen erfüllt sind, ist die Konstruktion von w nur in einigen Fällen einfach. Einer davon ist der Fall, daß auf allen Seitenflächen von $[0, 1]^d$ Dirichlet-Randwerte vorgegeben sind. Man kann dann ähnlich wie in Abschnitt 3.5 vorgehen, wo der Randwert durch Setzen der unmittelbar randnächsten Multiskalenkoeffizienten erzwungen wurde. Im multivariaten Fall ist der Randwert natürlich eine Funktion auf einer $(d-1)$ -dimensionalen Seitenfläche, so daß hier jeweils Werte auf einem $(d-1)$ -dimensionalen Teilgitter gesetzt werden müssen. Der komplette Algorithmus zum Setzen der Dirichlet-Randwerte ist unten angegeben. Darin bezeichnen \underline{l}, \bar{l} und \underline{l}, \bar{l} Teilindizes der Form $\underline{l} = (l_1, \dots, l_{i-1})$ bzw. $\bar{l} = (l_{i+1}, \dots, l_d)$. Mit ${}^+\Psi_{\mathcal{T}}$ sind wieder die Multiskalenfunktionen ohne Randbedingungen gemeint.

Algorithmus DirichletRandwerte

for $i = 1$ to d

Setze $\mathcal{I}_i := \{((\underline{l}, \bar{l}), (\underline{t}, \bar{t})) \text{ für die } ((\underline{l}, l_0, \bar{l}), (\underline{t}, 0, \bar{t})) \in \mathcal{T}\}$

Berechne den $(d-1)$ -dimensionalen Interpolant ${}^+\Psi_{\mathcal{I}_i} \cdot \mathbf{g}_{\mathcal{I}_i}$ von g_{i0}

Setze $w_{((\underline{l}, l_0, \bar{l}), (\underline{t}, 0, \bar{t}))} = g_{((\underline{l}, \bar{l}), (\underline{t}, \bar{t}))}$ für alle $((\underline{l}, \bar{l}), (\underline{t}, \bar{t})) \in \mathcal{I}_i$

Berechne den $(d-1)$ -dimensionalen Interpolant ${}^+\Psi_{\mathcal{I}_i} \cdot \mathbf{g}_{\mathcal{I}_i}$ von g_{i1}

Setze $w_{((\underline{l}, l_0, \bar{l}), (\underline{t}, 2^{l_0}, \bar{t}))} = g_{((\underline{l}, \bar{l}), (\underline{t}, \bar{t}))}$ für alle $((\underline{l}, \bar{l}), (\underline{t}, \bar{t})) \in \mathcal{I}_i$

end

Ebenfalls recht einfach ist der Fall, wo an einigen Seiten von $[0, 1]^d$ Dirichlet-Werte gegeben sind und an anderen Seiten periodische Randbedingungen. Hier geht man wie im Algorithmus DirichletRandwerte vor, bearbeitet aber nur die Dirichlet-Seiten.

Falls Neumann-Seiten vollständig an Dirichlet-Seiten oder Seiten mit periodischen Randbedingungen angrenzen, ist es möglich, w auf die nachfolgende Art zu konstruieren. Wir betrachten das Beispiel von Fall 2 aus Bild 6.3.

1. Zunächst werden wie in Algorithmus DirichletRandwerte die Dirichlet-Seitenflächen $y = 0$, $x = 1$ und $y = 1$ bearbeitet. Das Ergebnis dieses Zwischenschritts sei $w^1 \in \text{span}{}^+\Psi_{\mathcal{T}}$.
2. Wir bestimmen den $(d-1)$ -dimensionalen Interpolant $v = {}^+\Psi_{\mathcal{I}_1} \cdot \mathbf{v}_{\mathcal{I}_1}$ von $\partial_x w^1 - g_{10}$ auf der Seitenfläche $x = 0$.
3. Die Multiskalenkoeffizienten der eigentlich gesuchten Funktion w sind nun

$$w_{(\mathbf{1}, \mathbf{t})} := \begin{cases} \frac{v_{(l_2, t_2)}}{\partial_x + \phi_{(l_0, 0)}(0)} & \text{falls } l_1 = l_0 \text{ und } t_1 = 0 \\ w_{(\mathbf{1}, \mathbf{t})}^1 & \text{sonst} \end{cases}$$

Es ist bis jetzt noch nicht gelungen, einen einfachen Algorithmus für den Fall zu finden, daß Neumann-Seitenflächen aneinanderstoßen. Der Grund dafür ist, daß die unmittelbar randnächsten Multiskalenkoeffizienten, wie $w_{((\underline{l}, l_0, \bar{l}), (\underline{t}, 2^{l_0}, \bar{t}))}$, nicht mehr ausreichen, um alle Bedingungen zu erfüllen. Letztendlich führt dann die Konstruktion von w auf ein lineares Gleichungssystem, dessen Lösung man nicht direkt ablesen kann, wie in oben diskutierten Fällen.

Relativ einfach gestaltet sich hingegen der Basiswechsel von Multiskalenfunktionen mit homogenen Randbedingungen zu Multiskalenfunktionen ohne Randbedingungen, der für die Addition von w und der Lösung \hat{u} des homogenen Problems notwendig ist. Die multivariate Version

von Algorithmus AUP besteht ganz einfach aus dem sukzessiven Hintereinanderanwenden von $MV(AUP, \dots)$ bezüglich aller Koordinatenrichtungen. Wegen der einfachen Tensorproduktstruktur der anisotropen Multiskalenfunktionen $\psi_{(\mathbf{l}, \mathbf{t})}(\mathbf{x}) = \psi_{(l_1, t_1)}(x_1) \cdot \dots \cdot \psi_{(l_d, t_d)}(x_d)$ werden dabei im ersten Schritt, d.h. $MV(AUP, 1, \cdot)$, die Basisfunktionen $\psi_{(l_1, t_1)}(x_1)$ mit homogenen Randwerten durch ${}^{\sharp}\psi_{(l_1, t_1)}(x_1)$ ersetzt, so daß dann ein Koeffizientenvektor bezüglich Basisfunktionen der Form ${}^{\sharp}\psi_{(l_1, t_1)}(x_1) \cdot \psi_{(l_2, t_2)}(x_2) \cdot \dots \cdot \psi_{(l_d, t_d)}(x_d)$ resultiert. Im nächsten Schritt, $MV(AUP, 2, \cdot)$, wird dann auch noch $\psi_{(l_2, t_2)}$ durch ${}^{\sharp}\psi_{(l_2, t_2)}$ ersetzt u.s.w..

6.3 Finite-Differenzen–Diskretisierung

Die Finite-Differenzen–Methode auf adaptiven dünnen Gittern ist sehr stark an die Kollokationsmethode aus Abschnitt 6.1 angelehnt. Auch hier wird die gegebene Differentialgleichung in der Form $A_{\mathcal{T}} u_{\mathcal{T}} = I_{\mathcal{T}} f$ diskretisiert. Jedoch ist $A_{\mathcal{T}}$ nicht mehr die biorthogonale Projektion des Operators auf den Ansatzraum, sondern etwas anders definiert: Speziell für partielle Ableitungen $D^{r_i e_i}$ und $u_{\mathcal{T}} = \Psi_{\mathcal{T}} \cdot \mathbf{u}_{\mathcal{T}}$ ist

$$A_{\mathcal{T}} u_{\mathcal{T}} := \Psi_{\mathcal{T}} \cdot MV(AFD, i, \mathbf{u}_{\mathcal{T}}) .$$

Gemischte Ableitungen, wie $D^{r_i e_i + r_j e_j}$, werden mit

$$\Psi_{\mathcal{T}} \cdot MV(AFD, j, MV(AFD, i, \mathbf{u}_{\mathcal{T}}))$$

diskretisiert. D.h. im allgemeinen unterscheiden sich die Kollokationsmethode und die Finite-Differenzen–Technik nur darin, daß der univariate Algorithmus APG durch AFD ersetzt wurde. In Abschnitt 6.5 wird der Konsistenzfehler für beide Verfahren untersucht. Darin zeigt sich, daß sich beide Verfahren auch sehr ähnlich verhalten. Der große Vorteil der Finite-Differenzen–Technik ist jedoch die deutlich höhere Flexibilität. Bei der Kollokationsmethode hat man genau eine Diskretisierung, wogegen in AFD eine große Vielzahl von Finite-Differenzen–Schemen benutzt werden können, die optimal an den Typ der Gleichung (z.B. hyperbolisch, Sattelpunktprobleme,...) oder Eigenschaften der Lösung (glatt, Schocks u.s.w.) angepaßt werden können. Dieses deutliche Mehr an Flexibilität zeigt sich auch bei der Behandlung von Randbedingungen. Während für Dirichlet-Daten die Homogenisierungsmethode hinreichend einfach und damit brauchbar ist, stellen insbesondere inhomogene Neumann-Randbedingungen doch ein gewisses Problem für die Kollokationsmethode dar. Bei der Finite-Differenzen–Technik bereitet es dagegen keine großen Probleme, in den Randpunkten spezielle angepaßte Differenzen-Sterne zu benutzen, vergleiche Abschnitt 3.6 bzw. siehe [Sch98a]. Die komplizierte Konstruktion der Funktion w entfällt somit.

Für Vollgitter-Indexmengen und spezielle Finite-Differenzen–Verfahren sind die obigen Diskretisierungen äquivalent zu klassischen Diskretisierungen auf uniformen Gittern.

Lemma 6.4 *Sei $\mathcal{T} = \{(\mathbf{l}, \mathbf{t}) : |\mathbf{l}|_{\infty} \leq L\}$. Weiterhin sei für das in AFD benutzte Finite-Differenzen–Schema die Operatorauswertung eine lineare Operation. Da $\Omega_{\mathcal{T}}$ ein uniformes Gitter mit Maschenweite 2^{-L} ist, können wir das Finite-Differenzen–Schema aus AFD auch in klassischer Weise auf die Knotenwerte $\mathbf{u}_{\mathcal{T}}(\Omega_{\mathcal{T}})$ anwenden. Das Ergebnis sei mit $w(\Omega_{\mathcal{T}})$ bezeichnet. Dann gilt $A_{\mathcal{T}} u_{\mathcal{T}}(\Omega_{\mathcal{T}}) = w(\Omega_{\mathcal{T}})$.*

Der Beweis ist in [Kos00] angegeben.

6.4 Vergleich verschiedener Diskretisierungstechniken

Anhand des Modellproblems

$$\Delta u = f \text{ in }]0, 1[^d \quad \text{und} \quad u = 0 \text{ auf } \partial]0, 1[^d. \quad (6.4)$$

wollen wir das Kollokations-, das Finite-Differenzen- und das Galerkin-Verfahren bzgl. Anwendbarkeit und Konvergenztheorie vergleichen. Beim *Galerkin*-Verfahren ist $u_{\mathcal{T}} \in \text{span}\Psi_{\mathcal{T}}$ durch die Bedingung

$$\langle \psi_{(\mathbf{l}, \mathbf{t})}, \Delta u_{\mathcal{T}} \rangle = \langle \psi_{(\mathbf{l}, \mathbf{t})}, f \rangle \quad \forall (\mathbf{l}, \mathbf{t}) \in \mathcal{T}$$

eindeutig definiert. Damit das Galerkin-Verfahren wohldefiniert ist, muß $\Delta \psi_{(\mathbf{l}, \mathbf{t})}$ und f in einem Sobolev-Raum \mathcal{H}^s mit $s > -\gamma^*$, vergleiche Tabelle 2.1, sein. Der numerische Fehler kann wegen der \mathcal{H}^1 -Elliptizität des Laplace-Operators mit dem Lemma von Céa sehr einfach durch den Fehler der Best-Approximation abgeschätzt werden

$$\|u - u_{\mathcal{T}}\|_{\mathcal{H}^1} \lesssim \inf_{v_{\mathcal{T}} \in \text{span}\Psi_{\mathcal{T}}} \|u - v_{\mathcal{T}}\|_{\mathcal{H}^1}.$$

Auf diese Weise erhält man unmittelbar eine Beziehung zwischen der Konvergenz des numerischen Verfahrens und dem in Kapitel 5 diskutierten Problem der Approximation von Funktionen. Falls die Indexmengen \mathcal{T} ungefähr den Indexmengen entsprechen, mit denen die Best- n -Term Approximation von u realisiert wird, dann konvergiert das Galerkin-Verfahren mit der optimalen Rate. Untersuchungen zum Galerkin-Verfahren auf einfachen level-adaptiven dünnen Gittern findet man z.B. in [Zen91, Bun92, Pfl95].

Etwas problematisch ist die algorithmische Seite des Galerkin-Verfahrens. Im allgemeinen können die Matrix-Vektor-Multiplikationen oder die Skalarprodukte $\langle \psi_{(\mathbf{l}, \mathbf{t})}, f \rangle$ für die rechte Seite nicht exakt berechnet werden, sondern müssen auch wieder approximiert werden. Für das Quadraturproblem ist z.B. für isotrope Wavelets in [BCU00] ein Algorithmus entwickelt worden, der in seiner Struktur dem Algorithmus für die Auswertung von Finite-Differenzen-Operatoren im adaptiven Fall aus Abschnitt 3.6 gleicht. Die Quadraturfehler für z.B. $\langle \Psi_{\mathcal{T}}, f \rangle$ müssen natürlich auch bei der Konvergenzanalyse berücksichtigt werden. Das Hilfsmittel wäre hier das Lemma von Strang und Fix, siehe [Bra92].

Ebenfalls etwas problematisch ist die Steifigkeitsmatrix $\langle \Psi_{\mathcal{T}}, \Delta \Psi_{\mathcal{T}} \rangle$ selbst. Obwohl sie einigermaßen dünnbesetzt ist, sind selbst für den einfachen Laplace-Operator die Kosten für die Matrix-Vektor-Multiplikation deutlich höher als bei den Kollokations- oder Finite-Differenzen-Verfahren. Außerdem wächst für Interpolets auch nach diagonalem Skalieren die Konditionszahl sehr unangenehm in Abhängigkeit vom maximalen Level L und der Raum-Dimension d , siehe Kapitel 7. Der Grund dafür als auch für die teure Matrix-Vektor-Multiplikation ist der sogenannte Massmatrix-Anteil in der Steifigkeitsmatrix

$$\langle \psi_{(\mathbf{l}, \mathbf{t})}, \sum_{i=1}^d \partial_{x_i x_i} \psi_{(\mathbf{k}, \mathbf{s})} \rangle = \sum_{i=1}^d \langle \psi_{(l_i, t_i)}, \partial_{x_i x_i} \psi_{(k_i, s_i)} \rangle \cdot \underbrace{\prod_{j \neq i} \langle \psi_{(l_j, t_j)}, \psi_{(k_j, s_j)} \rangle}_{\text{Massmatrix-Anteil}}.$$

Bei z.B. der Kollokationsmethode ist wegen der Biorthogonalität von Ansatz- und Testfunktionen der äquivalente Term in der Steifigkeitsmatrix trivial, siehe Lemma 6.1.

Das Kollokations- und das Finite-Differenzen-Verfahren sind in ihrer Struktur so ähnlich, daß wir sie gemeinsam betrachten können. Beide Verfahren können in Operatorform als

$$A_{\mathcal{T}}u_{\mathcal{T}} = I_{\mathcal{T}}f$$

geschrieben werden. Eine Voraussetzung für die Wohldefiniertheit der beiden Verfahren ist, daß $f \in \mathcal{H}^s$ ist mit $s > d/2$, da die Testfunktionen aus $\mathcal{H}^{-d/2-\epsilon}$ sind. Für die Kollokationsmethode kommt noch die Forderung an die Ansatzfunktionen, $\Delta\psi_{(1,t)} \in \mathcal{H}^s$ mit $s > d/2$, dazu. Der Anwendungsbereich ist also kleiner als beim Galerkin-Verfahren. Deutliche Unterschiede gibt es auch in Bezug auf die Konvergenztheorie. Während beim Galerkin-Verfahren der numerische Fehler durch den Best-Approximationsfehler abgeschätzt werden konnte, ist dies beim Kollokations- bzw. Finite-Differenzen-Verfahren nicht ohne weiteres möglich. Hier bleibt im wesentlichen nur, den Fehler durch den Konsistenz-Fehler abzuschätzen. Ähnlich wie bei der klassischen Analyse von Finite-Differenzen-Verfahren könnte dies beispielsweise in der Maximum-Norm $\|v\|_{\Omega_{\mathcal{T}}} := \max_{\mathbf{x} \in \Omega_{\mathcal{T}}} |v(\mathbf{x})|$ bzgl. $\Omega_{\mathcal{T}}$ geschehen. Weiterhin braucht man noch die entsprechenden Stabilität des diskreten Operators.

Definition 6.1 (Stabilität) Die Operatoren $A_{\mathcal{T}}$ heißen stabil, falls es eine von \mathcal{T} unabhängige Konstante $\alpha > 0$ gibt, mit

$$\alpha \|v_{\mathcal{T}}\|_{\Omega_{\mathcal{T}}} \leq \|A_{\mathcal{T}}v_{\mathcal{T}}\|_{\Omega_{\mathcal{T}}} \quad \forall v_{\mathcal{T}} \in \text{span}\Psi_{\mathcal{T}} .$$

Für Vollgitter-Indexmengen und lineare Operatoren $A_{\mathcal{T}}$ kann man mit Hilfe von Lemma 6.4 bekannte Stabilitätsaussagen für z.B. klassische Finite-Differenzen-Verfahren [Hac96] auf $A_{\mathcal{T}}$ übertragen. Für den allgemeinen Fall scheint es noch keine Aussagen zu geben. Mit dem obigen Stabilitätsbegriff erhält man unmittelbar

$$\begin{aligned} \|u - u_{\mathcal{T}}\|_{\Omega_{\mathcal{T}}} = \|I_{\mathcal{T}}u - u_{\mathcal{T}}\|_{\Omega_{\mathcal{T}}} &\leq \frac{1}{\alpha} \|A_{\mathcal{T}}(I_{\mathcal{T}}u - u_{\mathcal{T}})\|_{\Omega_{\mathcal{T}}} = \frac{1}{\alpha} \|A_{\mathcal{T}}I_{\mathcal{T}}u - I_{\mathcal{T}}f\|_{\Omega_{\mathcal{T}}} \\ &= \frac{1}{\alpha} \|A_{\mathcal{T}}I_{\mathcal{T}}u - I_{\mathcal{T}}\Delta u\|_{\Omega_{\mathcal{T}}} = \frac{1}{\alpha} \|\Delta u - A_{\mathcal{T}}I_{\mathcal{T}}u\|_{\Omega_{\mathcal{T}}} . \end{aligned} \quad (6.5)$$

D.h. der numerische Fehler wird durch den Konsistenzfehler $\Delta u - A_{\mathcal{T}}I_{\mathcal{T}}u$ abgeschätzt. In der klassischen Analyse von Finite-Differenzen-Diskretisierungen wird dieser letzte Term weiter abgeschätzt, wobei vorausgesetzt wird, daß u hinreichend oft stetig oder Lipschitz-stetig differenzierbar ist. Entsprechende Aussagen für die hier beschriebenen Diskretisierungen sind in Abschnitt 6.5 angegeben. Leider sind diese Glattheitsforderungen oft nicht erfüllt, oder die so erzielten Abschätzungen für den Konsistenzfehler sind sehr schwach. Eine Möglichkeit dieses Problem zu umgehen, ist den Konsistenzfehler nicht in der starken $\|\cdot\|_{\Omega_{\mathcal{T}}}$ -Norm zu messen, sondern in *schwächeren* Sobolev-Normen. Entsprechend wird aber ein *stärkerer* Stabilitätsbegriff für $A_{\mathcal{T}}$ benötigt:

Definition 6.2 ($(s, s-2)$ -Stabilität, [BA72, Hac96, Sch98b]) Die Operatoren $A_{\mathcal{T}}$ heißen $(s, s-2)$ -stabil, falls es eine von \mathcal{T} unabhängige Konstante $\alpha_s > 0$ gibt, mit

$$\alpha_s \|v_{\mathcal{T}}\|_{\mathcal{H}^s} \leq \|A_{\mathcal{T}}v_{\mathcal{T}}\|_{\mathcal{H}^{s-2}} \quad \forall v_{\mathcal{T}} \in \text{span}\Psi_{\mathcal{T}} .$$

Diese Stabilitätseigenschaft ist ebenfalls schwierig nachzuweisen. Die einzigen dem Autoren bekannten Aussagen für (Spline-)Kollokations- oder Finite-Differenzen-Verfahren findet man in [AW85, Hac96, Sch98b] und den Zitaten darin. Allerdings wird dabei intensiver Gebrauch

von Fourier-Techniken gemacht, so daß die Aussagen letztlich nur für volle Gitter gültig sind. Auch die in der ersten Arbeit [AW85] enthaltenen Konvergenzaussagen beruhen darauf, daß in diesem einfachen Fall die exakte und die numerische Lösung explizit als Fourier-Entwicklung angegeben und gegeneinander abgeschätzt werden können. Im adaptiven Fall ist das natürlich nicht möglich. Stattdessen kann man hier folgendermaßen vorgehen:

$$\begin{aligned} \|u - u_{\mathcal{T}}\|_{\mathcal{H}^s} &\leq \|u - I_{\mathcal{T}}u\|_{\mathcal{H}^s} + \|I_{\mathcal{T}}u - u_{\mathcal{T}}\|_{\mathcal{H}^s} \\ &\leq \|u - I_{\mathcal{T}}u\|_{\mathcal{H}^s} + \frac{1}{\alpha_s} \|A_{\mathcal{T}}(I_{\mathcal{T}}u - u_{\mathcal{T}})\|_{\mathcal{H}^{s-2}} \\ &= \|u - I_{\mathcal{T}}u\|_{\mathcal{H}^s} + \frac{1}{\alpha_s} \|A_{\mathcal{T}}I_{\mathcal{T}}u - I_{\mathcal{T}}\Delta u\|_{\mathcal{H}^{s-2}} . \end{aligned} \quad (6.6)$$

D.h. der numerische Fehler wird mittels $(s, s - 2)$ -Stabilität durch den Interpolationsfehler in der \mathcal{H}^s -Norm und den Konsistenz-Fehler in der schwächeren \mathcal{H}^{s-2} -Norm abgeschätzt. Für die Konvergenzanalyse können wir somit die folgende Strategie benutzen:

1. Konstruiere Indexmengen \mathcal{T} , die an die lokale Glattheit von u angepaßt sind. Dies kann mittels eines Schwellwert-Ansatzes über die Multiskalenkoeffizienten von u geschehen.
2. Abschätzen des Interpolationsfehlers (bzw. Projektionsfehlers bei allgemeinen Petrov-Galerkin-Schemen) sowie des Konsistenzfehlers.

Durch den Schwellwert-Ansatz sind die Indexmengen \mathcal{T} unmittelbar an die Approximations-eigenschaften von u gekoppelt. Ideal wäre nun, wenn \mathcal{T} so gewählt werden könnte, daß sich der Interpolationsfehler ungefähr wie der Fehler der Best- $|\mathcal{T}|$ -Term-Approximation verhält: $\|u - I_{\mathcal{T}}u\|_{\mathcal{H}^s} \sim \sigma_{|\mathcal{T}|, \mathcal{H}^s}^{aniso}(u)$ und *gleichzeitig* der Konsistenzfehler von der gleichen Größenordnung ist $\|A_{\mathcal{T}}I_{\mathcal{T}}u - I_{\mathcal{T}}\Delta u\|_{\mathcal{H}^{s-2}} \sim \sigma_{|\mathcal{T}|, \mathcal{H}^s}^{aniso}(u)$. In diesem Fall könnte das numerische Verfahren theoretisch mit optimaler Ordnung konvergieren. Die Einschränkung 'theoretisch' ist notwendig, da man natürlich auch einen Algorithmus für die Konstruktion der optimalen Indexmengen bei unbekanntem u benötigt (Steuerung der Adaptivität, siehe Kapitel 8).

Soll der numerische Fehler $u - u_{\mathcal{T}}$ in der \mathcal{L}^2 -Norm gemessen werden, dann wäre dazu der Konsistenzfehler in der 'negativen' Sobolev-Norm $\|\cdot\|_{\mathcal{H}^{-2}}$ zu messen. Wie wir in Abschnitt 6.5 sehen werden, haben wir leider nur eine Darstellung von $A_{\mathcal{T}}I_{\mathcal{T}}u - I_{\mathcal{T}}\Delta u$ mittels der Interpolets, die jedoch keine einseitigen Abschätzungen für negative Sobolev-Normen erlauben. Aus diesem Grund gelingt es uns in dieser Arbeit noch nicht, durch den Übergang zu Sobolev-Normen bessere Fehlerabschätzungen für die Kollokationsmethode zu gewinnen als bezüglich der Maximum-Norm.

Prinzipiell scheint die oben skizzierte Vorgehensweise aber sehr vielversprechend, wie erste Untersuchungen für biorthogonale Petrov-Galerkin Verfahren mit stabilen Lifting-Wavelets bzw. ihren dualen Wavelets als Ansatz-/Testfunktionen gezeigt haben. Die entsprechenden Ergebnisse sind am Ende von Abschnitt 6.5 skizziert.

6.5 Konsistenz für partielle Ableitungsoperatoren D^{re_i}

In diesem Abschnitt untersuchen wir für die Finite-Differenzen- und die Kollokationsmethode den Konsistenzfehler $I_{\mathcal{T}}Au - A_{\mathcal{T}}I_{\mathcal{T}}u$ für partielle Ableitungsoperatoren der Form $A = D^{re_i}$. Es ist klar, daß damit sofort Abschätzungen für Operatoren folgen, die als Summe solcher partiellen Ableitungen geschrieben werden können. Speziell für die Maximum-Norm bzgl. $\Omega_{\mathcal{T}}$

folgen auch Abschätzungen für Operatoren bD^{re_i} mit nicht-konstanten Koeffizienten $b(\mathbf{x})$. Entsprechend Abschnitt 6.1 wird bD^{re_i} mittels $B_{\mathcal{T}} := I_{\mathcal{T}}(bA_{\mathcal{T}})$ diskretisiert. Wir können somit den Konsistenzfehler für $B_{\mathcal{T}}$ auf den für $A_{\mathcal{T}}$ zurückführen:

$$\|I_{\mathcal{T}}(bD^{re_i}u) - I_{\mathcal{T}}(bA_{\mathcal{T}}I_{\mathcal{T}}u)\|_{\Omega_{\mathcal{T}}} = \|I_{\mathcal{T}}(b(D^{re_i}u - A_{\mathcal{T}}I_{\mathcal{T}}u))\|_{\Omega_{\mathcal{T}}} \leq \|b\|_{\Omega_{\mathcal{T}}} \|D^{re_i}u - A_{\mathcal{T}}I_{\mathcal{T}}u\|_{\Omega_{\mathcal{T}}}.$$

Da wir in dieser Arbeit vornehmlich untersuchen wollen, wie weit man mit dem im vorangegangenen Abschnitt entwickelten Konzept für einen Konvergenzbeweis kommen kann, betrachten wir hier nicht den Konsistenzfehler für gemischte Ableitungen. Für erste Resultate dazu sei auf [Sch98a] verwiesen.

Im allgemeinen ist es recht schwierig, $I_{\mathcal{T}}Au - A_{\mathcal{T}}I_{\mathcal{T}}u$ für beliebige Indexmengen \mathcal{T} zu bestimmen. Wir gehen hier deshalb den Weg, zunächst bestimmte, an u angepaßte Indexmengen zu definieren, für die dann der Konsistenzfehler abgeschätzt werden kann. Wie auch in Kapitel 5 ist hierbei der Abfall von Multiskalenkoeffizienten der entscheidende Mechanismus, mit dem die Glattheit von u die Indexmengen \mathcal{T} beeinflusst. Es bietet sich dabei an, die Glattheit von u in sogenannten Lipschitz- bzw. Lipschitz-Nikolskii-Räumen zu messen.

Definition 6.3 (Lipschitz-Räume) *Der Raum $C^{\alpha}(\Omega)$, $\alpha > 0$, ist die Menge aller Funktionen $u : \Omega \subset \mathbb{R} \rightarrow \mathbb{R}$, die $[\alpha]_-$ -mal stetig differenzierbar sind und für die die Semi-Norm*

$$|u|_{\alpha} := \sup_{x, x+h \in \Omega} \frac{|\Delta_h v(x)|}{h^{(\alpha - [\alpha]_-)}} < \infty, \quad v := D^{[\alpha]_-}u$$

beschränkt ist. Δ_h ist der Differenzenoperator $\Delta_h v(x) = v(x) - v(x+h)$. Der Term $[\alpha]_-$ bezeichnet die größte ganze Zahl, die echt kleiner als α ist, z.B. ist $[1.1]_- = 1$ und $[1]_- = 0$.

Die für anisotrope Tensorprodukt-Basen passende Verallgemeinerung dieser Räume mißt wie bei den \mathcal{H}_{mix} -Räumen die Glattheit der gemischten Ableitungen.

Definition 6.4 (Lipschitz-Nikolskii-Räume [Tri78]) *Der Raum $C^{\alpha}(\Omega)$, $\alpha > 0$, ist die Menge aller Funktionen $u : \Omega \subset \mathbb{R}^d \rightarrow \mathbb{R}$, für die die gemischte Ableitung $v := D^{[\alpha]_-}u$ stetig ist und*

$$|u|_{\alpha} := \sup_{\mathbf{x}, \mathbf{x}+\mathbf{h} \in \Omega} \frac{|\Delta_{h_1 \mathbf{e}_1} \circ \dots \circ \Delta_{h_d \mathbf{e}_d} v(\mathbf{x})|}{h_1^{(\alpha_1 - [\alpha_1]_-)} \cdot \dots \cdot h_d^{(\alpha_1 - [\alpha_1]_-)}} < \infty.$$

Bemerkung 6.1 Falls alle Komponenten α_i ganzzahlig sind und die gemischte Ableitung $D^{\alpha}u$ stetig ist, dann folgt aus dem Mittelwertsatz

$$|u|_{\alpha} \leq \sup_{\mathbf{x}} |D^{\alpha}u(\mathbf{x})|.$$

Funktionen aus solchen Lipschitz-Räumen sind mindestens stetig. Lösungen von partiellen Differentialgleichungen haben aber im allgemeinen eine geringe globale Glattheit und können Singularitäten oder Sprünge enthalten. Es stellt sich deshalb die Frage, ob Lipschitz-Räume in diesem Zusammenhang überhaupt sinnvoll sind. Der Punkt ist nun, daß man fast immer theoretisch oder praktisch die singuläre Funktion durch eine regularisierte Variante ersetzen kann, die beliebig glatt ist. Natürlich nimmt diese mit kleiner werdendem Regularisierungsparameter mehr und mehr die Eigenschaften der ursprünglichen Funktion an: lokale Gradienten werden beliebig steil. Idealerweise würde ein adaptives Verfahren solche Situationen erkennen und in der Nähe kritischer Punkte mehr und mehr Freiheitsgrade verwenden. In diesem

Sinne können wir annehmen, daß formal die Voraussetzungen für die Kollokations-/Finite-Differenzen-Methode erfüllt sind, trotzdem aber auch Adaptivität sinnvoll und notwendig ist. In diesem Fall sind letztlich lokale Eigenschaften wie die lokale Semi-Norm $|u|_{\text{supp}\psi_{(\mathbf{l}, \mathbf{t})}}|_{\boldsymbol{\alpha}}$ entscheidend. Das nachfolgende Lemma zeigt, daß die Multiskalenkoeffizienten $u_{(\mathbf{l}, \mathbf{t})}$ als lokale Indikatoren für diese Größe dienen können.

Lemma 6.5 (Abfall von Multiskalenkoeffizienten) *Sei $u \in \mathcal{C}^{\boldsymbol{\alpha}}$. Dann gilt*

$$|\langle \tilde{\psi}_{(\mathbf{l}, \mathbf{t})}, u \rangle| \lesssim |u|_{\text{supp}\psi_{(\mathbf{l}, \mathbf{t})}}|_{\boldsymbol{\beta}} \cdot \prod_{i=1}^d 2^{-l_i \beta_i}, \quad \beta_i = \min(\alpha_i, N).$$

Dieses Lemma ist eine Verallgemeinerung von schon lange bekannten Abschätzungen [Bun92]. Der Beweis ist in Anhang A.14 zu finden. Mit diesen Abschätzungen für die Multiskalenkoeffizienten kann man in einem kleinen Gedankenexperiment Indexmengen \mathcal{T} herleiten, die für eine effiziente Approximation einer Funktion $u \in \mathcal{C}^{\boldsymbol{\alpha}}$ geeignet sind. Nimmt man an, daß außer $u \in \mathcal{C}^{\boldsymbol{\alpha}}$ keine weiteren Aussagen über die lokale Glattheit bekannt sind, ist \mathcal{T} als Level-adaptive Indexmenge zu wählen. Ein Ansatz wäre dann, alle Indizes (\mathbf{l}, \mathbf{t}) zu wählen, für die $|u|_{\boldsymbol{\beta}} \cdot \prod_{i=1}^d 2^{-l_i \beta_i}$ größer als ein vorgegebener Schwellwert δ ist. Wählt man $\delta \sim 2^{-L \min_j \beta_j}$, dann haben die Indexmengen die Form

$$\mathcal{T} = \left\{ (\mathbf{l}, \mathbf{t}) : \sum_j l_j \beta_j \leq L \min_j \beta_j \right\} \quad (6.7)$$

Mit diesen Indexmengen wird in Koordinatenrichtungen j , für die β_j klein ist, stark verfeinert (l_j kann groß werden) und für großes β_j bleibt l_j klein.

Für die Indexmengen (6.7) soll nun der Konsistenzfehler in der Maximum-Norm abgeschätzt werden. Da für Level-adaptive Indexmengen die Kollokationsmethode als ein spezielles Finite-Differenzen-Verfahren der Konsistenzordnung $K = N - r$ interpretiert werden kann, sind die Abschätzungen für beide Verfahren gültig.

Theorem 6.1 (Konsistenz für D^{re_i} und Level-adaptive Ansatz-Räume) *Sei $u \in \mathcal{C}^{\boldsymbol{\alpha}}$, $\alpha_i > r$, $\boldsymbol{\beta} \geq 0$ wie in Lemma 6.5 und \mathcal{T} von der Form (6.7). Das in AFD benutzte Finite-Differenzen-Schema für die Diskretisierung der r -ten Ableitung hat die Konsistenzordnung K . Dann gilt*

$$\|D^{re_i} u - \Psi_{\mathcal{T}} \cdot MV(\text{AFD}, i, \mathbf{u}_{\mathcal{T}})\|_{\Omega_{\mathcal{T}}} \leq C(u) L^{d-1} 2^{-\min(N, K, \alpha_i - r) \frac{\beta}{\beta_i} L} \quad \text{wobei } \beta := \min_j \beta_j.$$

Dabei ist $C(u)$ unabhängig von L . Die gleichen Abschätzungen gelten für eine Diskretisierung mit AFD2, da es für Level-adaptive Indexmengen keinen Unterschied zwischen $MV(\text{AFD}, \cdot)$ und $MV(\text{AFD2}, \cdot)$ gibt.

Bevor wir dieses Theorem beweisen, soll auf einige Folgerungen eingegangen werden.

Bemerkung 6.2 Für sehr glatte Funktionen, $\alpha_i - r \geq \min(N, K)$ und $\alpha_j \geq N$, ist $\beta_1 = \dots = \beta_d = \beta = N$, so daß die Konsistenzordnung gleich $\min(N, K)$ ist. D.h. es lohnt sich im allgemeinen nicht, ein univariates Finite-Differenzen-Verfahren höherer Ordnung als N zu benutzen. Die gegenüber Verfahren der Ordnung N zusätzlichen Kosten durch einen breiteren Differenzen-Stern schlagen sich nicht in einer höheren Konsistenzordnung nieder. Umgekehrt ist auch $K < N$ nicht optimal, da hier die höheren Kosten für die glatteren

Ansatzfunktionen keinen Gewinn in der Konsistenzordnung bringen. In diesem Sinne ist die Kollokationsmethode nicht optimal, da hier $K = N - r$ ist. In [Sch98a] ist behauptet worden, daß man auch für $K > N$ eine Gesamtkonsistenzordnung von K erhalten würde (Satz 4.13). Dies ist nicht richtig. Im Beweis sind bei den Abschätzungen die Terme mit $m = 1$ (Seite 217 unten) übersehen worden.

Bemerkung 6.3 Die Abschätzungen in Theorem 6.1 enthalten einen logarithmischen Term, der in der früheren Arbeit [Sch98a] fehlt. Tatsächlich ist in dem dortigen Beweis übersehen worden, daß aus der Rekursionsformel $\mathcal{F}(\mathbf{x}_{1,i}) = \frac{1}{2}(\mathcal{F}(\mathcal{H}_2^-(\mathbf{x}_{1,i})) + \mathcal{F}(\mathcal{H}_2^+(\mathbf{x}_{1,i}))) + O(h^2)$ nicht $\mathcal{F}(\mathbf{x}_{1,i}) = O(h^2)$, sondern $\mathcal{F}(\mathbf{x}_{1,i}) = O(l_1 h^2)$ folgt.

Beweis: (Theorem 6.1) Wir zeigen die Behauptung nur für $i = 1$. Da u mindestens stetig ist, kann u in die unendliche Multiskalendarstellung

$$u = \sum_{\mathbf{l}} \sum_{\mathbf{t} \in T(\mathbf{l})} u_{(\mathbf{l}, \mathbf{t})} \psi_{(\mathbf{l}, \mathbf{t})} \quad (6.8)$$

entwickelt werden. In einem ersten Schritt soll eine für den Beweis geeignete Darstellung des Ergebnisses der diskreten Operatorauswertung gefunden werden. Entsprechend des Vorgehens in $MV(\dots)$ zerlegen wir \mathcal{T} in Indexmengen der Form

$$\mathcal{T}_{\bar{\mathbf{l}}, \bar{\mathbf{t}}} := \{(l_1, t_1) : ((l_1, \bar{\mathbf{l}}), (t_1, \bar{\mathbf{t}})) \in \mathcal{T}\}.$$

Dabei sind $\bar{\mathbf{l}}$ und $\bar{\mathbf{t}}$ $(d-1)$ -dimensionale Teilindizes der Form $\bar{\mathbf{l}} = (l_2, \dots, l_d)$ bzw. $\bar{\mathbf{t}} = (t_2, \dots, t_d)$. Die Menge aller $\bar{\mathbf{l}}, \bar{\mathbf{t}}$, für die $\mathcal{T}_{\bar{\mathbf{l}}, \bar{\mathbf{t}}} \neq \emptyset$ ist, ist

$$\mathcal{T}_1 := \{(\bar{\mathbf{l}}, \bar{\mathbf{t}}) : \exists l_1, t_1 : (\mathbf{l}, \mathbf{t}) \in \mathcal{T}\}.$$

In der Definition von \mathcal{T}_1 haben wir implizit angenommen, daß z.B. $\mathbf{l} = (l_1, \bar{\mathbf{l}})$ ist. Von dieser Kurzschreibweise werden wir nachfolgend öfter Gebrauch machen. Für Indizes $(\bar{\mathbf{l}}, \bar{\mathbf{t}})$ definieren wir weiterhin

$$w_{\bar{\mathbf{l}}, \bar{\mathbf{t}}}(x) := \sum_{l_1 \geq l_0} \sum_{t_1 \in T(l_1)} u_{(\mathbf{l}, \mathbf{t})} \psi_{(l_1, t_1)}(x)$$

als die Teilfunktionen, auf die letztlich AFD wirkt. Konkret werden dabei natürlich nur die Koeffizienten $\mathbf{w}_{\mathcal{T}_{\bar{\mathbf{l}}, \bar{\mathbf{t}}}} := \{u_{(\mathbf{l}, \mathbf{t})}\}_{(l_1, t_1) \in \mathcal{T}_{\bar{\mathbf{l}}, \bar{\mathbf{t}}}}$ benutzt. Mit diesen Bezeichnungen ist das Resultat der diskreten Operatorauswertung gleich

$$v(\mathbf{x}) := \Psi_{\mathcal{T}}(\mathbf{x}) \cdot MV(AFD, 1, \mathbf{u}_{\mathcal{T}}) = \sum_{(\bar{\mathbf{l}}, \bar{\mathbf{t}}) \in \mathcal{T}_1} \psi_{(\bar{\mathbf{l}}, \bar{\mathbf{t}})}(\bar{\mathbf{x}}) \left[\Psi_{\mathcal{T}_{\bar{\mathbf{l}}, \bar{\mathbf{t}}}}(x_1) \cdot AFD(\mathbf{w}_{\mathcal{T}_{\bar{\mathbf{l}}, \bar{\mathbf{t}}}}) \right].$$

Konform zu unserer bisherigen Notation ist $\psi_{(\bar{\mathbf{l}}, \bar{\mathbf{t}})}(\bar{\mathbf{x}}) = \prod_{j=2}^d \psi_{(l_j, t_j)}(x_j)$. Wir betrachten nun den exakten Term $D^{r\mathbf{e}_1}u$. Da \mathcal{T} die multivariate Kegelbedingung erfüllt, ist speziell für $\mathbf{x} \in \Omega_{\mathcal{T}}$ der Wert $D^{r\mathbf{e}_1}u(\mathbf{x})$ gleich dem Wert des Interpolanten $I_{\mathcal{T}}D^{r\mathbf{e}_1}u$ in \mathbf{x} , d.h.

$$D^{r\mathbf{e}_1}u(\mathbf{x}) = \sum_{\bar{\mathbf{l}}, \bar{\mathbf{t}} \in \mathcal{T}_1} \sum_{(l_1, t_1) \in \mathcal{T}_{\bar{\mathbf{l}}, \bar{\mathbf{t}}}} \psi_{(\mathbf{l}, \mathbf{t})}(\mathbf{x}) \langle \tilde{\psi}_{(\mathbf{l}, \mathbf{t})}, D^{r\mathbf{e}_1} \sum_{\mathbf{k}, \mathbf{s}} u_{(\mathbf{k}, \mathbf{s})} \psi_{(\mathbf{k}, \mathbf{s})} \rangle.$$

Wegen der Tensorproduktstruktur und der Biorthogonalität der primalen und dualen Basisfunktionen erhalten wir wie im Beweis von Lemma 6.1

$$D^{r\mathbf{e}_1}u(\mathbf{x}) = \sum_{\bar{\mathbf{l}}, \bar{\mathbf{t}} \in \mathcal{T}_1} \psi_{(\bar{\mathbf{l}}, \bar{\mathbf{t}})}(\bar{\mathbf{x}}) \sum_{(l_1, t_1) \in \mathcal{T}_{\bar{\mathbf{l}}, \bar{\mathbf{t}}}} \psi_{(l_1, t_1)}(x_1) \langle \tilde{\psi}_{(l_1, t_1)}, D^r \sum_{k_1, s_1} u_{((k_1, \bar{\mathbf{l}}), (s_1, \bar{\mathbf{t}}))} \psi_{(k_1, s_1)} \rangle.$$

Die Summe über k_1, s_1 ist genau $w_{\bar{\mathbf{l}}, \bar{\mathbf{t}}}$. Somit entspricht die Summe über (l_1, t_1) gerade dem univariaten Interpolanten $I_{\mathcal{T}_{\bar{\mathbf{l}}, \bar{\mathbf{t}}}} D^r w_{\bar{\mathbf{l}}, \bar{\mathbf{t}}}$ in x_1 . Da \mathcal{T} die multivariate Kegelbedingung erfüllt, erfüllt auch $\mathcal{T}_{\bar{\mathbf{l}}, \bar{\mathbf{t}}}$ die Kegelbedingung und außerdem ist $x_1 \in \Omega_{\mathcal{T}_{\bar{\mathbf{l}}, \bar{\mathbf{t}}}}$, so daß $I_{\mathcal{T}_{\bar{\mathbf{l}}, \bar{\mathbf{t}}}} D^r w_{\bar{\mathbf{l}}, \bar{\mathbf{t}}}(x_1) = D^r w_{\bar{\mathbf{l}}, \bar{\mathbf{t}}}(x_1)$ folgt. Also ist für $\mathbf{x} \in \Omega_{\mathcal{T}}$

$$D^{re_1} u(\mathbf{x}) = \sum_{\bar{\mathbf{l}}, \bar{\mathbf{t}} \in \mathcal{T}} \psi_{(\bar{\mathbf{l}}, \bar{\mathbf{t}})}(\bar{\mathbf{x}}) D^r w_{\bar{\mathbf{l}}, \bar{\mathbf{t}}}(x_1) .$$

Damit kann der Konsistenzfehler in \mathbf{x} durch

$$|(D^{re_1} u - v)(\mathbf{x})| \leq \sum_{(\bar{\mathbf{l}}, \bar{\mathbf{t}}) \in \mathcal{T}} |\psi_{(\bar{\mathbf{l}}, \bar{\mathbf{t}})}(\bar{\mathbf{x}})| \left| D^r w_{\bar{\mathbf{l}}, \bar{\mathbf{t}}}(x_1) - \Psi_{\mathcal{T}_{\bar{\mathbf{l}}, \bar{\mathbf{t}}}}(x_1) \cdot AFD(\mathbf{w}_{\mathcal{T}_{\bar{\mathbf{l}}, \bar{\mathbf{t}}}}) \right| . \quad (6.9)$$

abgeschätzt werden. Nun gilt aber für $\mathbf{x} \in \Omega_{\mathcal{T}}$:

$$\sum_{(\bar{\mathbf{l}}, \bar{\mathbf{t}}) \in \mathcal{T}} |\psi_{(\bar{\mathbf{l}}, \bar{\mathbf{t}})}(\bar{\mathbf{x}})| \lesssim L^{d-1} . \quad (6.10)$$

Der Grund dafür ist, daß es für einen beliebigen Punkt $x \in [0, 1]$ auf jedem Level l nur $O(1)$ Indizes t mit $\psi_{(l,t)}(x) \neq 0$ gibt. Da in unserem Fall die Zahl der Level für jede Koordinatenrichtung durch L beschränkt ist, gibt es maximal $O(L^{d-1})$ Indizes $\bar{\mathbf{l}}, \bar{\mathbf{t}}$, für die $\psi_{(\bar{\mathbf{l}}, \bar{\mathbf{t}})}(\bar{\mathbf{x}}) \neq 0$ sein kann. Die einzelnen Basisfunktionen haben eine Maximum-Norm von 1, so daß (6.10) folgt.

Es ist also nur noch zu zeigen, daß der (univariate) Konsistenzfehler $|D^r w_{\bar{\mathbf{l}}, \bar{\mathbf{t}}}(x_1) - \Psi_{\mathcal{T}_{\bar{\mathbf{l}}, \bar{\mathbf{t}}}}(x_1) \cdot AFD(\mathbf{w}_{\mathcal{T}_{\bar{\mathbf{l}}, \bar{\mathbf{t}}}})|$ gleichmäßig in $\bar{\mathbf{l}}, \bar{\mathbf{t}}$ durch $2^{-\min(N, K, \alpha_1 - r)\beta/\beta_1 L}$ beschränkt ist. Zur Vereinfachung der Notation lassen wir hier die tiefgestellten Indizes $\bar{\mathbf{l}}, \bar{\mathbf{t}}$ in $w_{\bar{\mathbf{l}}, \bar{\mathbf{t}}}$ bzw. $\mathcal{T}_{\bar{\mathbf{l}}, \bar{\mathbf{t}}}$ weg und schreiben einfach nur w, \mathcal{T} oder $\mathbf{w}_{\mathcal{T}}$. Da \mathcal{T} eine Level-adaptive Indexmenge ist, enthält \mathcal{T} alle Indizes (l, t) , $t \in T(l)$, bis zu einem bestimmten Level l_{max} , das von $\bar{\mathbf{l}}$ abhängt,

$$l_{max} := \frac{1}{\beta_1} [\beta L - \sum_{j=2}^d l_j \beta_j] .$$

D.h. \mathcal{T} ist nicht adaptiv und $\Omega_{\mathcal{T}}$ ist ein gewöhnliches, äquidistantes Gitter mit Maschenweite $2^{-l_{max}}$. Damit entspricht AFD im wesentlichen einer inversen Multiskalenttransformation, der Auswertung des Finite-Differenzen-Schemas auf $\Omega_{\mathcal{T}}$ und einer Multiskalenttransformation. Mit den klassischen Aussagen zum Konsistenzfehler für Finite-Differenzen-Verfahren, siehe z.B. [Kos00], folgt somit

$$|D^r w(x_1) - \Psi_{\mathcal{T}}(x_1) \cdot AFD(\mathbf{w}_{\mathcal{T}})| \lesssim 2^{-l_{max} \gamma} |w|_{\gamma+r} , \quad \gamma := \min(K, \alpha_1 - r) .$$

Vollkommen analog wie Lemma 6.5 kann man nun für $u \in \mathcal{C}^{\alpha} \subset \mathcal{C}^{(\gamma+r, \beta_2, \dots, \beta_d)}$ zeigen, daß

$$|w|_{\gamma+r} \lesssim |u|_{(\gamma+r, \bar{\boldsymbol{\beta}})} \cdot \prod_{j=2}^d 2^{-l_j \beta_j} = |u|_{(\gamma+r, \bar{\boldsymbol{\beta}})} \cdot 2^{-\bar{\mathbf{l}} \cdot \bar{\boldsymbol{\beta}}} .$$

Dabei haben wir der Einfachheit halber $\bar{\boldsymbol{\beta}} = (\beta_2, \dots, \beta_d)$ gesetzt sowie die übliche Definition des Skalarproduktes $\bar{\mathbf{l}} \cdot \bar{\boldsymbol{\beta}} = \sum_{j=2}^d l_j \beta_j$ benutzt. Somit ist

$$|D^r w(x_1) - \Psi_{\mathcal{T}}(x_1) \cdot AFD(\mathbf{w}_{\mathcal{T}})| \lesssim 2^{-(\gamma l_{max} + \bar{\mathbf{l}} \cdot \bar{\boldsymbol{\beta}})} |u|_{(\gamma+r, \bar{\boldsymbol{\beta}})} .$$

Wir müssen nur noch zeigen, daß $\gamma l_{max} + \bar{\mathbf{l}} \cdot \bar{\boldsymbol{\beta}} \geq \min(N, K, \alpha_1 - r)\beta/\beta_1 L$ ist. Wir setzen $q = \gamma/\beta_1$, dann ist

$$\gamma l_{max} + \bar{\mathbf{l}} \cdot \bar{\boldsymbol{\beta}} = \frac{\gamma}{\beta_1}(\beta L - \bar{\mathbf{l}} \cdot \bar{\boldsymbol{\beta}}) + \bar{\mathbf{l}} \cdot \bar{\boldsymbol{\beta}} = q\beta L + \bar{\mathbf{l}} \cdot \bar{\boldsymbol{\beta}}(1 - q).$$

Falls $q \leq 1$, dann ist der obige Ausdruck $\geq q\beta L$. Falls $q > 1$, dann wird der obige Ausdruck minimal, falls $\bar{\mathbf{l}} \cdot \bar{\boldsymbol{\beta}}$ maximal wird. Wegen (6.7) ist jedoch $\bar{\mathbf{l}} \cdot \bar{\boldsymbol{\beta}} \leq \beta L - \beta_1 l_0$, so daß

$$q\beta L + \bar{\mathbf{l}} \cdot \bar{\boldsymbol{\beta}}(1 - q) \geq q\beta L + (1 - q)(\beta L - \beta_1 l_0) = \beta L + (q - 1)\beta_1 l_0 \geq \beta L$$

Sowohl für $q \geq 1$ als auch für $q < 1$ ist damit $\gamma l_{max} + \bar{\mathbf{l}} \cdot \bar{\boldsymbol{\beta}} \geq \min(1, q)\beta L$. Da

$$\min(1, q) = \min\left(1, \frac{\min(K, \alpha_1 - r)}{\beta_1}\right) = \frac{\min(N, K, \alpha_1 - r)}{\beta_1},$$

folgt somit die Behauptung. \square

Herzstück des Beweises für Theorem 6.1 ist, den Konsistenzfehler auf den Konsistenzfehler für die univariate Funktion $w_{\bar{\mathbf{l}}, \bar{\mathbf{t}}}$ zurückzuführen. Die für die Abschätzung dieses Konsistenzfehlers notwendigen Glattheitsannahmen konnten aus der a priori bekannten Glattheit $u \in \mathcal{C}^\alpha$ hergeleitet werden.

Auch für den adaptiven Fall ist es möglich, den Konsistenzfehler auf den Konsistenzfehler für univariate Funktionen $w_{\bar{\mathbf{l}}, \bar{\mathbf{t}}}$ abzuwälzen. Entscheidend dafür ist, daß die Darstellung des Konsistenzfehlers (6.9) und die Abschätzung (6.10) für beliebige Indexmengen \mathcal{T} gelten, die die multivariate Kegelbedingung erfüllen. Für die Abschätzungen des Konsistenzfehlers für $w_{\bar{\mathbf{l}}, \bar{\mathbf{t}}}$ können wir ausnutzen, daß die Approximation von $D^r w_{\bar{\mathbf{l}}, \bar{\mathbf{t}}}$ in AFD durch ein normales Finite-Differenzen-Verfahren, allerdings bezüglich lokal variabler Maschenweiten, realisiert wird. Dabei geht dann aber die lokale Glattheit von u ein. Idealerweise ist nun die Indexmenge \mathcal{T} so definiert, daß in Bereichen, wo die lokalen Ableitungen höherer Ordnung sehr groß sind, sehr kleine Maschenweiten für den lokalen Finite-Differenzen-Stern entstehen, so daß trotz geringer Glattheit der punktweise Konsistenzfehler klein bleibt. Ein erster Prototyp für eine entsprechende Aussage zum Konsistenzfehler auf angepaßten Indexmengen ist das folgende Theorem. Um den Kern der Aussage nicht unter einer Vielzahl von Parametern zu verschütten, nehmen wir dabei an, daß u formal aus \mathcal{C}^{N1} ist.

Theorem 6.2 (Konsistenzfehler für adaptive Finite-Differenzen-Diskretisierung)

Sei $u \in \mathcal{C}^{N1}$. Das univariate Finite-Differenzen-Schema in AFD bzw. AFD2 zur Berechnung der r -ten Ableitung habe die Konsistenzordnung $N - r$ und erfüllt die Bedingung

- (i) Die Aufsetzpunkte des Finite-Differenzen-Sterns auf Level l für einen Punkt $2^{-l}s$, $s \in S(l)$, $l_0 \leq l \leq L$ sind im Träger eines Interpolants $\psi_{(l, \mathbf{t})}$ enthalten.

Für ein beliebiges festes $\delta > 0$ sei \mathcal{T} durch die folgende Bedingung definiert:

- (ii) \mathcal{T} enthält alle Indizes (\mathbf{l}, \mathbf{t}) , für die

$$2^{-\|\mathbf{l}\|_1 N + l_i r} \cdot \sup_{\mathbf{x} \in \text{supp}\psi_{(\mathbf{l}, \mathbf{t})}} |D^{N1} u(\mathbf{x})| > \delta \quad (6.11)$$

sowie die Söhne solcher Indizes bezüglich der i -ten Koordinatenrichtung. D.h. auch die Indizes $(\mathbf{l} + \mathbf{e}_i, \mathbf{s})$ mit $s_j = t_j$ für $j \neq i$, sowie $\text{supp}\psi_{(l_i+1, s_i)} \cap \text{supp}\psi_{(l_i, t_i)} \neq \emptyset$. Außerdem soll \mathcal{T} die multivariate Kegelbedingung erfüllen.

Mit L sei wieder das maximale Level von \mathcal{T} bezeichnet. Dann gilt

$$\|D^{re_i}u - \Psi_{\mathcal{T}} \cdot MV(AFD, i, \mathbf{u}_{\mathcal{T}})\|_{\Omega_{\mathcal{T}}} \lesssim L^{d-1}\delta .$$

Die gleichen Abschätzungen gelten für $MV(AFD2, \dots)$.

Einige Bemerkungen sind an dieser Stelle angebracht:

Bemerkung 6.4 Bedingung (i) ist für Interpolets und die üblichen zentralen Finite-Differenzen-Schemen mit Konsistenzordnung $N - r$ erfüllt.

Bemerkung 6.5 Es stellt sich die Frage, welche Beziehung zwischen den Indexmengen \mathcal{T} aus Theorem 6.2 und den Indexmengen für die Best- n -Term Approximation von u besteht. Das Problem ist hier, daß \mathcal{T} in Bezug auf die Approximation von $D^{re_i}u$ gewählt wurde, d.h. der Konsistenzfehler soll klein sein. Dagegen würde man bei der Best- n -Term-Approximation die Indexmenge so definieren, daß z.B. der Interpolationsfehler möglichst klein wird. In diesem Sinne sind die Indexmengen aus Theorem 6.2 nicht optimal. Für die Approximation von Funktionen $u \in \mathcal{C}^{N1}$ mit kleiner Semi-Norm $|u|_{N1}$ bis zur Genauigkeit $\sim 2^{-LN}$ würde man beispielsweise die Level-adaptiven Indexmengen $\{(\mathbf{l}, \mathbf{t}) : |\mathbf{l}|_1 \leq L\}$ wählen. Für den Konsistenzfehler mit Genauigkeit $\sim 2^{-LN}$ hat die Indexmenge aber die Form $\mathcal{T} = \{(\mathbf{l}, \mathbf{t}) : |\mathbf{l}|_1 - l_i r / N \leq L\}$. \mathcal{T} wäre also deutlich größer. Das Problem stellt hier der $2^{l_i r}$ -Anteil im Schwellwert-Ansatz (6.11) dar, der durch den Operator verursacht wird, jedoch nicht von der Norm $\|\cdot\|_{\Omega_{\mathcal{T}}}$ kompensiert wird.

Beweis: (Theorem 6.2) Wir zeigen die Aussage wieder nur für $i = 1$ und gehen wie im Beweis für Theorem 6.1 vor, einschließlich der vereinfachten Bezeichnungen. Wie dort erhält man für $\mathbf{x} \in \Omega_{\mathcal{T}}$ die Abschätzungen

$$|(D^{re_1}u - \Psi_{\mathcal{T}} \cdot MV(AFD, i, \mathbf{u}_{\mathcal{T}}))(\mathbf{x})| \leq \sum_{(\bar{\mathbf{l}}, \bar{\mathbf{t}}) \in \mathfrak{T}} |\psi_{(\bar{\mathbf{l}}, \bar{\mathbf{t}})}(\bar{\mathbf{x}})| \left| D^r w(x_1) - \Psi_{\mathcal{T}}(x_1) \cdot AFD(\mathbf{w}_{\mathcal{T}}) \right| ,$$

sowie (6.10). Innerhalb von AFD bzw. AFD2 werden aus den Koeffizienten $\mathbf{w}_{\mathcal{T}}$ die Knotenwerte $\mathbf{w}^{S(l)}$ bzw. $\mathbf{w}^{S^f(l)}$ auf den Levels $l = l_0, \dots, L$ berechnet, mit denen dann die Finite-Differenzen-Approximationen $\mathbf{v}^{S(l)}$ bzw. $\mathbf{v}^{S^e(l)}$ von $D^r w$ berechnet werden. Die Δ -Korrektur im anschließenden Aufruf von $AWT(\mathbf{v}^S, \mathcal{S}, \Delta)$ bzw. $AWT(\mathbf{v}^{S^e}, \mathcal{S}^e, \Delta^e)$ sorgt dann dafür, daß $\Psi_{\mathcal{T}}(x_1) \cdot AFD(\mathbf{w}_{\mathcal{T}})$ gleich dem Wert $v^{(l,s)}$ vom höchsten zur Verfügung stehenden Level ist. D.h. (l, s) ist durch

$$2^{-l}s = x_1 \quad \text{und} \quad s \in \mathcal{S}(l) \setminus \Delta(l) \quad \text{bzw.} \quad s \in \mathcal{S}^e(l) \setminus \Delta^e(l) \quad (6.12)$$

charakterisiert. Wenn wir mit Σ das kleinste Intervall bezeichnen, in dem alle Aufsetzpunkte des Finite-Differenzen-Sterns für $v^{(l,s)}$ liegen, dann folgt aus der Konsistenz des univariaten Finite-Differenzen-Schemas

$$|D^r w(x_1) - \Psi_{\mathcal{T}}(x_1) \cdot AFD(\mathbf{w})| \lesssim 2^{-l(N-r)} \sup_{x \in \Sigma} |D^N w(x)| . \quad (6.13)$$

Nach Voraussetzung (i) ist Σ im Träger mindestens eines Interpolets $\psi_{(l,t)}$ enthalten. Analog Lemma 6.5 folgt damit

$$\sup_{x \in \Sigma} |D^N w(x)| \lesssim 2^{-|\bar{\mathbf{l}}|_1 N} \sup_{\mathbf{y} \in \text{supp} \psi_{(\mathbf{l}, \mathbf{t})}} |D^{N1} u(\mathbf{y})| \quad \text{wobei} \quad \mathbf{l} := (l, \bar{\mathbf{l}}), \quad \mathbf{t} := (t, \bar{\mathbf{t}}) .$$

Aus Bedingung (ii) können wir eine Abschätzung für den Term auf der rechten Seite folgern. Angenommen, $2^{-|\mathbf{l}|_1 N} 2^{lr} \sup\{|D^{N\mathbf{1}}u(\mathbf{y})| : \mathbf{y} \in \text{supp}\psi_{(\mathbf{l}, \mathbf{t})}\}$ wäre größer gleich δ . Dann wäre $(\mathbf{l}, \mathbf{t}) \in \mathcal{T}$ und somit $(l, t) \in \mathcal{T}$. Außerdem wären auch alle Söhne von (\mathbf{l}, \mathbf{t}) entlang der ersten Koordinatenrichtung in \mathcal{T} , so daß auch alle Söhne von (l, t) in \mathcal{T} wären. Daraus würde aber folgen, daß $x_1 \in 2^{-(l+1)}\mathcal{S}(l+1)$ ist. Somit ist $2s \in \mathcal{S}(l+1)$ und somit $s \in \Delta(l)$ bzw. $2s \in \mathcal{S}^e(l+1)$ und $s \in \Delta^e(l)$, was ein Widerspruch zur Definition von (l, s) in (6.12) ist. Also ist

$$\begin{aligned} \sup_{x \in \Sigma} |D^N w(x)| &\lesssim 2^{-|\mathbf{l}|_1 N} \sup_{\mathbf{y} \in \text{supp}\psi_{(\mathbf{l}, \mathbf{t})}} |D^{N\mathbf{1}}u(\mathbf{y})| \\ &\leq 2^{-|\mathbf{l}|_1 N} \delta 2^{-lr} 2^{|\mathbf{l}|_1 N} = \delta 2^{l(N-r)}. \end{aligned}$$

Daraus folgt mit (6.13) $|D^r w(x_1) - \Psi_{\mathcal{T}}(x_1) \cdot AFD(\mathbf{w})| \lesssim \delta$. Insgesamt erhalten wir wie im Beweis für Theorem 6.1 $|(D^{r\mathbf{e}_1}u - v)(\mathbf{x})| \lesssim \delta L^{d-1}$, und damit die Behauptung. \square

Wie wir bereits in Bemerkung 6.21 erwähnt hatten, sind die Forderungen an die Indexmenge \mathcal{T} sehr stark. Die Ursache dafür ist, daß der Konsistenzfehler in der starken $\|\cdot\|_{\Omega_{\mathcal{T}}}$ -Norm gemessen wird. Dies führt zu dem 2^{lr} -Anteil des Operators in der Definition von \mathcal{T} , siehe (6.11). Die Idee ist nun, dieses Problem dadurch zu lösen, daß man schwächere Normen für den Fehler betrachtet. Entsprechend Ungleichung (6.6) ist dafür dann aber neben dem Konsistenzfehler $\|I_{\mathcal{T}}Au - A_{\mathcal{T}}I_{\mathcal{T}}u\|_{\mathcal{H}^{s-r}}$ auch der Interpolationsfehler $\|u - I_{\mathcal{T}}u\|_{\mathcal{H}^s}$ zu berücksichtigen. Eine Möglichkeit ist, mittels eines Schwellwert-Ansatzes Indexmengen \mathcal{T} zu definieren, die in Bezug auf den Interpolationsfehler gewählt wurden. Entsprechend Lemma 4.2 bzw. Lemma 6.5 bieten sich dabei für z.B. $\|u - I_{\mathcal{T}}u\|_{\mathcal{L}^2}$ die Indexmengen

$$\mathcal{T} = \{(\mathbf{l}, \mathbf{t}) : |u_{(\mathbf{l}, \mathbf{t})}| 2^{-|\mathbf{l}|_1/2} > \delta\} \quad \text{bzw.} \quad \mathcal{T} = \{(\mathbf{l}, \mathbf{t}) : 2^{-|\mathbf{l}|_1(N+1/2)} \cdot \sup_{\mathbf{x} \in \text{supp}\psi_{(\mathbf{l}, \mathbf{t})}} |D^{N\mathbf{1}}u(\mathbf{x})| > \delta\} \quad (6.14)$$

an. Der resultierende Interpolationsfehler muß dann in Abhängigkeit von δ mittels Aussagen über die Besov-Regularität von u abgeschätzt werden. Für isotrope Wavelets gibt es dazu eine Reihe von Arbeiten [CDDD99, Don92]. Für anisotrope Wavelets scheint es solche Untersuchungen noch nicht zu geben. Sinnvoll wäre hier, den Zusammenhang zwischen den anisotropen Besov-Räumen aus Kapitel 5 und Schwellwert-Ansätzen zu betrachten.

In dieser Arbeit wollen wir jedoch nur den Konsistenzfehler untersuchen. Dieser kann ähnlich wie in Theorem 6.2 allein über die in der Definition von \mathcal{T} enthaltenen lokalen Glattheitsaussagen bestimmt werden. Ein generelles Problem dabei ist, daß wir zwar relativ einfache Abschätzungen für die Multiskalenkoeffizienten $e_{(\mathbf{l}, \mathbf{t})}$ des Konsistenzfehlers $e := I_{\mathcal{T}}D^{r\mathbf{e}_1}u - \Psi_{\mathcal{T}} \cdot MV(AFD, i, \mathbf{u}_{\mathcal{T}}) = \Psi_{\mathcal{T}} \cdot \mathbf{e}_{\mathcal{T}}$ bzgl. der Interpolet-Basis erhalten können, jedoch erlauben die Interpolets keine einseitigen Norm-Abschätzungen für negative Sobolev-Normen, siehe Lemma 4.2. Was man somit eigentlich benötigen würde, wäre eine Darstellung von e mittels der stabilen Lifting-Wavelets, zusammen mit Abschätzungen für die entsprechenden Lifting-Waveletkoeffizienten. In diese Richtung konnten noch keine Fortschritte erzielt werden. Um zumindest einen ersten Eindruck davon zu bekommen, wie man Abschätzungen des Konsistenzfehlers erhält, betrachten wir in den nachfolgenden zwei Theoremen den Konsistenzfehler für die Finite-Differenzen- bzw. Kollokationsmethode in der \mathcal{L}^2 -Norm. Der Einfachheit halber betrachten wir auch nicht die Indexmengen aus (6.14), sondern die aus Theorem 6.2.

Theorem 6.3 (Konsistenzfehler in der $\|\cdot\|_{\mathcal{L}^2}$ -Norm für Finite-Differenzen)

Unter den gleichen Voraussetzungen an u , das Finite-Differenzen-Schema sowie \mathcal{T} aus Theorem 6.2 gilt für beliebiges $\delta > 0$

$$\|I_{\mathcal{T}}D^{r\mathbf{e}_i}u - \Psi_{\mathcal{T}} \cdot MV(AFD, i, \mathbf{u}_{\mathcal{T}})\|_{\mathcal{L}^2} \lesssim L^d \delta ,$$

bzw. die gleichen Abschätzung für $MV(AFD2, \dots)$.

Beweis: Wir folgen dem Beweis von Theorem 6.2 und benutzen auch wieder die gleichen Bezeichnungen. Der Konsistenzfehler ist $e = I_{\mathcal{T}}D^{r\mathbf{e}_i}u - \Psi_{\mathcal{T}} \cdot MV(AFD, 1, \mathbf{u}_{\mathcal{T}})$ mit der Multiskalendarstellung $\Psi_{\mathcal{T}} \cdot \mathbf{e}_{\mathcal{T}}$. Wegen der einseitigen Norm-Abschätzung aus Lemma 4.2 für Interpolets gilt

$$\|e\|_{\mathcal{L}^2}^2 \lesssim L^d \sum_{(\mathbf{l}, \mathbf{t}) \in \mathcal{T}} |e_{(\mathbf{l}, \mathbf{t})}|^2 2^{-|\mathbf{l}|_1} . \quad (6.15)$$

Die einzelnen Koeffizienten des Fehlers sind Linearkombinationen von punktweisen Werten des (univariaten) Konsistenzfehlers, konkret

$$e_{(\mathbf{l}, \mathbf{t})} = \sum_{s \in \mathcal{S}(l_1)} \tilde{G}_{s, t_1}^{l_1} \left(D^r w(2^{-l_1} s) - \Psi_{\mathcal{T}}(2^{-l_1} s) \cdot AFD(\mathbf{w}_{\mathcal{T}}) \right) .$$

Da es aber unabhängig von \mathbf{l} oder \mathbf{t} nur maximal $2N$ Indizes s mit $\tilde{G}_{s, t_1}^{l_1} \neq 0$ gibt, und diese Werte auch gleichmäßig in l_1 beschränkt sind, folgt

$$|e_{(\mathbf{l}, \mathbf{t})}| \lesssim \|D^r w - \Psi_{\mathcal{T}} \cdot AFD(\mathbf{w}_{\mathcal{T}})\|_{\Omega_{\mathcal{T}}} \lesssim \delta \quad (\text{siehe Beweis von Theorem 6.2}) .$$

Daraus folgt $\|e\|_{\mathcal{L}^2} \lesssim L^d \sum_{(\mathbf{l}, \mathbf{t}) \in \mathcal{T}} \delta^2 2^{-|\mathbf{l}|_1} \lesssim L^{2d} \delta^2$, also die Behauptung. \square

Ähnlich wie bei den Finite-Differenzen-Verfahren kann auch für die Kollokationsmethode der Konsistenzfehler durch geeignete Indexmengen kontrolliert werden. Für die Kollokationsmethode ist der Fehler gleich

$$\begin{aligned} e &= I_{\mathcal{T}}D^{r\mathbf{e}_i}(u - I_{\mathcal{T}}u) \\ &= \sum_{(\mathbf{l}, \mathbf{t}) \in \mathcal{T}} \psi_{(\mathbf{l}, \mathbf{t})} \underbrace{\sum_{(\mathbf{k}, \mathbf{s}) \notin \mathcal{T}} \langle \tilde{\psi}_{(\mathbf{l}, \mathbf{t})}, D^{r\mathbf{e}_i} \psi_{(\mathbf{k}, \mathbf{s})} \rangle u_{(\mathbf{k}, \mathbf{s})}}_{=: e_{(\mathbf{l}, \mathbf{t})}} \end{aligned}$$

Die einzelnen Multiskalenkoeffizienten des Fehlers können mittels

$$|e_{(\mathbf{l}, \mathbf{t})}| \leq \sum_{(\mathbf{k}, \mathbf{s}) \notin \mathcal{T}} |\langle \tilde{\psi}_{(\mathbf{l}, \mathbf{t})}, D^{r\mathbf{e}_i} \psi_{(\mathbf{k}, \mathbf{s})} \rangle| |u_{(\mathbf{k}, \mathbf{s})}| \quad (6.16)$$

abgeschätzt werden. Darin gehen sowohl Abschätzungen für die Koeffizienten $\{u_{(\mathbf{k}, \mathbf{s})}\}_{(\mathbf{k}, \mathbf{s}) \notin \mathcal{T}}$ ein, die aber auf die Definition von \mathcal{T} abgewälzt werden können, sowie Abschätzungen für die Operatorkoeffizienten $\langle \tilde{\psi}_{(\mathbf{l}, \mathbf{t})}, A\psi_{(\mathbf{k}, \mathbf{s})} \rangle$. Da wir solche Abschätzungen für die Operatorkoeffizienten noch bei weiteren Gelegenheiten benötigen, betrachten wir in nachfolgendem Lemma eine sehr allgemeine Situation.

Lemma 6.6 (Operatorkoeffizienten) Seien $\{\eta_{(k,s)}\}$ bzw. $\{\tilde{\eta}_{(k,s)}\}$ univariate Multiskalenbasen, die nicht biorthogonal zueinander sein müssen. Die maximale Sobolev-Regularität von $\eta_{(k,s)}$ bzw. $\tilde{\eta}_{(k,s)}$ sei γ^* bzw. $\tilde{\gamma}^*$. Für ein p mit $p < \gamma^*$ und $-p < \tilde{\gamma}^*$ seien die Basisfunktionen zu $\|\eta_{(k,s)}\|_{\mathcal{H}^p} \sim 2^{k(p-1/2)}$ bzw. $\|\tilde{\eta}_{(k,s)}\|_{\mathcal{H}^{-p}} \sim 2^{k(-p+1/2)}$ normiert. Sei nun $r \geq 0$ mit $\gamma^* + \tilde{\gamma}^* > r$, dann gilt für hinreichend kleines $\epsilon > 0$

$$|\langle \tilde{\eta}_{(l,t)}, D^r \eta_{(k,s)} \rangle| \leq 2^{kr} \cdot \begin{cases} 0 & \text{falls } \text{supp} \tilde{\eta}_{(l,t)} \cap \text{supp} \eta_{(k,s)} = \emptyset \\ 2^{(l-k)/2} 2^{-(\gamma^* - \epsilon - r)(l-k)} & \text{falls } l \geq k \\ 2^{(l-k)/2} 2^{-(\tilde{\gamma}^* - \epsilon)(k-l)} & \text{falls } l < k \end{cases}.$$

Speziell für Interpolets und ihre dualen Wavelets gilt

$$|\langle \tilde{\psi}_{(l,t)}, D^r \psi_{(k,s)} \rangle| \lesssim 2^{kr} \cdot \begin{cases} 0 & \text{falls } \text{supp} \tilde{\eta}_{(l,t)} \cap \text{supp} \eta_{(k,s)} = \emptyset \\ 2^{-(\theta^* - \epsilon - r)(l-k)} & \text{falls } l \geq k \\ 1 & \text{falls } l < k \end{cases}.$$

wobei $\theta^* := \sup\{s : \phi \in \mathcal{C}^s\}$ größer als r sein muß.

Der Beweis ist in Anhang A.15 angegeben.

Theorem 6.4 (Konsistenzfehler in der $\|\cdot\|_{\mathcal{L}^2}$ -Norm für die Kollokationsmethode)

Sei $u \in \mathcal{C}^{N^1}$. Für beliebiges $\delta > 0$ sei \mathcal{T} wie in Theorem 6.2 definiert. Dann gilt

$$\|I_{\mathcal{T}} D^{r_{e_i}}(u - I_{\mathcal{T}} u)\|_{\mathcal{L}^2} \lesssim L^d \delta.$$

Das heißt wir erhalten unter gleichen Voraussetzungen an \mathcal{T} wie in Theorem 6.3 eine ähnliche Fehlerschranke wie bei der Finite-Differenzen-Technik.

Beweis: (Theorem 6.4) Wir gehen ähnlich wie im Beweis für Theorem 6.3 vor. Der Konsistenzfehler $\|e\|_{\mathcal{L}^2}$ kann wieder durch (6.15) abgeschätzt werden. Wegen der Biorthogonalität von Ansatz- und Testfunktionen, vergleiche Lemma 6.1, vereinfacht sich (6.16) zu

$$|e_{(l,t)}| \leq \sum_{(k_1, s_1) \notin \mathcal{T}_{\bar{l}, \bar{t}}} |\langle \tilde{\psi}_{(l_1, t_1)}, D^r \psi_{(k_1, s_1)} \rangle| |u_{(k_1, \bar{l}, s_1, \bar{t})}| \quad (6.17)$$

Es ist klar, daß in der obigen Summe nur über Indizes (k_1, s_1) aus der Menge

$$Q := \{(k_1, s_1) : (k_1, s_1) \notin \mathcal{T}_{\bar{l}, \bar{t}} \text{ und } \text{supp} \tilde{\psi}_{(l_1, t_1)} \cap \text{supp} \psi_{(k_1, s_1)} \neq \emptyset\}$$

summiert werden muß. Da $\tilde{\psi}_{(l_1, t_1)}$ eine Linearkombination von maximal $2N$ Dirac-Funktionen ist, ist für alle Level k_1 die Anzahl der Interpolets $\psi_{(k_1, s_1)}$ mit $\text{supp} \tilde{\psi}_{(l_1, t_1)} \cap \text{supp} \psi_{(k_1, s_1)} \neq \emptyset$ gleichmäßig durch ein $c < \infty$ beschränkt. Da weiterhin der Träger eines jeden Interpolets $\psi_{(k_1, s_1)}$, $k_1 > l_0$, im Träger mindestens eines Interpolets des darunterliegenden Levels enthalten ist, kann die Menge Q in maximal c kettenförmige, nicht notwendigerweise disjunkte Teilmengen S der folgenden Form zerlegt werden:

1. Für jedes Level k_1 enthält S maximal einen Index (k_1, s_1) .
2. Für $(k_1, s_1) \in S$ gibt es entweder $(k_1 - 1, q) \in S$ mit $\text{supp} \psi_{(k_1, s_1)} \subset \text{supp} \psi_{(k_1 - 1, q)}$ oder es gibt einen Index $(k_1 - 1, q) \in \mathcal{T}_{\bar{l}, \bar{t}}$ mit $\text{supp} \psi_{(k_1, s_1)} \subset \text{supp} \psi_{(k_1 - 1, q)}$. Diesen speziellen Index aus $\mathcal{T}_{\bar{l}, \bar{t}}$ nennen wir (k_0, s_0) .

Wir wollen nun den Fehlerbeitrag in (6.17) durch eine Teilmenge S abschätzen. Nach Konstruktion von S gilt für alle $(k_1, s_1) \in S$ $\text{supp}\psi_{(k_1, s_1)} \subset \text{supp}\psi_{(k_0, s_0)}$. Analog zu Lemma 6.5 erhält man damit die Abschätzung

$$\begin{aligned} |u_{(k_1\bar{l}, s_1\bar{t})}| &\lesssim 2^{-k_1 N} 2^{-|\bar{l}|_1 N} \sup_{\mathbf{x} \in \text{supp}\psi_{(k_0\bar{l}, s_0\bar{t})}} |D^{N\mathbf{1}}u(\mathbf{x})| \\ &\lesssim 2^{-k_1 N} 2^{-|\bar{l}|_1 N} \delta 2^{k_0 N + |\bar{l}|_1 N - k_0 r} \\ &= \delta 2^{-k_1 N + k_0 N - k_0 r}. \end{aligned}$$

Dabei haben wir wieder ausgenutzt, daß \mathcal{T} nicht nur die Indizes mit (6.11) enthält, sondern auch deren Söhne, für die der Term auf der linken Seite von (6.11) $\leq \delta$ sein muß. Nach Konstruktion von S muß aber $(k_0\bar{l}, s_0\bar{t})$ gerade so ein Sohn sein.

Letztlich erhält man die folgende Abschätzung für den Fehlerbeitrag durch S :

$$\begin{aligned} \sum_{(k_1, s_1) \in S} |\langle \tilde{\psi}_{(l_1, t_1)}, D^r \psi_{(k_1, s_1)} \rangle| |u_{(k_1\bar{l}, s_1\bar{t})}| &\lesssim \\ &\lesssim \delta 2^{k_0(N-r)} \left[\sum_{k_1 \geq \max(k_0, l_1)} 2^{-k_1 N} 2^{k_1 r} + \sum_{k_0 \leq k_1 < \max(k_0, l_1)} 2^{-k_1 N} 2^{l_1 r - \theta^*(l_1 - k_1)} \right] \\ &\lesssim \delta \end{aligned}$$

Für den letzten Schritt wurde $\theta^* > r$ benutzt. Also ist auch $|e_{(1, \mathbf{t})}| \lesssim \delta$. Daraus folgt wie im Beweis für Theorem 6.3 die Behauptung. \square

Zum Abschluß dieses Abschnittes soll noch folgendes Resultat erwähnt werden: Benutzt man für den Petrov-Galerkin-Ansatz Lifting-Interpolets $\bar{\Psi}_{\mathcal{T}}$ und ihre dualen Wavelets $\tilde{\Psi}_{\mathcal{T}}$ als Ansatz- bzw. Testfunktionen, dann erhält man wegen Lemma 6.1 eine Darstellung für die Multiskalenkoeffizienten des Konsistenzfehlers $e := \bar{I}_{\mathcal{T}} D^{r\mathbf{e}_i}(u - \bar{I}_{\mathcal{T}}u)$, wobei $\bar{I}_{\mathcal{T}}v := \langle \tilde{\Psi}_{\mathcal{T}}, v \rangle \cdot \bar{\Psi}_{\mathcal{T}}$, die analog zu (6.17) ist

$$\bar{e}_{(1, \mathbf{t})} = \sum_{(k_1, s_1) \notin \mathcal{T}_{\bar{l}, \bar{t}}} \langle \tilde{\psi}_{(l_1, t_1)}, D^r \bar{\psi}_{(k_1, s_1)} \rangle \bar{u}_{(k_1\bar{l}, s_1\bar{t})}.$$

Da die dualen Wavelets nun einigermaßen glatt sind, erhält man für die Operatorkoeffizienten jedoch bessere Abschätzungen als bei der Kollokationsmethode. Man hat nun einen Abfall der Operatorkoeffizienten auch für $l \leq k$, vergleiche Lemma 6.6, so daß die Operatormatrix sehr viel stärker einer Diagonalmatrix ähnelt, als bei der Kollokationsmethode. Es kann dann analog wie im Beweis zu Theorem 6.4 gezeigt werden, daß bereits für die Indexmenge

$$\mathcal{T} = \left\{ (1, \mathbf{t}) : 2^{-|\bar{l}|_1 N} \cdot \sup_{\mathbf{x} \in \text{supp}\psi_{(1, \mathbf{t})}} |D^{N\mathbf{1}}u(\mathbf{x})| > \delta \right\} \cup \text{Söhne} \quad (6.18)$$

der Konsistenzfehler $\|e\|_{\mathcal{H}^{-r}} \lesssim \delta$ ist. Somit gelingt es hier, durch die Norm $(\|\cdot\|_{\mathcal{H}^{-r}})$ zumindest den Operator-Anteil vollständig aus der Definition der Indexmenge zu eliminieren. Allerdings unterscheidet sich obige Indexmenge \mathcal{T} doch noch um den Faktor $2^{-|\bar{l}|_1/2}$ von den Indexmengen in (6.14), die bezüglich der \mathcal{L}^2 -Norm des Approximationsfehlers $u - \bar{I}_{\mathcal{T}}u$ optimiert waren. Somit ist zu vermuten, daß man mit \mathcal{T} aus (6.18) noch nicht die optimale Konvergenzordnung erzielt. Ein weiteres Problem ist, daß wir für die Abschätzung $\|e\|_{\mathcal{H}^{-r}} \lesssim \delta$ die recht

starke Glattheitsannahme $\tilde{\gamma}^* > r + 1$ an die dualen Wavelets benötigen. D.h. die in Abschnitt 2.5 konstruierten Lifting-Wavelets mit $\overline{M} = 2$ sind dafür noch nicht glatt genug, siehe Tabelle 2.3. Es ist jedoch zu erwarten, daß für Lifting-Wavelets mit einer hinreichend großen Zahl an verschwindenden Momenten \overline{M} die Glattheitsforderungen erfüllt sind. Untersuchungen, wie die Forderung $\tilde{\gamma}^* > r + 1$ abgeschwächt werden kann, sind noch nicht abgeschlossen. Die Vermutung ist hier, daß mit besseren Abschätzungen von $|\langle \tilde{\psi}_{(l,t)}, D^r \overline{\psi}_{(k,s)} \rangle|$, bei denen auch der Abfall im Ort berücksichtigt wird, $\tilde{\gamma}^* > r$ ausreichend ist.

Kapitel 7

Vorkonditionieren von elliptischen Gleichungssystemen

In diesem Kapitel werden zwei Vorkonditionierer für die linearen Gleichungssysteme vorgestellt, die bei der Finite-Differenzen- bzw. der Kollokationsdiskretisierung der Poisson-Gleichung auf adaptiven dünnen Gittern entstehen. Dabei handelt es sich um eine einfache Diagonalskalierung sowie eine Kombination aus Lifting-Wavelettransformation und Diagonalskalierung. Mit dem etwas komplizierteren Vorkonditionierer erzielt man gute bis sehr gute Konvergenzraten beim iterativen Lösen der Gleichungssysteme. Der einfache Diagonal-Vorkonditionierer funktioniert zumindest für glatte Interpolets mit $N \geq 4$ gut. Interessanterweise beobachtet man keine Dimensionsabhängigkeit der Konvergenzraten, wie dies beim Galerkin-Verfahren der Fall wäre.

7.1 Diagonal- und Lifting-Vorkonditionierer

Mit \mathbf{A} wird die Koeffizientenmatrix des linearen Gleichungssystems bezeichnet. \mathbf{C} sei die Matrix-Darstellung des Basiswechsels "hierarchisch \rightarrow lifting". \mathbf{D} und \mathbf{R} sind Diagonalmatrizen mit den Einträgen

$$\mathbf{D}_{(1,t)(1,t)} = \sum_{i=1}^d 4^{l_i} \quad \text{und} \quad \mathbf{R}_{(1,t)(1,t)} = 2^{|\mathbf{l}|_1/2} .$$

Der Diagonal- bzw. der Lifting-Vorkonditionierer ist so definiert, daß die vorkonditionierten Koeffizientenmatrizen die Form

$$\mathbf{R}\mathbf{A}\mathbf{R}^{-1}\mathbf{D}^{-1} \quad \text{bzw.} \quad \mathbf{C}\mathbf{R}\mathbf{A}\mathbf{R}^{-1}\mathbf{C}^{-1}\mathbf{D}^{-1} \quad (7.1)$$

haben. Durch die Multiplikation mit \mathbf{R} bzw. \mathbf{R}^{-1} werden die Ansatzfunktionen in gewisser Weise auf $\|\cdot\|_{\mathcal{L}^2} \sim 1$ normiert, was für das Vorkonditionieren günstiger ist, als $\|\psi_{(1,t)}\|_{\mathcal{L}^2} \sim 2^{-|\mathbf{l}|_1/2}$. Die Algorithmen für das Anwenden der Vorkonditionierer bestehen somit aus einem einfachen Umskalieren von Multiskalenkoeffizienten und gegebenenfalls dem Anwenden der Algorithmen AL2H und AH2L bezüglich aller Koordinatenrichtungen. Diese Algorithmen sind in den vorangegangenen Kapiteln ausführlich beschrieben worden.

Da sowohl \mathbf{A} als auch die Matrizen aus (7.1) unsymmetrisch sind, kommen als iterative Lösungsverfahren unter anderem das CGNE-, das QMR-, das GMRES- oder das BiCGStab-Verfahren in Betracht [BBC⁺94]. Verwendet man beispielsweise das CGNE-Verfahren, dann

würde die Konvergenzrate direkt von der l_2 -Konditionszahl κ abhängen, welche durch

$$\kappa(\mathbf{E}) := \sqrt{\frac{\sup_{\|\mathbf{u}\|_{l_2}=1} \mathbf{u}^T \cdot (\mathbf{E}^T \mathbf{E}) \cdot \mathbf{u}}{\inf_{\|\mathbf{u}\|_{l_2}=1} \mathbf{u}^T \cdot (\mathbf{E}^T \mathbf{E}) \cdot \mathbf{u}}}, \quad \mathbf{E} = \mathbf{R} \mathbf{A} \mathbf{R}^{-1} \mathbf{D}^{-1} \quad \text{bzw.} \quad \mathbf{E} = \mathbf{C} \mathbf{R} \mathbf{A} \mathbf{R}^{-1} \mathbf{C}^{-1} \mathbf{D}^{-1}$$

definiert ist. Allerdings ist das CGNE-Verfahren unattraktiv, da Matrix-Vektor-Multiplikationen mit der Transponierten der vorkonditionierten Koeffizientenmatrix notwendig sind. Dementsprechend wären neben z.B. AFD, AH2L oder AL2H auch Algorithmen zu entwickeln, mit denen die entsprechenden transponierten Operatoren angewendet werden können. Aus dem gleichen Grund ist auch das QMR-Verfahren unattraktiv. In diesem Fall ist unter gewissen Voraussetzungen der Quotient aus maximalem und minimalem Eigenwert der Koeffizientenmatrix für die Konvergenz entscheidend [FN91]. Dieser Quotient kann aber durch κ abgeschätzt werden: $\lambda_{max}/\lambda_{min} \leq \kappa$. Algorithmisch günstiger ist das GMRES-Verfahren, da hier die Transponierte nicht benötigt wird. Für das nach unserer Erfahrung effektivste der oben genannten Verfahren, BiCGStab und seine Varianten, gibt es keine zufriedenstellende Konvergenztheorie. Die numerischen Experimente in Abschnitt 7.2 zeigen aber, daß es eine enge Korrelation zwischen κ und der Konvergenzrate gibt. Unklar ist dabei allerdings, ob primär κ oder z.B. der Quotient der Eigenwerte entscheidend ist.

Wie dem auch sei, an dieser Stelle soll die l_2 -Konditionszahl κ der Matrizen (7.1) in Abhängigkeit vom maximalen Level L abgeschätzt werden. Diese Abschätzungen sind im allgemeinen sehr schwierig zu erhalten. Aus diesem Grund müssen wir uns für die nachfolgende Analyse auf den einfachen Fall von vollen Gittern, d.h. $\mathcal{T} = \{(\mathbf{l}, \mathbf{t}) : |\mathbf{l}|_\infty \leq L\}$ beschränken. Wegen der einfacheren Notation beschränken wir uns bei Beschreibungen außerdem auf den zweidimensionalen Fall. Die wesentlichen Aussagen werden aber für beliebige Dimensionen formuliert.

Für volle Gitter hat die mit \mathbf{R} umskalierte Koeffizientenmatrix $\mathbf{R} \mathbf{A} \mathbf{R}^{-1}$ bzw. die modifizierte Koeffizientenmatrix $\mathbf{C} \mathbf{R} \mathbf{A} \mathbf{R}^{-1} \mathbf{C}^{-1}$ eine relativ einfache Struktur:

$$\begin{aligned} \mathbf{A}_h &:= \mathbf{R} \mathbf{A} \mathbf{R}^{-1} = (T_h^{-1} B T_h) \otimes I + I \otimes (T_h^{-1} B T_h) \\ \text{bzw. } \mathbf{A}_l &:= \mathbf{C} \mathbf{R} \mathbf{A} \mathbf{R}^{-1} \mathbf{C}^{-1} = (T_l^{-1} B T_l) \otimes I + I \otimes (T_l^{-1} B T_l). \end{aligned} \quad (7.2)$$

Dabei ist B die Diskretisierungsmatrix bzgl. der Einzelskalenbasis für den eindimensionalen Laplace-Operator, also beispielsweise

$$B = -\langle \tilde{\Phi}^{S(L)}, \partial_{xx} \Phi^{S(L)} \rangle \quad \text{bzw.} \quad B = -4^L [1 \quad -2 \quad 1]$$

für die Kollokationsmethode bzw. das klassische Finite-Differenzen-Schema 2. Ordnung. T_h bzw. T_l sind die Transformationsmatrizen für die inverse Multiskalenttransformation, welche durch

$$\begin{aligned} 2^{L/2} \Phi^{S(L)} \cdot T_h &= (2^{l_0/2} \Psi_{T(l_0)}, \dots, 2^{L/2} \Psi_{T(L)}) \quad (\text{hierarchisch}) \\ 2^{L/2} \Phi^{S(L)} \cdot T_l &= (2^{l_0/2} \bar{\Psi}_{T(l_0)}, \dots, 2^{L/2} \bar{\Psi}_{T(L)}) \quad (\text{lifting}) \end{aligned} \quad (7.3)$$

definiert sind. Somit ist die Transformationsmatrix für den Basiswechsel ‘‘hierarchisch \rightarrow lifting’’ durch $\mathbf{C} = (T_l^{-1} T_h) \otimes (T_l^{-1} T_h)$ gegeben. Wegen der identischen Struktur von \mathbf{A}_h und \mathbf{A}_l , vgl. rechte Seite von (7.2), können die beiden Vorkonditionierer weitgehend mit den gleichen Techniken analysiert werden. Das Ziel ist dabei, Abschätzungen der Form

$$c(L) \|\mathbf{D} \cdot \mathbf{u}\|_{l_2} \leq \|\mathbf{A}_j \cdot \mathbf{u}\|_{l_2} \leq C(L) \|\mathbf{D} \cdot \mathbf{u}\|_{l_2} \quad \forall \mathbf{u}$$

für $j = h, l$ zu erhalten. Daraus folgt dann $\kappa \leq C(L)/c(L)$.

Neben der Einschränkung auf volle Gitter benötigen wir als weitere Voraussetzungen:

1. die Symmetrie von B
2. die Stabilität der univariaten Diskretisierung, d.h.

$$B \sim 2^L \langle \partial_x \Phi^{S(L)}, \partial_x \Phi^{S(L)} \rangle \quad \text{und falls } \phi \in \mathcal{H}^2 : B^T B \sim 2^L \langle \partial_{xx} \Phi^{S(L)}, \partial_{xx} \Phi^{S(L)} \rangle \quad (7.4)$$

Die Symmetrie wird beim Abschätzen des numerischen Wertebereiches [Zha99, Hac91] von Tensorprodukten von Matrizen durch den numerischen Wertebereich der einzelnen Matrizen gebraucht. Wegen der Symmetriebedingung betrachten wir hier nur *periodische* Randbedingungen. In diesem Fall ist auch die Stabilitätsbedingung (7.4) für die Kollokationsmethode bzw. für Finite-Differenzen-Schemen einfach nachzuweisen:

Sowohl B als auch $F := 2^L \langle \partial_x \Phi^{S(L)}, \partial_x \Phi^{S(L)} \rangle$ sind zirkulant und haben deshalb eine gemeinsame Basis von orthogonalen Eigenvektoren. Die entsprechenden Eigenwerte, λ_k von B und μ_k von F , sind aber für die Kollokationsdiskretisierung und die üblichen Finite-Differenzen-Schemen äquivalent, d.h. $\lambda_k \sim \mu_k \sim k^2$ für $0 \leq k \leq 2^{L-1}$. Analog kann man $B^T B \sim 2^L \langle \partial_{xx} \Phi^{S(L)}, \partial_{xx} \Phi^{S(L)} \rangle$ zeigen. Hier müssen die Eigenwerte $\sim k^4$ sein. Man beachte aber, daß es auch instabile Finite-Differenzen-Diskretisierungen gibt. Ein interessantes Beispiel ist der Fall, wo der ∂_{xx} -Operator durch Hintereinanderschalten von zwei zentralen Differenzen-Schemen für die erste Ableitung diskretisiert wird. Solche zusammengesetzten Diskretisierungen entstehen z.B. bei Sattelpunktproblemen, wie in Kapitel 10. Hier entspricht die Stabilität im wesentlichen einer LBB-Bedingung.

Im ersten Zwischenschritt werden die Konditionszahlen von gegebenenfalls vorkonditionierten univariaten Operatoren untersucht. Dazu werden die bekannten Normäquivalenzen aus Abschnitt 2.2 mittels der Stabilitätseigenschaften (7.4) auf die entsprechenden Matrizen übertragen.

Theorem 7.1 (Stabilität der Operator-Matrizen) *Wir setzen voraus, daß $N \geq 2$, $\gamma^* > 1$ und $\tilde{\gamma}^* > 0$. Weiterhin sei D die Diagonalmatrix mit Einträgen $D_{(l,t)(l,t)} = 4^l$. Dann gilt*

(i) *(Transformationsmatrizen) Für $j = h, l$ gibt es Konstanten $0 < \underline{a}_j(L) \leq \bar{a}_j(L)$ mit*

$$\underline{a}_j(L) \mathbf{u}^T \cdot \mathbf{u} \leq \mathbf{u}^T \cdot T_j^T T_j \cdot \mathbf{u} \leq \bar{a}_j(L) \mathbf{u}^T \cdot \mathbf{u} . \quad (7.5)$$

Dabei gilt $2^{-L} \lesssim \underline{a}_h(L)$, $\bar{a}_h(L) \lesssim L$ sowie $\underline{a}_l(L) \sim \bar{a}_l(L) \sim 1$.

(ii) *(Steifigkeitsmatrizen) Für Koeffizientenvektoren $\mathbf{u} = \{u_{(l,t)}\}_{(l,t): l_0 \leq l \leq L, t \in T(l)}$ mit $\mathbf{1}^T \cdot T_j \cdot \mathbf{u} = 0$, was bei periodischen Randbedingungen einem verschwindenden Mittelwert*

$$\int_0^1 \sum_{l=l_0}^L \Psi_l \cdot \mathbf{u}_l dx = 0 \quad \text{bzw.} \quad \int_0^1 \sum_{l=l_0}^L \bar{\Psi}_l \cdot \mathbf{u}_l dx = 0 \quad (7.6)$$

entspricht, gilt für $j = h, l$

$$\mathbf{u}^T \cdot D \cdot \mathbf{u} \sim \mathbf{u}^T \cdot T_j^T B T_j \cdot \mathbf{u} . \quad (7.7)$$

(iii) (Bi-Steifigkeitsmatrizen) Für $j = h, l$ gibt es Konstanten $0 < \underline{b}_j(L) \leq \bar{b}_j(L)$, so daß für Koeffizientenvektoren \mathbf{u} mit (7.6) gilt

$$\underline{b}_j(L) \mathbf{u}^T \cdot D^2 \cdot \mathbf{u} \leq \mathbf{u}^T \cdot (BT_j)^T (BT_j) \cdot \mathbf{u} \leq \bar{b}_j(L) \mathbf{u}^T \cdot D^2 \cdot \mathbf{u} . \quad (7.8)$$

Falls $\gamma^* > 2$, dann ist $\underline{b}_j(L) \sim \bar{b}_j(L) \sim 1$. Ansonsten gilt für beliebiges $\epsilon > 0$

$$\bar{b}_j(L) \lesssim 4^{(2-\gamma^*+\epsilon)L} \quad \text{und} \quad \underline{b}_j(L) \gtrsim \begin{cases} L^{-1} & \text{falls } N = 2 \\ 1 & \text{falls } N > 2 \end{cases} .$$

(iv) (Quadrate der Steifigkeitsmatrizen) Für $j = h, l$ gibt es Konstanten $0 < \underline{c}_j(L) \leq \bar{c}_j(L)$, so daß für alle Koeffizientenvektoren \mathbf{u} mit (7.6)

$$\underline{c}_j(L) \mathbf{u}^T \cdot D^2 \cdot \mathbf{u} \leq \mathbf{u}^T \cdot (T_j^{-1} BT_j)^T (T_j^{-1} BT_j) \cdot \mathbf{u} \leq \bar{c}_j(L) \mathbf{u}^T \cdot D^2 \cdot \mathbf{u} .$$

gilt. Dabei ist $\frac{\underline{b}_j(L)}{\underline{a}_j(L)} \lesssim \underline{c}_j(L)$ und $\bar{c}_j(L) \lesssim \frac{\bar{b}_j(L)}{\bar{a}_j(L)}$.

Die Aussagen sind die diskreten Analogons der Normäquivalenzen für \mathcal{H}^0 , \mathcal{H}^1 bzw. \mathcal{H}^2 . Die Beweise sind in Anhang A.16 aufgeführt. In Tabelle 7.1 sind numerische Schätzwerte zu $\underline{a}_j(L), \dots, \bar{c}_j(L)$ für Interpolets und Lifting-Interpolets mit $\bar{M} = N = 2$ abgebildet. In diesem Fall ist $\gamma^* = 1.5$, so daß sich die auftretenden Werte in etwa folgendermaßen verhalten sollten:

$\underline{a}_h \sim 2^{-L}$	$\bar{a}_h \sim L$	$\underline{a}_l \sim 1$	$\bar{a}_l \sim 1$
$\underline{b}_h \sim L^{-1}$	$\bar{b}_h \sim 2^L$	$\underline{b}_l \sim L^{-1}$	$\bar{b}_l \sim 2^L$
$\underline{c}_h \sim L^{-2}$	$\bar{c}_h \sim 4^L$	$\underline{c}_l \sim L^{-1}$	$\bar{c}_l \sim 2^L$

Die numerischen Werte wurden mit MatLab aus den Eigenwerten der Operatormatrizen berechnet. Die Werte stimmen sehr gut mit der Theorie überein.

Nach diesen Vorarbeiten kann nun für $j = h, l$ $\|\mathbf{A}_j \cdot \mathbf{u}\|_{l_2}$ durch $\|\mathbf{D} \cdot \mathbf{u}\|_{l_2}$ abgeschätzt werden. Dabei werden wir an einigen Stellen von der folgenden Modifikation von Theorem 6.19 aus [Zha99] Gebrauch machen.

Lemma 7.1 (Numerischer Wertebereich von Tensorprodukten von Matrizen)

Seien $M_1, M_2 \in M(n, n)$ symmetrische, positiv semidefinite Matrizen deren Kern jeweils von den Vektoren \mathbf{n}_1 bzw. $\mathbf{n}_2 \in \mathbb{R}^n$ aufgespannt wird, d.h. $M_i \cdot \mathbf{n}_i = 0$ ($\mathbf{n}_i = 0$ ist erlaubt). Die kleinsten bzw. größten nicht-verschwindenden Eigenwerte seien $0 < \underline{\alpha}_i < \bar{\alpha}_i$, $i = 1, 2$. Dann gilt für alle $\mathbf{u} \in \mathbb{R}^{n \times n}$ mit

$$\mathbf{u}^T \cdot (\mathbf{n}_1 \otimes \mathbf{q}) = \mathbf{u}^T \cdot (\mathbf{q} \otimes \mathbf{n}_2) = 0 \quad \text{für alle } \mathbf{q} \in \mathbb{R}^n \quad (7.9)$$

die Abschätzung

$$\underline{\alpha}_1 \underline{\alpha}_2 \mathbf{u}^T \cdot \mathbf{u} \leq \mathbf{u}^T \cdot (M_1 \otimes M_2) \cdot \mathbf{u} \leq \bar{\alpha}_1 \bar{\alpha}_2 \mathbf{u}^T \cdot \mathbf{u} .$$

Beweis: Eine Orthogonalbasis von Eigenvektoren von $M_1 \otimes M_2$ ist durch Tensorprodukte orthogonaler Eigenvektoren von M_1 bzw. M_2 gegeben [Zha99]. Stellt man \mathbf{u} als Linearkombination dieser orthogonalen Eigenvektoren von $M_1 \otimes M_2$ dar, dann tauchen darin wegen (7.9) nur Eigenvektoren zu nicht-verschwindenden Eigenwerten auf. Diese Eigenwerte sind aber durch $\underline{\alpha}_1 \underline{\alpha}_2$ bzw. $\bar{\alpha}_1 \bar{\alpha}_2$ beschränkt. Daraus erhält man die Behauptung. \square

L	$\underline{a}_h(L)$	$\underline{a}_h(L)2^L$	$\bar{a}_h(L)$	$\bar{a}_h(L)/L$	L	$\underline{a}_l(L)$		$\bar{a}_l(L)$	
4	1.7e-1	2.71	2.21	0.55	4	0.34		1.00	
6	4.0e-2	2.55	3.21	0.54	6	0.18		1.00	
8	9.8e-3	2.52	4.22	0.53	8	0.13		1.00	
10	2.4e-3	2.51	5.24	0.52	10	0.11		1.00	
L	$\underline{b}_h(L)$	$\underline{b}_h(L)L$	$\bar{b}_h(L)$	$\bar{b}_h(L)/2^L$	L	$\underline{b}_l(L)$	$\underline{b}_l(L)L$	$\bar{b}_l(L)$	$\bar{b}_l(L)/2^L$
4	2.17	8.66	66.1	4.13	4	1.89	7.56	67.3	4.21
6	1.41	8.49	270.4	4.22	6	1.27	7.60	276.9	4.33
8	1.04	8.34	1087.5	4.25	8	0.95	7.64	1115.2	4.36
10	0.82	8.24	4356.0	4.25	10	0.77	7.66	4468.2	4.36
L	$\underline{c}_h(L)$	$\underline{c}_h(L)L^2$	$\bar{c}_h(L)$	$\bar{c}_h(L)/4^L$	L	$\underline{c}_l(L)$	$\underline{c}_l(L)L$	$\bar{c}_l(L)$	$\bar{c}_l(L)/2^L$
4	1.17	1.88	3.0e+2	1.17	4	2.45	9.8	182	11.4
6	0.50	1.80	5.3e+3	1.30	6	1.73	10.4	1165	18.2
8	0.27	1.74	8.7e+4	1.33	8	1.31	10.5	5789	22.6
10	0.17	1.70	1.4e+6	1.34	10	1.05	10.5	25373	24.8

Tabelle 7.1: Werte $\underline{a}_j(L), \dots, \bar{c}_j(L)$ für Interpolets(links), Lifting-Interpolets(rechts), $N = 2$, $B = 4^L[1 \ -2 \ 1]$.

Theorem 7.2 Für Koeffizientenvektoren \mathbf{u} mit

$$\int_{[0,1]^d} \sum_1 \Psi_1 \cdot \mathbf{u}_1 d\mathbf{x} = 0 \quad \text{bzw.} \quad \int_{[0,1]^d} \sum_1 \bar{\Psi}_1 \cdot \mathbf{u}_1 d\mathbf{x} = 0 \quad (\text{verschwindender Mittelwert})$$

gilt

$$\|\mathbf{A}_j \cdot \mathbf{u}\|_{l_2}^2 \lesssim \bar{c}_j(L) \|\mathbf{D} \cdot \mathbf{u}\|_{l_2}^2, \quad j = h, l. \quad (7.10)$$

Man beachte, daß zwar die in \lesssim verborgenen Konstanten von der Dimension d abhängen können, der möglicherweise vom maximalen Level L abhängige Wert $\bar{c}_j(L)$ geht aber unabhängig von d in die Abschätzungen ein.

Beweis: (Theorem 7.2) Wir betrachten nur den zweidimensionalen Fall. Der Beweis ist für beliebige Dimensionen analog, aber notationstechnisch aufwendiger. Wir benutzen die Darstellungen (7.2) von \mathbf{A}_j . Sei $E = T_j^{-1} B T_j$, dann ist

$$\begin{aligned} \|\mathbf{A}_j \cdot \mathbf{u}\|_{l_2}^2 &= \mathbf{u}^T \cdot (E \otimes I + I \otimes E)^T (E \otimes I + I \otimes E) \cdot \mathbf{u} \\ &= \mathbf{u}^T \cdot (E^T E \otimes I + I \otimes E^T E) \cdot \mathbf{u} + \mathbf{u}^T \cdot (E^T \otimes E + E \otimes E^T) \cdot \mathbf{u} \\ &\leq 2\mathbf{u}^T \cdot (E^T E \otimes I + I \otimes E^T E) \cdot \mathbf{u} \end{aligned} \quad (7.11)$$

Die einzelnen Terme $\mathbf{u}^T \cdot (E^T E \otimes I) \cdot \mathbf{u}$ bzw. $\mathbf{u}^T \cdot (I \otimes E^T E) \cdot \mathbf{u}$ können nun mittels Lemma 7.1 und Theorem 7.1 durch $\mathbf{u}^T \cdot (D^2 \otimes I) \cdot \mathbf{u}$ bzw. $\mathbf{u}^T \cdot (I \otimes D^2) \cdot \mathbf{u}$ abgeschätzt werden. Die dafür notwendige Voraussetzung des verschwindenden Mittelwertes entlang jeder Koordinatenrichtung erfordert allerdings einen Zwischenschritt. Dazu betrachten wir die folgende Zerlegung des Koeffizientenvektors $\mathbf{u} = \mathbf{v} + \mathbf{n} \otimes \mathbf{w}$. Darin ist \mathbf{w} der Vektor mit Komponenten $w_{(l,t)} = \sum_{t_1 \in T(l_0)} u_{((l_0,l),(t_1,t))}$ und \mathbf{n} ist der Vektor $T_j^{-1} \cdot \mathbf{1}$ mit Komponenten

$$n_{(l,t)} = \begin{cases} 2^{-l_0/2} & \text{für } l = l_0 \text{ und } t \in T(l_0) \\ 0 & \text{sonst.} \end{cases} \quad (7.12)$$

Da $B \cdot \mathbf{1} = 0$, ist der Vektor \mathbf{n} der Nullvektor von E . Außerdem gilt auch $D \cdot \mathbf{n} = 4^{l_0} \mathbf{n}$. Das Tensorprodukt $\mathbf{p} := \mathbf{n} \otimes \mathbf{w}$ ist ein 'multi-dimensionaler' Vektor mit Komponenten $p_{((l_1, l_2), (t_1, t_2))} = n_{(l_1, t_1)} w_{(l_2, t_2)}$. Die Zerlegung von \mathbf{u} ist gerade so definiert, daß für alle 'ein-dimensionalen' Vektoren \mathbf{q} gilt: $(\mathbf{n} \otimes \mathbf{q})^T \cdot \mathbf{v} = 0$. Mit diesen Eigenschaften erhalten wir

$$\begin{aligned}
& (\mathbf{v} + \mathbf{n} \otimes \mathbf{w})^T \cdot (E^T E \otimes I) \cdot (\mathbf{v} + \mathbf{n} \otimes \mathbf{w}) = \\
&= \mathbf{v}^T \cdot (E^T E \otimes I) \cdot \mathbf{v} + \underbrace{\mathbf{v}^T \cdot (E^T E \cdot \mathbf{n} \otimes \mathbf{w}) + (E^T E \cdot \mathbf{n} \otimes \mathbf{w})^T \cdot \mathbf{v}}_{=0} + (\mathbf{n}^T \cdot E^T E \cdot \mathbf{n}) \mathbf{w}^T \cdot \mathbf{w} \\
&\leq \bar{c}_j(L) \mathbf{v}^T \cdot (D^2 \otimes I) \cdot \mathbf{v} \\
&\leq \bar{c}_j(L) \mathbf{v}^T \cdot (D^2 \otimes I) \cdot \mathbf{v} + \underbrace{\mathbf{v}^T \cdot (D^2 \cdot \mathbf{n} \otimes \mathbf{w}) + (D^2 \cdot \mathbf{n} \otimes \mathbf{w})^T \cdot \mathbf{v}}_{=0} + \underbrace{(\mathbf{n}^T \cdot D^2 \cdot \mathbf{n}) \mathbf{w}^T \cdot \mathbf{w}}_{\geq 0} \\
&= \bar{c}_j(L) (\mathbf{v} + \mathbf{n} \otimes \mathbf{w})^T \cdot (D^2 \otimes I) \cdot (\mathbf{v} + \mathbf{n} \otimes \mathbf{w}) \\
&= \bar{c}_j(L) \mathbf{u}^T \cdot (D^2 \otimes I) \cdot \mathbf{u} .
\end{aligned} \tag{7.13}$$

Insgesamt erhält man aus (7.11) und (7.13) die Behauptung (7.10). Im dreidimensionalen Fall würde man an Stelle von (7.11) $\|\mathbf{A}_j \cdot \mathbf{u}\|_{l_2}^2$ durch $3\mathbf{u}^T (E^T E \otimes I \otimes I + I \otimes E^T E \otimes I + I \otimes I \otimes E^T E) \cdot \mathbf{u}$ abschätzen und dann analog zum zweidimensionalen Fall fortfahren. \square

Leider kann die im Beweis für Theorem 7.2 benutzte Technik nicht ohne weiteres auch für Abschätzungen von $\|\mathbf{A}_j \cdot \mathbf{u}\|_{l_2}$ nach unten benutzt werden. Der kritische Punkt dabei ist, die 'gemischten' Terme $\mathbf{u}^T \cdot E^T \otimes E \cdot \mathbf{u}$ nach unten abzuschätzen. Ideal wäre z.B. die Existenz von $c < 1$, so daß

$$\mathbf{u}^T \cdot (E^T \otimes E + E \otimes E^T) \cdot \mathbf{u} \geq -c \mathbf{u}^T \cdot (E^T E \otimes I + I \otimes E^T) \cdot \mathbf{u} . \tag{7.14}$$

In diesem Fall wäre

$$\begin{aligned}
\|(E \otimes I + I \otimes E) \cdot \mathbf{u}\|_{l_2}^2 &\gtrsim \mathbf{u}^T \cdot (E^T E \otimes I + I \otimes E^T E) \cdot \mathbf{u} \\
&\gtrsim \underline{c}_j(L) \mathbf{u}^T \cdot (D^2 \otimes I + I \otimes D^2) \cdot \mathbf{u} \\
&\geq \frac{1}{2} \underline{c}_j(L) \mathbf{u}^T \cdot (D \otimes I + I \otimes D)^2 \cdot \mathbf{u} .
\end{aligned}$$

D.h. in diesem Fall würde der Level-abhängige Term $\underline{c}_j(L)$ ebenfalls *unabhängig von der Dimension* in die Abschätzungen eingehen. Insbesondere wäre die Kondition der vorkonditionierten Koeffizientenmatrizen nur sehr schwach über die in ' \gtrsim ' enthaltenen Konstanten von d abhängig: $\kappa(\mathbf{A}_j \mathbf{D}^{-1}) \lesssim \bar{c}_j(L) / \underline{c}_j(L)$. Dies wäre ein sehr wichtiger Unterschied zum Galerkin-Verfahren: Hier erhält man mit Lemma 7.1 aus den Abschätzungen für $T_j^T T_j$ und $T_j^T B T_j$ in Theorem 7.1 Konditionszahlen der Ordnung $(\bar{a}_j(L) / \underline{a}_j(L))^{d-1}$ für die diagonal vorkonditionierte Steifigkeitsmatrix. Für Interpolets ergibt das $\sim (L^2 L)^{d-1}$. Bei der Finite-Differenzen-Technik würde also die einfachere Struktur der Koeffizientenmatrix

$$\sum_{i=1}^d \underbrace{I \otimes \dots \otimes I}_{(i-1)\text{-mal}} \otimes (T_h^{-1} B T_h) \otimes \underbrace{I \otimes \dots \otimes I}_{(d-i)\text{-mal}}$$

im Vergleich zu

$$\sum_{i=1}^d \underbrace{T_h^T T_h \otimes \dots \otimes T_h^T T_h}_{(i-1)\text{-mal}} \otimes (T_h^T B T_h) \otimes \underbrace{T^T T \otimes \dots \otimes T_h^T T_h}_{(d-i)\text{-mal}}$$

für das Galerkin-Verfahren, zu einem günstigeren Verhalten der Konditionszahlen führen. Obgleich noch kein Beweis für (7.14) bekannt ist, deuten die Resultate der numerischen Experimente im nächsten Abschnitt doch darauf hin, daß die Konditionszahlen tatsächlich weitgehend unabhängig von d sind. Für die folgenden Abschätzungen von $\|\mathbf{A}_j \cdot \mathbf{u}\|_{l_2}$ nach unten wird (7.14) nicht benötigt, allerdings erhält man auch nur für \mathbf{A}_l befriedigende Aussagen.

Theorem 7.3 *Unter den Voraussetzungen von Theorem 7.2 gilt*

$$\frac{\underline{b}_j(L)}{\bar{a}_j(L)} \left(\frac{\underline{a}_j(L)}{\bar{a}_j(L)} \right)^{d-1} \|\mathbf{D} \cdot \mathbf{u}\|_{l_2}^2 \lesssim \|\mathbf{A}_j \cdot \mathbf{u}\|_{l_2}^2, \quad j = h, l. \quad (7.15)$$

Für den Lifting-Vorkonditionierer, d.h. \mathbf{A}_l , ist $\underline{a}_l \sim \bar{a}_l \sim 1$, so daß die Abschätzung weitgehend mit $\|\mathbf{A}_l\|_{l_2}^2 \gtrsim \underline{c}_j(L) \|\mathbf{D} \cdot \mathbf{u}\|_{l_2}^2$ übereinstimmt. Für den einfachen Diagonal-Vorkonditionierer ist allerdings $\underline{a}_j(L)/\bar{a}_j(L) \sim (L2^L)^{-1}$, so daß die Abschätzung (7.15) sehr schwach ist.

Beweis: (Theorem 7.3) Wir beweisen wieder nur den zweidimensionalen Fall, benutzen aber an den entsprechenden Stellen die allgemeinen dimensionsabhängigen Terme. Aus der Darstellung (7.2) folgt, daß \mathbf{A}_j auch als

$$\mathbf{A}_j = (T_j^{-1} \otimes T_j^{-1})(B \otimes I + I \otimes B)(T_j^{-1} \otimes T_j^{-1})$$

geschrieben werden kann, so daß $\|\mathbf{A}_j \cdot \mathbf{u}\|_{l_2}^2 =$

$$\begin{aligned} &= \mathbf{u}^T \cdot (T_j \otimes T_j)^T (B \otimes I + I \otimes B)^T (T_j T_j^T)^{-1} \otimes (T_j T_j^T)^{-1} (B \otimes I + I \otimes B) (T_j \otimes T_j) \cdot \mathbf{u} \\ &\geq \bar{a}_j(L)^{-d} \mathbf{u}^T \cdot (T_j \otimes T_j)^T (B \otimes I + I \otimes B)^T (B \otimes I + I \otimes B) (T_j \otimes T_j) \cdot \mathbf{u} \\ &= \bar{a}_j(L)^{-d} \mathbf{u}^T \cdot (T_j^T B^T B T_j \otimes T_j^T T_j + 2T_j^T B T_j \otimes T_j^T B T_j + T_j^T T_j \otimes T_j^T B^T B T_j) \cdot \mathbf{u} \quad (7.16) \end{aligned}$$

Die einzelnen Terme können wie im Beweis von Theorem 7.2 mit Lemma 7.1 und Theorem 7.1 abgeschätzt werden:

$$\begin{aligned} \mathbf{u}^T \cdot T_j^T B^T B T_j \otimes T_j^T B T_j \cdot \mathbf{u} &\gtrsim \mathbf{u}^T \cdot D \otimes D \cdot \mathbf{u} \\ \mathbf{u}^T \cdot T_j^T B^T B T_j \otimes T_j^T T_j \cdot \mathbf{u} &\geq \underline{b}_j(L) \underline{a}_j(L)^{d-1} \mathbf{u}^T \cdot D^2 \otimes I \cdot \mathbf{u} \end{aligned}$$

Da $\underline{b}_j(L) \underline{a}_j(L)^{d-1} \lesssim 1$ folgt daraus zusammen mit (7.16)

$$\|\mathbf{A}_j \cdot \mathbf{u}\|_{l_2}^2 \gtrsim \frac{\underline{b}_j(L)}{\bar{a}_j(L)} \left(\frac{\underline{a}_j(L)}{\bar{a}_j(L)} \right)^{d-1} \mathbf{u}^T \cdot (D^2 \otimes I + 2D \otimes D + I \otimes D^2) \cdot \mathbf{u} = \frac{\underline{b}_j(L)}{\bar{a}_j(L)} \left(\frac{\underline{a}_j(L)}{\bar{a}_j(L)} \right)^{d-1} \mathbf{u}^T \cdot \mathbf{D}^2 \cdot \mathbf{u},$$

also die Behauptung. \square

Bemerkung 7.1 Benutzt man für das Petrov-Galerkin-Verfahren direkt die Lifting-Wavelets und ihre dualen Wavelets als Ansatz- bzw. Testfunktionen, dann hat (zumindest im nichtadaptiven Fall) die Steifigkeitsmatrix unmittelbar die Form von \mathbf{A}_l aus (7.2). D.h. für die stabilen Wavelets genügt ein diagonales Skalieren für ein $O(1)$ -Vorkonditionieren. Dies ist also ein weiterer Vorteil von stabilen Lifting-Wavelets im Zusammenhang mit biorthogonalen Petrov-Galerkin-Verfahren, siehe auch die Diskussion am Ende von Abschnitt 6.5.

Neben den in diesem Kapitel vorgestellten Vorkonditionierern gibt es zwei weitere Techniken zur Beschleunigung iterativer Löser bzw. zum schnellen Lösen der bei der Finite-Differenzen-Methode entstehenden linearen Gleichungssysteme.

In [Sch98a] ist beispielsweise ebenfalls eine einfache diagonale Skalierung benutzt worden. Da dort aber die primären Unbekannten die Knotenwerte der Funktion sind, werden nicht wie hier die Multiskalenkoeffizienten, sondern Knotenwerte umskaliert. Die Skalierungsfaktoren setzen sich dabei aus den lokalen Maschenweiten bzgl. der d Koordinatenrichtungen zusammen. Für volle Gitter sind somit die Faktoren für alle Knoten gleich, so daß der Vorkonditionierer in diesem Fall *keinen* Effekt hat. Für dünne Gitter wird allerdings ein gutes Vorkonditionierungsergebnis erreicht, bei dem die Zahl der zum Erreichen einer bestimmten Genauigkeit notwendigen Iterationen etwa wie $\sqrt{2^L}$ anwächst (2D Beispiel). Das gleiche Verhalten beobachtet man auch für den hier vorgestellten Diagonal-Vorkonditionierer bei stückweise linearen Interpolets, siehe nächster Abschnitt. Im Gegensatz zu [Sch98a] ist dieser Vorkonditionierer für volle Gitter aber *genauso* wirksam. Der Lifting-Vorkonditionierer ist sogar noch effizienter, hier ist die Zahl der für eine bestimmte Genauigkeit notwendigen Iterationen fast unabhängig von L .

Ein etwas anderer Ansatz ist in [Spr01, HS99] verfolgt worden. Darin wird ein Mehrgitterartiges Schema konstruiert, das auf multivariaten Vergrößerungsstrategien für reguläre dünne Gitter aufbaut [HP97]. Allerdings sind die mit dieser Technik erzielten Ergebnisse nicht sehr gut.

7.2 Numerische Experimente

Die im vorangegangenen Abschnitt dargestellte Analyse des Diagonal- bzw. Lifting-Vorkonditionierers war durch die folgenden Einschränkungen gekennzeichnet: nur volle Gitter, nur periodische Randbedingungen, schlechte Abschätzungen für den Diagonal-Vorkonditionierer, dabei insbesondere eine unangenehme Dimensionsabhängigkeit für κ . Aus diesem Grund sollen nachfolgend die Resultate einiger Experimente dargestellt werden, die über die obigen Einschränkungen hinausgehen. Dabei stehen die folgenden Fragen im Mittelpunkt:

1. Gibt es Unterschiede im Verhalten der Vorkonditionierer zwischen vollen und dünnen Gittern ?
2. Gibt es Probleme mit z.B. Dirichlet-Randbedingungen und den resultierenden unsymmetrischen Diskretisierungen ?
3. Gibt es eine Dimensionsabhängigkeit im Verhalten der Vorkonditionierer ?
4. Die im Abschnitt 7.1 gefundenen Abschätzungen lassen vermuten, daß die Konvergenzraten für glatte Interpolets, $\gamma^* > 2$, besser sein müssen als für weniger glatte Interpolets, wie die hierarchische Basis mit $N = 2$ und $\gamma^* = 1.5$. Ist das der Fall ?

Als Ergebnis unserer Experimente können wir diese Fragen folgendermaßen beantworten:

1. Beide Vorkonditionierer funktionieren für volle und dünne Gitter ähnlich gut bis sehr gut.
2. Es gibt keine Probleme mit z.B. Dirichlet-Randbedingungen.
3. Es gibt so gut wie keine Dimensionsabhängigkeit beim Vorkonditionieren.
4. Glattere Interpolets bewirken beim Diagonal-Vorkonditionierer ein deutlich besseres Konvergenzverhalten. Beim Lifting-Vorkonditionierer sind die Konvergenzraten ohnehin so gut, daß die Glattheit nur einen unbedeutenden Einfluß hat.

Für die Experimente betrachten wir das Modellproblem

$$-\Delta u = -d9\pi^2 \prod_{i=1}^d \sin(3\pi x_i), \quad \mathbf{x} \in]0, 1[^d \quad (7.17)$$

für $d = 2, 3$ auf vollen und auf regulären dünnen Gittern mit homogenen Dirichlet-Randbedingungen. Als iterativer Löser wurde das BiCGStab-Verfahren benutzt. Für die Berechnung der Konvergenzhistorien wurde jeweils das entsprechende lineare Gleichungssystem mit hoher Genauigkeit gelöst. Diese Lösung sei mit $u_{\mathcal{T}}$ bezeichnet. Anschließend wurde das lineare Gleichungssystem erneut mit dem BiCGStab-Verfahren gelöst (Anfangswert $u_{\mathcal{T}}^{(0)} = \mathbf{0}$). Diesmal wird aber eine diskrete Variante $e^{(it)}$ der \mathcal{L}^2 -Norm des Fehlers protokolliert:

$$e^{(it)} := \left(\int_{[0,1]^d} I_{\mathcal{T}}[(u_{\mathcal{T}} - u_{\mathcal{T}}^{(it)})^2] d\mathbf{x} \right)^{1/2}.$$

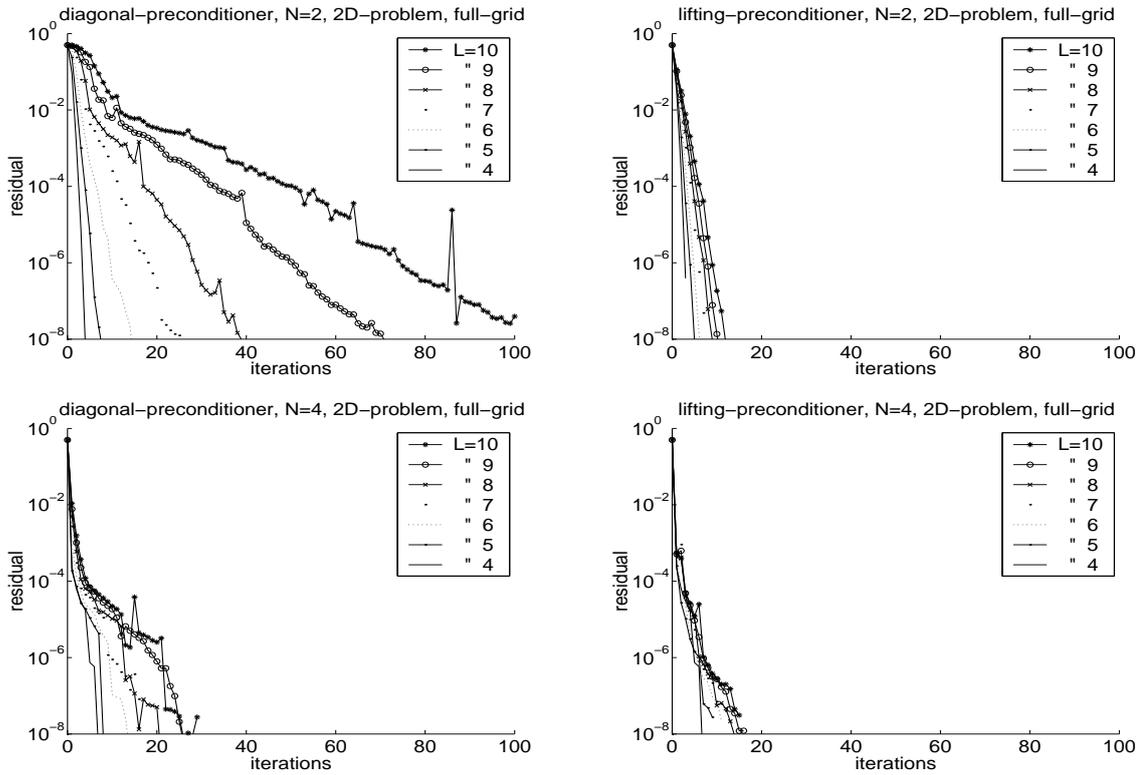
Da die Lösung $u_{\mathcal{T}}$ als auch die Iterierten $u_{\mathcal{T}}^{(it)}$ glatte Funktionen sind, stimmen die diskreten \mathcal{L}^2 -Normen sehr gut mit den kontinuierlichen Normen überein. Eigentlich braucht man sogar nur, daß die diskrete \mathcal{L}^2 -Norm und die kontinuierliche Norm des Fehlers bis auf konstante Faktoren äquivalent sind.

In Abbildung 7.1 sind die Ergebnisse für den 2D Fall dargestellt und in Abbildung 7.2 die Ergebnisse für den 3D Fall. In den einzelnen Diagrammen ist jeweils $e^{(it)}$ mit einer logarithmischen Skala über it aufgetragen. Es gibt offensichtlich keine großen Unterschiede zwischen 2D und 3D für die einzelnen Experimente. Man beachte, daß die Voll-Gitter-Experimente in 3D nur bis Level 7 durchgeführt werden konnten.

Sowohl in 2D als auch in 3D gibt es keinen signifikanten Unterschied zwischen vollen und dünnen Gittern. Die Dünn-Gitter Resultate sind jeweils in der unteren Bildhälfte zu finden. Einen deutlichen Unterschied gibt es nur zwischen dem Diagonal- und dem Lifting-Vorkonditionierer für stückweise lineare Interpolets/Lifting-Interpolets mit $N = 2$. Während beim Lifting-Vorkonditionierer die Konvergenzraten nahezu Level-unabhängig sind, werden diese beim Diagonal-Vorkonditionierer deutlich schlechter. Bei den glatten Interpolets mit $N = 4$ ist der Unterschied zwischen Lifting- und Diagonal-Vorkonditionierer allerdings gering. Hier erzielen beide Vorkonditionierer ähnliche, sehr gute Konvergenzraten, so daß sich hier der Lifting-Vorkonditionierer kaum noch lohnt.

(Zusammenfassung) Die in diesem Kapitel vorgestellten Vorkonditionierer gestatten ein effizientes Lösen der bei der Finite-Differenzen/Kollokationsmethode entstehenden linearen Gleichungssysteme. Die Konvergenzraten scheinen weitgehend unabhängig von Raumdimension und Typ des Gitters zu sein. Für glatte Interpolets mit $N \geq 4$, können bereits mit dem einfachen Diagonal-Vorkonditionierer sehr gute Konvergenzraten erzielt werden.

VOLLES GITTER 2D



DÜNNES GITTER 2D

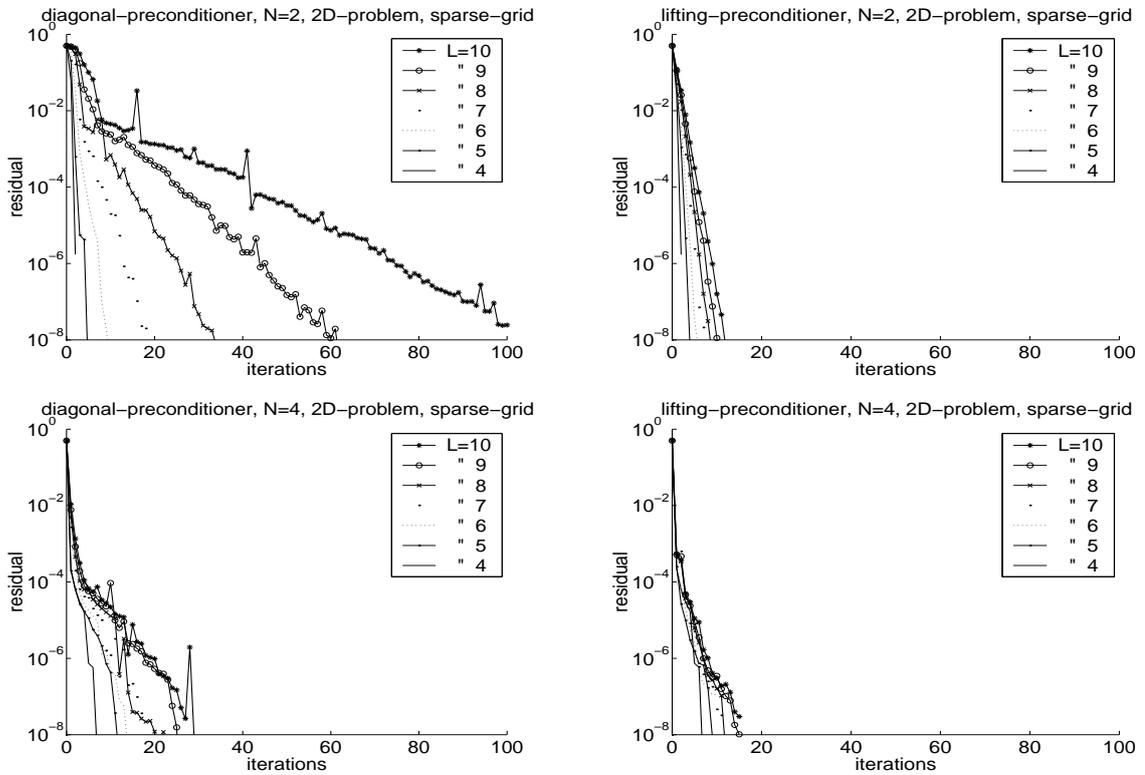
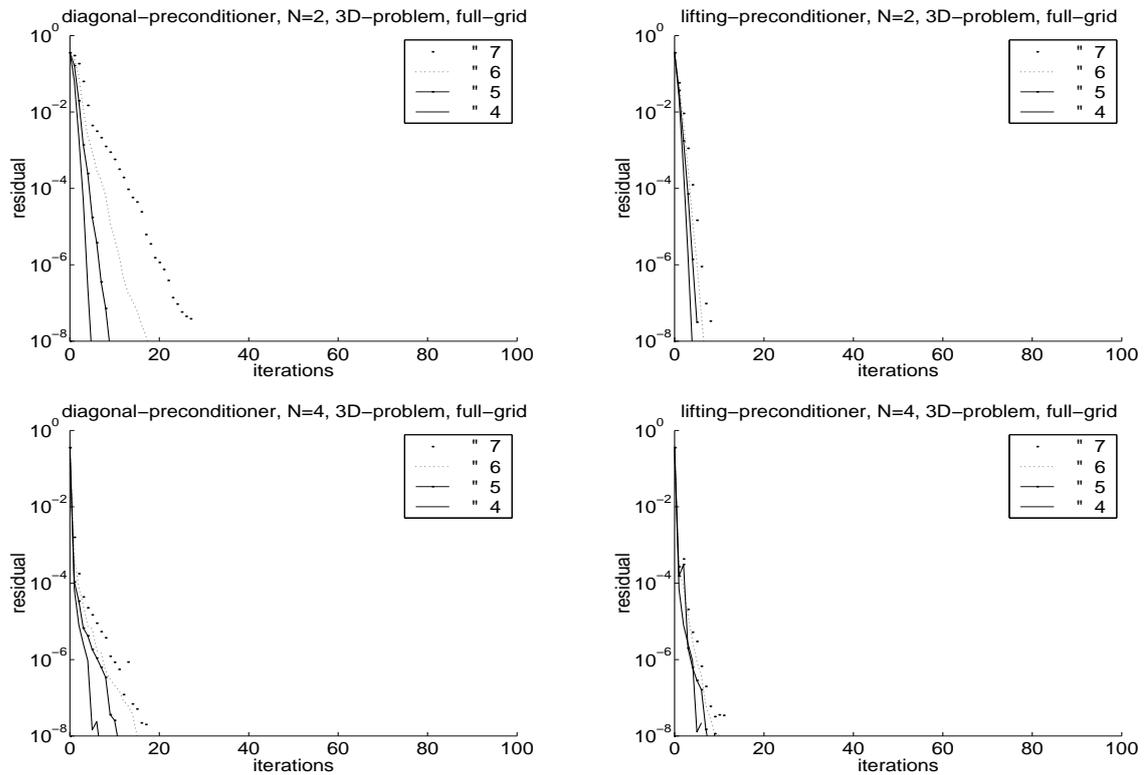


Abbildung 7.1: Links: Diagonal-, Rechts: Lifting-Vorkonditionierer. Erste und dritte Zeile: $N = 2$. Zweite und vierte Zeile $N = 4$. Problem (7.17) mit $d = 2$.

VOLLES GITTER 3D



DÜNNES GITTER 3D

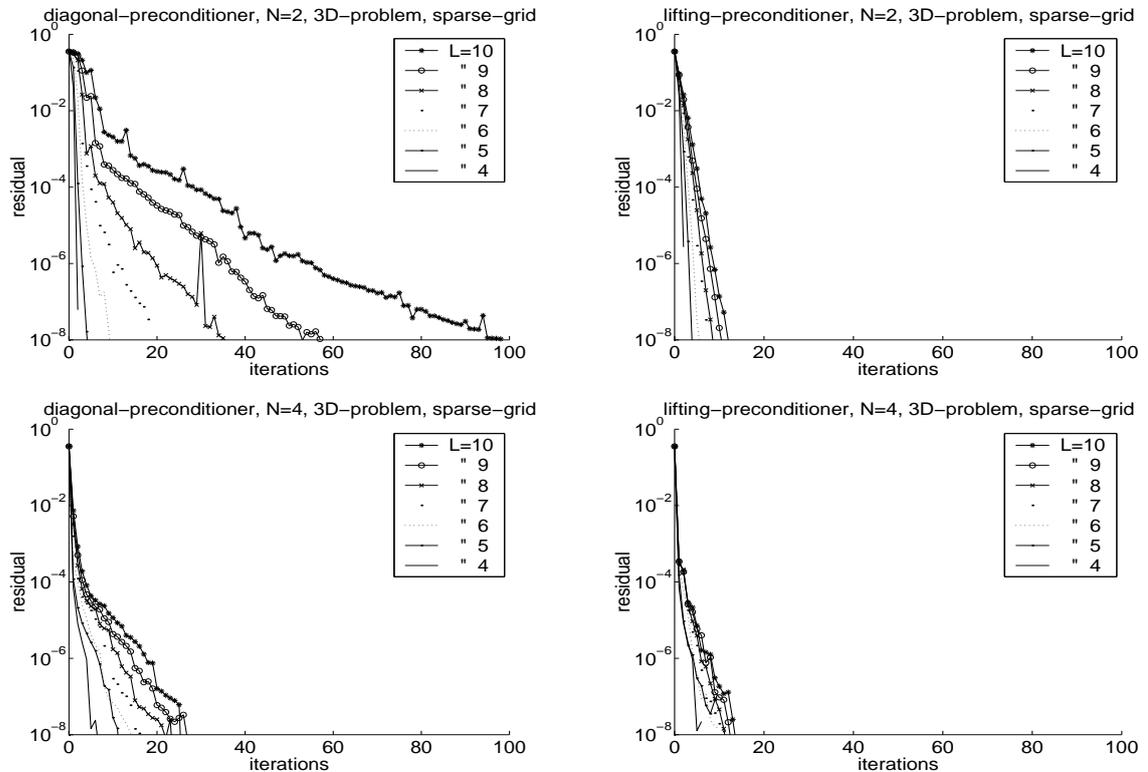


Abbildung 7.2: Links: Diagonal-, Rechts: Lifting-Vorkonditionierer. Erste und dritte Zeile: $N = 2$. Zweite und vierte Zeile $N = 4$. Problem (7.17) mit $d = 3$.

Kapitel 8

Elliptische Gleichungen

In diesem Kapitel wird eine Verfeinerungsstrategie für die adaptive Lösung von elliptischen Differentialgleichungen mit den Dünngitterdiskretisierungen aus Kapitel 6 vorgestellt. Zunächst betrachten wir in Abschnitt 8.1 die Prinzipien einiger Verfeinerungsstrategien bei Finite-Element- oder Wavelet-Galerkin-Verfahren nach z.B. [BEK96, Doe96, CDD01]. Algorithmisch bestehen diese Schemen immer aus dem iterierten Lösen und Verfeinern in Abhängigkeit von der aktuellen Lösung. Obgleich es auch schon für Finite-Element-Methoden gelungen war, die Konvergenz solcher Verfeinerungsschemen zu zeigen [Doe96], scheint das Wavelet-basierte Verfahren aus [CDD01] das erste Verfahren zu sein, für das auch gezeigt wurde, daß die Konvergenzraten im Sinne der Best- n -Term-Approximation optimal sind. Der entscheidende Trick dabei ist, daß es unter gewissen Voraussetzungen möglich ist, das aktuelle Residuum bis zu einer vorher festgelegten garantierten Genauigkeit zu berechnen. Die Theorie und die konkrete Steuerung des Verfahrens basieren auf der Galerkin-Orthogonalität des Fehlers, der Stabilität der Wavelets und die entsprechenden Konstanten in den Norm-Abschätzungen sowie auf a priori Informationen über die Größe der einzelnen Waveletkoeffizienten der rechten Seite und des Operators. Es ist somit nicht möglich, diese Strategie unmittelbar auf die Finite-Differenzen- oder Kollokationsdiskretisierungen zu übertragen, da weder die Galerkin-Orthogonalität noch die Stabilität der Ansatzfunktionen erfüllt ist. Vielmehr benutzen wir für die in Abschnitt 8.2 beschriebene Verfeinerungsstrategie das bekannte Schema, mittels eines Schwellwertes signifikante Indizes zu bestimmen und diese sowie ihre Nachbarn einzufügen [Har95, Bun98, Sch98a]. Dieses Schema kann in gewissem Sinne als eine vereinfachte Version von [CDD01] verstanden werden. In den Abschnitten 8.3 und 8.4 benutzen wir das adaptive Schema für die Behandlung von Poisson-Problemen mit einer quasi-singulären bzw. einer glatten Lösung. Für die glatte Lösung wird die optimale Konvergenzrate erreicht, bei der quasi-singulären Lösung gelingt dies nur für die Finite-Differenzen-Diskretisierung mit dem AFD Algorithmus.

8.1 Verfeinerungsstrategien für Galerkin-Verfahren

Wir betrachten das Modell-Problem

$$Au = f \quad \text{in }]0, 1[^d \quad \text{sowie} \quad u = 0 \quad \text{auf} \quad \partial[0, 1]^d, \quad (8.1)$$

wobei A ein symmetrischer $\mathcal{H}^{r/2}$ -elliptischer Operator mit Energie-Norm $\|u\|_A^2 := \langle u, Au \rangle$ ist. Diese Norm ist somit äquivalent zur $\mathcal{H}^{r/2}$ -Norm, d.h. $\|u\|_A \sim \|u\|_{\mathcal{H}^{r/2}}$.

Durch einen iterativen Verfeinerungsalgorithmus werden sukzessive endlich-dimensionale Räume $\dots \subset \mathcal{V}^k \subset \mathcal{V}^{k+1} \subset \dots$ aus Finite-Element- oder Wavelet-Ansatzfunktionen sowie Galerkin-Lösungen $u^k \in \mathcal{V}^k$ generiert, siehe Bild 8.1.

Allgemeines iteratives Verfeinerungsschema

Wähle einen geeigneten ersten Ansatzraum \mathcal{V}^1 und setze $k := 1$
do
 Löse $\langle v, Au^k \rangle = \langle v, f \rangle \quad \forall v \in \mathcal{V}^k$
 Verfeinere mittels $u^k: \mathcal{V}^k \rightarrow \mathcal{V}^{k+1}$
 $k := k + 1$
while ($\|u - u^k\|_A > \epsilon$)
 $k_{end} := k$

Abbildung 8.1: *Allgemeines iteratives Verfeinerungsschema.*

Für ein solches Schema benötigt man einen Fehlerschätzer, der einerseits bestimmt, wann die Iteration terminiert und andererseits auch den Verfeinerungsschritt $\mathcal{V}^k \rightarrow \mathcal{V}^{k+1}$ steuert. Ideal wäre, wenn dabei die Ansatzräume \mathcal{V}^k so generiert würden, daß das Verfahren im Sinne der Best- n -Term-Approximation mit optimaler Rate konvergiert, z.B.

$$\|u - u^k\|_A \sim \|u - u^k\|_{\mathcal{H}^{r/2}} \lesssim (\dim \mathcal{V}^k)^{-\frac{s-r/2}{d}} \|u\|_{B_{\mathcal{H}^{r/2}}^s}. \quad (8.2)$$

Bei Verfeinerungsstrategien für Finite-Elemente findet man oft direkt oder indirekt das folgende Prinzip [BEK96]:

1. Ausgehend von \mathcal{V}^k wählt man einen erweiterten Ansatzraum $\mathcal{V}^k \oplus \mathcal{Q}$. Üblicherweise nach einem von u^k unabhängigen Schema. Beispiele sind die Hinzunahme von lokalen Bubble-Funktionen (stückweise definierte Polynome höheren Grades) für jedes Dreieck [BR78] oder die konforme Unterteilung jedes Dreiecks und die Hinzunahme von entsprechenden hierarchischen Hutfunktionen [BS93].
2. In einem zweiten Schritt wird der Defekt $d^{\mathcal{Q}} \in \mathcal{V}^k \oplus \mathcal{Q}$, der über das Residuum $r^k := f - Au^k$ durch $\langle v, Ad^{\mathcal{Q}} \rangle = \langle v, r^k \rangle \quad \forall v \in \mathcal{V}^k \oplus \mathcal{Q}$ definiert ist, näherungsweise durch lokale, entkoppelte Probleme gelöst. Wurden beispielsweise für Dreieck T die Bubble-Funktionen b_1, \dots, b_p gewählt, dann wird der lokale Anteil $d_T^{\mathcal{Q}}$ des Defekts durch $\langle v, Ad_T^{\mathcal{Q}} \rangle = \langle v, r^k \rangle \quad \forall v \in \text{span}\{b_1, \dots, b_p\}$ berechnet. Für die üblichen Wahlen von \mathcal{Q} kann gezeigt werden, daß $\|d^{\mathcal{Q}}\|_A^2 \sim \sum_T \|d_T^{\mathcal{Q}}\|_A^2$. Letzterer Term ist numerisch auswertbar. Für eine Abschätzung des aktuellen Fehlers $u - u^k$ nimmt man nun an, daß die Räume $\mathcal{V}^k \oplus \mathcal{Q}$ die sogenannte *Saturationseigenschaft* erfüllen. D.h. es gibt $\theta < 1$, so daß für die Galerkin Lösung $u^{\mathcal{Q}} := u^k + d^{\mathcal{Q}}$ in $\mathcal{V}^k \oplus \mathcal{Q}$ gilt

$$\|u - u^{\mathcal{Q}}\|_A^2 \leq \theta \|u - u^k\|_A^2 \quad (\text{Saturationseigenschaft}). \quad (8.3)$$

Mittels dieser Kontraktion und der Galerkin-Orthogonalität $\|u - u^k\|_A^2 = \|u - u^{\mathcal{Q}}\|_A^2 + \|u^{\mathcal{Q}} - u^k\|_A^2$ folgt

$$(1 - \theta) \|u - u^k\|_A^2 \leq \|d^{\mathcal{Q}}\|_A^2 \lesssim \sum_T \|d_T^{\mathcal{Q}}\|_A^2. \quad (8.4)$$

D.h. man erhält eine Abschätzung für den aktuellen Fehler. Natürlich benötigt man z.B. für ein verlässliches Abbruchkriterium in Algorithmus 8.1 noch quantitative Aussagen über θ oder die in \lesssim verborgenen Konstanten, siehe [BEK96]. Üblicherweise kann man zeigen, daß (8.3) erfüllt ist, falls der Ansatzraum \mathcal{V}^k hinreichend groß ist.

3. Mit der obigen Abschätzung (8.4) kann nicht nur der Abbruch der Iteration sondern auch die lokale Verfeinerung gesteuert werden. Dazu nimmt man an, daß für lokale Fehler $\|(u - u^k)|_T\|_A \lesssim \|d_T^{\mathcal{Q}}\|_A$ gilt. Da A elliptisch ist, kann der lokale Fehler wiederum durch Verfeinern von T reduziert werden. Im Sinne einer Optimierung von Aufwand zu Genauigkeit ist es nun erstrebenswert, daß die lokalen Beiträge $\|d_T^{\mathcal{Q}}\|_A^2$ für alle Dreiecke T ungefähr gleich groß sind [BR78]. Deshalb verfeinert man z.B. alle Dreiecke mit einem besonders großen lokalen Beitrag. Auf diese Weise erhält man aus u^k und \mathcal{V}^k die neue Indexmenge \mathcal{V}^{k+1} .

Mit Hilfe von sogenannten dualen Fehlerschätzern ist es auch möglich, die Verfeinerung so zu steuern, daß ein bestimmtes Fehlerfunktional, wie Funktionswerte/Gradienten in einem Punkt, minimiert werden, siehe [BR96].

Während für $u^{\mathcal{Q}}$ aus der Saturationseigenschaft (8.3) sofort die Kontraktion $\|u - u^{\mathcal{Q}}\|_A \leq \sqrt{\theta}\|u - u^k\|_A$ folgt, ist jedoch a priori unklar, ob für die Galerkin-Lösung u^{k+1} bzgl. \mathcal{V}^{k+1} auch $\|u - u^{k+1}\|_A < \|u - u^k\|_A$ gilt. Letzten Endes ist deshalb die Konvergenz des Verfeinerungsalgorithmus nicht abgesichert. In [Doe96] ist aber unter einer zur Saturationseigenschaft analogen Bedingung gezeigt worden, daß zumindest für die dort beschriebene Modifikation der obigen Verfeinerungsstrategie die Konvergenz garantiert ist. D.h. es gibt ein $\kappa < 1$, das unabhängig von z.B. der Maschenweite ist, so daß $\|u - u^{k+1}\|_A < \kappa\|u - u^k\|_A$. Offen bleibt aber die genaue Größe von κ bzw. die tatsächliche Konvergenzrate. Insbesondere, ob das Verfahren mit der optimalen Rate im Sinne von (8.2) konvergiert.

Ein Verfahren, das diese Bedingung erfüllt ist in [CDD01] für $\mathcal{H}^{r/2}$ stabile Wavelets angegeben worden. Dieses soll nachfolgend stark vereinfacht skizziert werden. Wir bezeichnen mit $\{\psi_\lambda\}_{\lambda \in \Lambda}$ bzw. $\{\tilde{\psi}_\lambda\}_{\lambda \in \Lambda}$ wieder die generischen (dualen) Wavelets. Diese Wavelets sind *keine* hierarchischen Wavelets, die ja nicht $\mathcal{H}^{r/2}$ -stabil sind. Insbesondere sind die dualen Wavelets mindestens aus \mathcal{L}^2 . Die Ansatzräume \mathcal{V}^k sind dann durch Indexmengen Λ^k bestimmt: $\mathcal{V}^k = \text{span}\{\psi_\lambda\}_{\lambda \in \Lambda^k}$. Mit u^k werden wieder die entsprechenden Galerkin-Lösungen bezeichnet. Im Gegensatz zur Finite-Element-Vorgehensweise wird nun direkt aus Λ^k und u^k eine Indexmenge Λ^{k+1} konstruiert, so daß

$$\|u - u^{k+1}\|_A^2 \leq \kappa \|u - u^k\|_A^2 ,$$

für ein $\kappa < 1$. Die dafür entscheidende Aussage ist:

Lemma 8.1 ([CDD01]) *Sei $\Lambda^{k+1} \supset \Lambda^k$ eine Indexmenge, so daß die Projektion $I_{\Lambda^{k+1}}^* r^k = \langle \Psi_{\Lambda^{k+1}}, r^k \rangle \cdot \tilde{\Psi}_{\Lambda^{k+1}}$ des Residuums $r^k = f - Au^k$ hinreichend gut mit r^k übereinstimmt, genauer*

$$\|r^k - I_{\Lambda^{k+1}}^* r^k\|_{\mathcal{H}^{-r/2}} \leq \frac{1}{2} \|r^k\|_{\mathcal{H}^{-r/2}} , \quad (8.5)$$

dann gilt für die Galerkin-Lösung u^{k+1} bzgl. Λ^{k+1}

$$\|u - u^{k+1}\|_A^2 \leq \left(1 - \frac{\underline{c}}{2\bar{c}}\right) \|u - u^k\|_A^2 .$$

Dabei sind \underline{c} bzw. \bar{c} die die Elliptizitätskonstanten von A , $\underline{c}\|u\|_{\mathcal{H}^{r/2}}^2 \leq \|u\|_A^2 \leq \bar{c}\|u\|_{\mathcal{H}^{r/2}}^2$.

Für einen Beweis siehe Anhang A.17. Schranken für $\|r^k\|_{\mathcal{H}^{-r/2}}$ können im Algorithmus mitberechnet werden [CDD01]. Man nimmt außerdem an, daß a priori Abschätzungen für die Koeffizienten der rechten Seite, $|\langle \psi_\lambda, f \rangle| \leq f_\lambda$, bzw. für die Koeffizienten der Operatormatrix, $|\langle \psi_\lambda, A\psi_\mu \rangle| \leq a_{\lambda\mu}$, bekannt sind [Dah97]. Mit diesen a priori Informationen kann Λ^{k+1} so bestimmt werden, daß (8.5) erfüllt ist. Das läuft dann auf so etwas wie

$$\Lambda^{k+1} = \Lambda^k \cup \left\{ \lambda : 2^{-r|\lambda|} \left(f_\lambda^2 + \left(\sum_{\mu \in \Lambda^k} a_{\lambda\mu} |u_\mu^k|^2 \right) \right) > \delta \right\} \quad (8.6)$$

hinaus, wobei δ eine von $\|r^k\|_{\mathcal{H}^{-r/2}}$ abhängige Schranke ist. Wir wollen nachfolgend das FE- und das Wavelet-Verfahren miteinander vergleichen.

Bemerkung 8.1 (Saturationseigenschaft) Für die FE-Strategie war zu fordern, daß \mathcal{V}^k hinreichend groß ist, damit die Saturationseigenschaft (8.3) erfüllt ist. Für das Wavelet-Verfahren gibt es formal keine solche Bedingung. Allerdings benötigt man hier die Abschätzungen für $\langle \psi_\lambda, f \rangle$, so daß letztlich *a priori* bekannt sein muß, wo und welche Singularitäten auf Grund von f in der Lösung u auftreten können bzw. wo und wie wegen f verfeinert werden muß. Konkret benötigt man eine Vorschrift, die explizit alle Indizes λ liefert, für die z.B. f_λ multipliziert mit dem Level-abhängigen Gewichtungsfaktor größer als eine vorgegebene Schranke ist.

Bemerkung 8.2 (Defektgleichung) Bei der FE-Strategie basiert der Fehlerschätzer auf Approximationen d_T^Q des lokalen Defekts $(u - u^k)|_T$. Diese werden mittels entkoppelter Gleichungssysteme berechnet. Dort wo $\|d_T^Q\|_A^2$ groß ist, wird verfeinert. Bei der Wavelet-Strategie geschieht im Grunde das Gleiche: Wegen der Normäquivalenzen gilt

$$\|u - u^k\|_A^2 = \|A^{-1}r^k\|_A^2 = \langle r^k, A^{-1}r^k \rangle \sim \sum_{\lambda \notin \Lambda^k} 2^{-r|\lambda|} |\langle \psi_\lambda, r^k \rangle|^2$$

D.h. auch hier wird der Fehler durch eine Summe lokaler Beiträge abgeschätzt, vergleiche (8.4). Überall wo die lokalen Beiträge groß sind, sollte verfeinert werden. Bei der Definition (8.6) von Λ^{k+1} werden nun die lokalen Beiträge $|\langle \psi_\lambda, r^k \rangle|^2$ durch einfacher zu berechnende Schranken $(f_\lambda^2 + (\cdot)^2)$ ersetzt.

Bemerkung 8.3 (Verfeinerung) Bei der FE-Strategie wird in jeder Iteration lokal maximal einmal verfeinert, so daß das Verfeinerungslevel maximal um eins größer wird. Bei der Wavelet-Strategie kann dagegen ein großer Koeffizient u_λ^k via (8.6) das Einfügen von Indizes auf deutlich höheren Leveln induzieren. Das maximale Verfeinerungslevel kann somit sehr schnell ansteigen. Es ist aber zu vermuten, daß nach einer Startphase ein Einfügen von neuen Indizes nur noch in der unmittelbaren Nachbarschaft von schon aktiven Indizes erfolgt, da die großen Koeffizienten u_λ^k auf niedrigen Leveln leben dürften. In diesem Fall gibt es keine großen Unterschiede mehr zum einfachen Abschneiden mit Schwellwert und Einfügen von Nachbarn mehr.

Bis hier hin ist wegen der Kontraktion des Fehlers in Lemma 8.1 klar, daß das Verfeinerungsschema auf jeden Fall konvergiert. Unklar ist, ob die optimale Konvergenzrate im Sinne von (8.2) erreicht wird. Um dies zu prüfen, betrachtet man $C_k := \|u - u^k\|_A \cdot |\Lambda_k|^{\frac{s-r/2}{d}}$. Falls diese Terme gleichmäßig beschränkt sind, würde das Verfahren mit optimaler Rate konvergieren. Wir benötigen eine Abschätzung für $|\Lambda_k|$. Unter der Voraussetzung, daß die Schranken f_λ bzw. $a_{\lambda\mu}$ hinreichend genau das Verhalten der Koeffizienten der rechten Seite bzw. des Operators beschreiben, können wir annehmen, daß mit (8.6) nahezu die kleinste Indexmenge

bestimmt wird, für die $\|r^k - I_{\Lambda_{k+1}}^* r^k\|_{\mathcal{H}^{-r/2}} \leq 1/2 \|r^k\|_{\mathcal{H}^{-r/2}}$ ist, so daß man mittels einer geeigneten Besov-Norm zu $\frac{1}{2} \|r^k\|_{\mathcal{H}^{-r/2}} \lesssim \|r^k\|_{\mathcal{B}_{\mathcal{H}^{-r/2}}^{s-r}} |\Lambda_{k+1}|^{-\frac{s-r/2}{d}}$ kommt. Daraus folgt

$$\begin{aligned} C_{k+1} = \|u - u^{k+1}\|_A \cdot |\Lambda_{k+1}|^{\frac{s-r/2}{d}} &\lesssim \sqrt{1 - \frac{\underline{c}}{2\bar{c}}} \|u - u^k\|_A \cdot \frac{2 \|r^k\|_{\mathcal{B}_{\mathcal{H}^{-r/2}}^{s-r}}}{\|r^k\|_{\mathcal{H}^{-r/2}}} \\ &\lesssim \sqrt{1 - \frac{\underline{c}}{2\bar{c}}} \cdot \frac{2}{\underline{c}} \|r^k\|_{\mathcal{B}_{\mathcal{H}^{-r/2}}^{s-r}} . \end{aligned}$$

Für den letzten Schritt sei auf Anhang A.17 verwiesen. Die obige Abschätzung bedeutet, daß die optimale Konvergenzrate erreicht würde, falls die Besov-Normen der Residuen gleichmäßig beschränkt wären. Das zweite Merkmal des Algorithmus aus [CDD01] ist nun, daß durch einen Aufräumschritt, bei dem unwichtige Waveletkoeffizienten von u^k abgeschnitten werden, genau diese Normen $\|r^k\|_{\mathcal{B}_{\mathcal{H}^{-r/2}}^{s-r}}$ gleichmäßig beschränkt werden können. Führt man einen solches Abschneiden alle K Iterationen aus, dann wird der Fehler zunächst K -mal um der Faktor $\theta := \sqrt{1 - \underline{c}/(2\bar{c})}$ reduziert und dann bei geeignetem Abschneiden um maximal einen Faktor $\gamma > 1$ vergrößert. Deshalb muß K so gewählt werden, daß $\theta^K \gamma < 1$. Hier ist ein genaues Balancieren des Verfahrensparameters γ notwendig.

8.2 Verfeinerungsstrategien für Dünngitter-Diskretisierungen

Auch für die Finite-Differenzen- oder die Kollokationsmethode auf dünnen Gittern kann ein iteratives Verfeinerungsschema analog zu Abbildung 8.1 für das adaptive Lösen des Modellproblems (8.1) benutzt werden. Wir bezeichnen mit $\{\mathcal{T}^k\}$ die Folge der dabei sukzessive generierten Indexmengen und mit $u^k = \Psi_{\mathcal{T}^k} \cdot \mathbf{u}_{\mathcal{T}^k}^k$ die diskrete Lösung des k -ten Iterationsschrittes. Gegenstand dieses Abschnittes sind nun Verfeinerungsstrategien mit denen aus \mathcal{T}^k und u^k die nächste Indexmenge \mathcal{T}^{k+1} erzeugt werden kann. Leider kann man dafür die ausgefeilte Strategie aus [CDD01] nicht direkt übernehmen, da es z.B. keine Galerkin-Orthogonalität des Fehlers gibt. Wir können jedoch die Argumentation bzgl. des Schätzens des Residuums übertragen, da dieses Problem weitgehend unabhängig von der eigentlichen Diskretisierung ist. Aus den Abschätzungen kann dann abgelesen werden, welche Bereiche des aktuellen Ansatzraumes besonders stark zum Fehler beitragen und wo und wie demnach verfeinert werden sollte.

Als Beispiel betrachten wir das Poisson-Problem mit homogenen Dirichlet-Randwerten auf dem Gebiet $[0, 1]^d$. Der Laplace-Operator ist dann im Sinne von Definition 6.2 stabil und stetig, so daß für $0 \leq s \leq 2$ gilt

$$\|u\|_{\mathcal{H}^s} \sim \|Au\|_{\mathcal{H}^{s-2}} .$$

Damit sind Fehlerabschätzungen in z.B. der \mathcal{H}^0 -Norm äquivalent zum Schätzen des aktuellen Residuums $r^k := f - Au^k$ in der \mathcal{H}^{-2} -Norm, d.h. $\|u - u^k\|_{\mathcal{H}^0} \sim \|r^k\|_{\mathcal{H}^{-2}}$. Das Residuum $\|r^k\|_{\mathcal{H}^{-2}}$ soll mittels einseitiger Normabschätzungen über die Multiskalenkoeffizienten von r^k kontrolliert werden. Damit dies für die \mathcal{H}^{-2} -Norm möglich ist, wird r^k in einer Basis von Lifting-Wavelets mit mindestens drei verschwindenden Momenten dargestellt, vergleiche Lemma 4.1

$$r^k = \sum_{(\mathbf{1}, \mathbf{t})} \bar{r}_{(\mathbf{1}, \mathbf{t})}^k \bar{\psi}_{(\mathbf{1}, \mathbf{t})} .$$

Unter der Voraussetzung, daß die Lifting-Waveletkoeffizienten $\langle \tilde{\psi}_{(\mathbf{l}, \mathbf{t})}, f \rangle$ beliebig genau berechnet werden können, kann dann $\|r^k\|_{\mathcal{H}^{-2}}$ in einen exakt berechenbaren und einen abzuschätzenden Anteil zerlegt werden:

$$\|r^k\|_{\mathcal{H}^{-2}} \leq \underbrace{\|r^k - \bar{I}_\Lambda r^k\|_{\mathcal{H}^{-2}}}_{\text{abzuschätzen}} + \underbrace{\|\bar{I}_\Lambda r^k\|_{\mathcal{H}^{-2}}}_{\text{berechenbar}} .$$

Dabei ist $\bar{I}_\Lambda v := \langle \tilde{\Psi}_\Lambda, v \rangle \cdot \bar{\Psi}_\Lambda$ die biorthogonale Projektion auf einen noch zu bestimmenden endlich-dimensionalen Ansatzraum von Lifting-Wavelets. Entsprechend der Abschätzung

$$\|r^k - \bar{I}_\Lambda r^k\|_{\mathcal{H}^{-2}}^2 \lesssim \sum_{(\mathbf{l}, \mathbf{t}) \notin \Lambda} |\bar{r}_{(\mathbf{l}, \mathbf{t})}^k|^2 4^{-2|\mathbf{l}|_\infty} 2^{-|\mathbf{l}|_1}$$

wird Λ so gewählt, daß alle Indizes (\mathbf{l}, \mathbf{t}) mit großen Koeffizienten $|\bar{r}_{(\mathbf{l}, \mathbf{t})}^k|^2 4^{-2|\mathbf{l}|_\infty} 2^{-|\mathbf{l}|_1/2}$ enthalten sind. Genau wie auch in [CDD01] benötigt man für Abschätzungen von $|\bar{r}_{(\mathbf{l}, \mathbf{t})}^k|$ scharfe Abschätzungen für die Lifting-Waveletkoeffizienten der rechten Seite

$$|\langle \tilde{\psi}_{(\mathbf{l}, \mathbf{t})}, f \rangle| \leq f_{\mathbf{l}\mathbf{t}} , \quad (8.7)$$

was in praktischen Anwendungen etwas problematisch ist, wie bereits in Bemerkung 8.1 erwähnt. Außerdem benötigt man auch Abschätzungen für den Anteil von Δu^k am Residuum. Unter Ausnutzung der Tensorproduktstruktur der Basisfunktionen und mittels Lemma 6.6 erhält man für die Operatorkoeffizienten

$$|\langle \tilde{\psi}_{(\mathbf{l}, \mathbf{t})}, \Delta \psi_{(\mathbf{k}, \mathbf{s})} \rangle| \lesssim 2^{2|\mathbf{k}|_\infty} 2^{(|\mathbf{l}|_1 - |\mathbf{k}|_1)/2} a_{\mathbf{l}\mathbf{t}, \mathbf{k}\mathbf{s}} \quad (8.8)$$

wobei $a_{\mathbf{l}\mathbf{t}, \mathbf{k}\mathbf{s}} := \begin{cases} 0 & \text{falls } \text{supp } \tilde{\psi}_{(\mathbf{l}, \mathbf{t})} \cap \text{supp } \psi_{(\mathbf{k}, \mathbf{s})} = \emptyset \\ \sum_{i=1}^d 2^{-\min(\gamma^* - 2, \tilde{\gamma}^*)|l_i - k_i|} \prod_{j \neq i} 2^{-\gamma^* \max(l_j - k_j, 0)} \end{cases} .$

Die Werte γ^* bzw. $\tilde{\gamma}^*$ sind die Sobolev-Regularitäten der Interpolets bzw. der dualen Lifting-Wavelets. Wir nehmen hier an, daß Interpolets der Ordnung $N \geq 4$ benutzt werden, so daß $\gamma^* > 2$ und $\tilde{\gamma}^* > 0$. Zusammen mit (8.7) erhält man daraus für die Multiskalenkoeffizienten des Residuums

$$|\bar{r}_{(\mathbf{l}, \mathbf{t})}^k| 2^{-2|\mathbf{l}|_\infty - |\mathbf{l}|_1/2} \leq 2^{-2|\mathbf{l}|_\infty - |\mathbf{l}|_1/2} f_{\mathbf{l}\mathbf{t}} + \sum_{(\mathbf{k}, \mathbf{s}) \in \mathcal{T}^k} 2^{|\mathbf{k}|_\infty - |\mathbf{l}|_\infty} a_{\mathbf{l}\mathbf{t}, \mathbf{k}\mathbf{s}} |u_{(\mathbf{k}, \mathbf{s})}^k| 2^{-|\mathbf{k}|_1/2} \quad (8.9)$$

Die Indexmenge Λ kann nun so gewählt werden, daß z.B. alle Indizes (\mathbf{l}, \mathbf{t}) enthalten sind, für die die rechte Seite von (8.9) größer als ein vorgegebener Schwellwert ist, der den Fehler $\|r^k - \bar{I}_\Lambda r^k\|_{\mathcal{H}^{-2}}^2$ bestimmt. Für Einzelheiten, wie dies in einem Algorithmus realisiert werden kann, sei wieder auf [CDD01] verwiesen.

Für die Verfeinerungsstrategie $\mathcal{T}^k \rightarrow \mathcal{T}^{k+1}$ nehmen wir gegenüber der Bestimmung von Λ einige Vereinfachungen vor. Zum einen wollen wir auf die Abschätzung $f_{\mathbf{l}\mathbf{t}}$ für die rechte Seite verzichten, da diese für jede rechte Seite erst aufwendig hergeleitet werden müsste. Zum zweiten wollen wir immer nur unmittelbare Nachbarn $\mathcal{N}(\mathbf{k}, \mathbf{s})$ von bestimmten signifikanten Indizes $\mathcal{T}_s \subseteq \mathcal{T}^k$ in \mathcal{T}^{k+1} einfügen:

$$\mathcal{T}^{k+1} := \mathcal{T}^k \cup \bigcup_{(\mathbf{k}, \mathbf{s}) \in \mathcal{T}_s} \mathcal{N}(\mathbf{k}, \mathbf{s}) , \quad (8.10)$$

$$\text{wobei } \mathcal{N}(\mathbf{k}, \mathbf{s}) := \{(\mathbf{k}, \mathbf{t}) : |\mathbf{t} - \mathbf{s}|_\infty \leq R_{\text{ort}}\} \cup \{(\mathbf{l}, \mathbf{t}) : \mathbf{l} \geq \mathbf{k}, |\mathbf{l} - \mathbf{k}|_\infty \leq R_{\text{level}} \text{ und } |\mathbf{2t} - \mathbf{s}|_\infty \leq R_{\text{ort}}\} .$$

R_{ort} und R_{level} sind vorgegebene Parameter. Wir haben in unseren Experimenten $R_{ort} = R_{level} = 1$ gewählt. Im Allgemeinen kann man natürlich auch größere Werte nehmen bzw. die Werte sogar in Abhängigkeit von $|u_{(\mathbf{k},\mathbf{s})}^k|$ wählen, wie dies z.B. in [CKMP00] vorgeschlagen wird. Zumindest die Wahl $R_{ort} = R_{level} = 2$ führt jedoch nach unseren Erfahrungen eher zu etwas schlechteren als besseren Ergebnissen, da die Gefahr größer ist, daß überflüssige Indizes eingefügt werden.

Entsprechend den Abschätzungen (8.9) und (8.8) bietet es sich an, für \mathcal{T}_s in (8.10) die Indizes (\mathbf{k}, \mathbf{s}) zu wählen, für die $2^0 a_{\mathbf{k},\mathbf{s}} |u_{(\mathbf{k},\mathbf{s})}^k| 2^{-|\mathbf{k}|/2} = |u_{(\mathbf{k},\mathbf{s})}^k| 2^{-|\mathbf{k}|/2}$ signifikant ist, also

$$\mathcal{T}_s = \{(\mathbf{k}, \mathbf{s}) \in \mathcal{T}^k : |u_{(\mathbf{k},\mathbf{s})}^k| 2^{-|\mathbf{k}|/2} > \delta\} . \quad (8.11)$$

Der Schwellwert δ kontrolliert auch hier wieder die Zielgenauigkeit. Das Verfeinerungskriterium (8.11) kann man auch sehr viel einfacher motivieren. Entsprechend der einseitigen Normabschätzungen aus Lemma 4.1 haben quasi-optimale Indexmengen für die Approximation der exakten Lösung $u = \sum_{(\mathbf{k},\mathbf{s})} u_{(\mathbf{k},\mathbf{s})} \psi_{(\mathbf{k},\mathbf{s})}$ in der \mathcal{L}^2 -Norm bei einem Schwellwert-Ansatz die Form $\{(\mathbf{k}, \mathbf{s}) : |u_{(\mathbf{k},\mathbf{s})}^k| 2^{-|\mathbf{k}|/2} > \delta\}$. Das Kriterium (8.11) kann also so interpretiert werden, daß damit versucht wird, solche quasi-optimale Indexmengen anzustreben. Eine Überlegung ist nun, daß das Verfeinerungskriterium nicht auf den Koeffizienten $u_{(\mathbf{k},\mathbf{s})}$ bzgl. der Interpolets basieren sollte, sondern auf Koeffizienten bzgl. Lifting-Interpolets, da man für diese ja schärfere einseitige Normabschätzungen ohne logarithmische Vorfaktoren hat. In diesem Fall wäre also

$$\mathcal{T}_s = \{(\mathbf{k}, \mathbf{s}) : |\bar{u}_{(\mathbf{k},\mathbf{s})}^k| 2^{-|\mathbf{k}|/2} > \delta\} \quad \text{wobei} \quad u^k = \sum_{(\mathbf{k},\mathbf{s}) \in \mathcal{T}^k} \bar{u}_{(\mathbf{k},\mathbf{s})}^k \bar{\psi}_{(\mathbf{k},\mathbf{s})} . \quad (8.12)$$

Die Indexmengen $\{\mathcal{T}^k\}$ müssen dann auch die multivariate Lifting-Bedingung erfüllen. Eine weitere Überlegung ist, daß das Verfeinerungskriterium nicht auf Fehlerabschätzungen in der \mathcal{L}^2 -Norm, sondern auf Fehlerabschätzungen in z.B. der \mathcal{L}^∞ -Norm basieren sollte. Da man für die exakte Lösung u und einen gegebenen endlich-dimensionalen Interpolanten $\Psi_{\mathcal{T}} \cdot \mathbf{u}_{\mathcal{T}}$ die Abschätzung

$$\|u - \Psi_{\mathcal{T}} \cdot \mathbf{u}_{\mathcal{T}}\|_\infty \leq \sup_{\mathbf{x}} \sum_{(\mathbf{1},\mathbf{t}) \notin \mathcal{T} : \mathbf{x} \in \text{supp} \psi_{(\mathbf{1},\mathbf{t})}} |u_{(\mathbf{1},\mathbf{t})}|$$

hat, bietet sich hier

$$\mathcal{T}_s := \{(\mathbf{k}, \mathbf{s}) : |u_{(\mathbf{k},\mathbf{s})}^k| > \delta\} \quad (8.13)$$

an. Auch dabei könnte man gegebenenfalls $|u_{(\mathbf{k},\mathbf{s})}^k|$ durch $|\bar{u}_{(\mathbf{k},\mathbf{s})}^k|$ ersetzen. Allgemein erhält man somit das folgende iterative Verfeinerungsschema:

Iteratives Verfeinerungsschema für adaptive dünne Gitter

Wähle eine geeignete Indexmenge \mathcal{T}^1 , $k := 1$

do

Löse $A_{\mathcal{T}^k} u^k = I_{\mathcal{T}^k} f$

Bestimme \mathcal{T}_s mittels (8.11) oder (8.12) oder (8.13)

$\mathcal{T}^{k+1} := \mathcal{T}^k \cup \bigcup_{(\mathbf{1},\mathbf{t}) \in \mathcal{T}_s} \mathcal{N}(\mathbf{1}, \mathbf{t})$

Auffüllen von \mathcal{T}^{k+1} für die Kegel-/Lifting-Bedingung

$k := k + 1$

while ($\mathcal{T}^k \neq \mathcal{T}^{k-1}$)

$k_{end} := k$

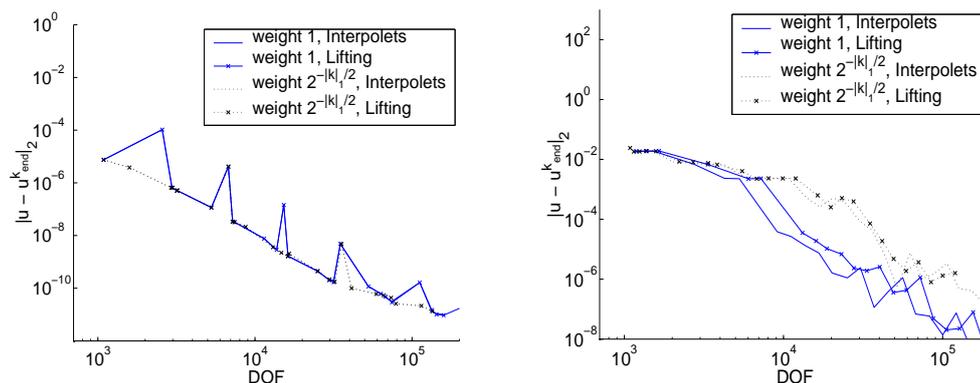


Abbildung 8.2: Vergleich der Ergebnisse des adaptiven Verfeinerungsschemas mit einer Finite-Differenzen-Diskretisierung (AFD, Konsistenzordnung 4) und Interpolets $N = 4$ für die unterschiedlichen Verfeinerungskriterien. Links: Poisson-Problem mit glatter Lösung. Rechts: quasi-singuläre Lösung.

Dieses Schema wurde für die verschiedenen Verfeinerungskriterien und verschiedene Schwellwerte δ auf Poisson-Probleme mit einer glatten bzw. einer quasi-singulären Lösung

$$u(x, y) = \sin(2\pi x) \sin(2\pi y) \quad \text{bzw.} \quad u(x, y) = (10^{-10} + (x - \frac{3}{9})^2 + (y - \frac{4}{9})^2)^{1/4}$$

angewendet. In Bild 8.2 sind jeweils die Fehler in der \mathcal{L}^2 -Norm gegen die Zahl der Freiheitsgrade der finalen Lösung u^k_{end} aufgetragen. Für Einzelheiten der Rechnungen sowie Darstellungen für andere Fehlernormen sei auf die nächsten beiden Abschnitte verwiesen. Sowohl für die glatte als auch für die quasi-singuläre Lösung gibt es jeweils kaum einen Unterschied, ob Interpolet-Koeffizienten $|u^k_{(\mathbf{k},\mathbf{s})}|$ oder Lifting-Waveletkoeffizienten $|\bar{u}^k_{(\mathbf{k},\mathbf{s})}|$ für das Verfeinerungskriterium benutzt wurden. Man braucht mit den Lifting-Waveletkoeffizienten sogar etwas mehr Freiheitsgrade, um die gleiche Genauigkeit zu erreichen, wie mit den Interpolet-basierten Kriterien, da die Indextmengen zusätzlich die Lifting-Bedingung erfüllen müssen. Der geringe Unterschied zwischen den Ergebnissen kann damit erklärt werden, daß zu einem guten Teil die Blätter von \mathcal{T}^k die Verfeinerung bestimmen. Ein Blatt ist dabei ein Index (\mathbf{k}, \mathbf{s}) , für den es *keinen* Index (\mathbf{l}, \mathbf{t}) gibt mit $\mathbf{l} \geq \mathbf{k}, \mathbf{l} \neq \mathbf{k}$ und $\text{supp}\psi_{(\mathbf{l},\mathbf{t})} \cap \text{supp}\psi_{(\mathbf{k},\mathbf{s})} \neq \emptyset$. Für solche Indizes stimmen aber die Multiskalenkoeffizienten $u^k_{(\mathbf{k},\mathbf{s})}$ und $\bar{u}^k_{(\mathbf{k},\mathbf{s})}$ überein, siehe Abschnitt 3.3. Einen deutlichen Unterschied gibt es in den Ergebnissen nur für die verschiedenen Gewichtungsfaktoren $2^{-|\mathbf{k}|_1/2}$ bzw. 1. Der Faktor $2^{-|\mathbf{k}|_1/2}$ führt in dem Problem mit glatter Lösung zu einem etwas besseren Konvergenzverhalten mit weniger Sprüngen. Jedoch ist das Konvergenzverhalten für die quasi-singuläre Lösung deutlich schlechter. Aus diesem Grund wurde für die nachfolgenden numerischen Experimente ausschließlich Kriterium (8.13) benutzt. Unklar ist, warum die Gewichtung mit 1 scheinbar besser funktioniert als die Gewichtung mit $2^{-|\mathbf{k}|_1/2}$. Hier sind in Zukunft weitere Untersuchungen notwendig.

8.3 Numerische Experimente I

Als erstes Beispiel betrachten wir das Dirichlet-Problem

$$-\Delta u = f \quad \text{in }]0, 1]^2 \quad \text{sowie} \quad u = g \quad \text{auf } \partial[0, 1]^2, \quad (8.14)$$

wobei f und g so definiert sind, daß die Lösung durch

$$u(\mathbf{x}) = |\mathbf{x} - \mathbf{x}_0|^{\frac{1}{2}}, \quad \mathbf{x}_0 := \left(\frac{3}{9}, \frac{4}{9}\right)$$

gegeben ist. Diese Funktion hat in \mathbf{x}_0 eine Wurzelsingularität, so daß weder die Finite-Differenzen- noch die Kollokationsmethode unmittelbar angewendet werden können. Aus diesem Grund wird stattdessen ein regularisiertes Dirichlet-Problem behandelt, dessen exakte Lösung für vorgegebenen Regularisierungsparameter $\epsilon > 0$

$$u_\epsilon(\mathbf{x}) := (\epsilon^2 + |\mathbf{x} - \mathbf{x}_0|^2)^{\frac{1}{4}}$$

ist. Die Differenz zwischen der ursprünglichen Lösung u und u_ϵ genügt den folgenden Abschätzungen

$$\|u - u_\epsilon\|_{\mathcal{L}^2} \sim \epsilon \quad \text{bzw.} \quad \|u - u_\epsilon\|_{\mathcal{L}^\infty} = \epsilon^{1/2}.$$

Um die Qualität der adaptiven Diskretisierungen bzw. des Verfeinerungsalgorithmus abschätzen zu können, benötigen wir Aussagen über das Verhalten der Best- n -Term-Approximation von u_ϵ . Da u_ϵ eine regularisierte Variante von u ist, dürften sich die Approximationseigenschaften von u und u_ϵ nicht wesentlich voneinander unterscheiden, sofern die Zahl der Freiheitsgrade bzw. das maximal belegte Level hinreichend klein sind. Ab einem gewissen Punkt müßte aber zum Tragen kommen, daß u_ϵ eine sehr glatte Funktion ist, so daß dann sogar die Approximationsraten für Funktionen mit beschränkter gemischter Ableitung erreicht werden können. Entsprechend unserer Betrachtungen aus Kapitel 5 dürfte somit der Approximationsfehler den folgenden Abschätzungen genügen:

$$\sigma_{n, \mathcal{L}^2}^{aniso}(u_\epsilon) \lesssim \begin{cases} n^{-N/2} \|u\|_{\mathcal{B}_{\mathcal{L}^2}^N} & \text{falls } n \leq n(\epsilon) \\ (n / \log(n))^{-N} \|u\|_{\mathcal{H}_{mix}^N} & \text{sonst} \end{cases}. \quad (8.15)$$

Dabei bezeichnet $n(\epsilon)$ die Grenze, ab der der Charakter der Approximation wechselt. Man beachte, daß in der preasymptotischen Phase, $n \leq n(\epsilon)$, der Approximationsfehler gleichmäßig in ϵ abgeschätzt werden kann, da dieser durch den Best- n -Term-Fehler für u majorisiert werden kann (u_ϵ ist eine geglättete Version von u).

Unser numerisches Experiment soll nun die folgenden Fragen klären:

1. Produziert der Verfeinerungsalgorithmus Indexmengen $\mathcal{T}^{k_{end}}$ und Lösungen $u^{k_{end}}$, für die der Fehler wie oben beschrieben abfällt ?
2. Beobachtet man in der preasymptotischen Phase einen schnelleren Abfall als $|\mathcal{T}|^{-N/2}$, da u eventuell mit anisotropen Wavelets besser approximiert werden kann, als mit isotropen Wavelets ?
3. Ist der Gesamtrechnaufwand proportional zu $|\mathcal{T}^{k_{end}}|$, d.h. $\sum_{k=1}^{k_{end}} |\mathcal{T}^k| \sim |\mathcal{T}^{k_{end}}|$?

Für die Beantwortung dieser Fragen wurde der Verfeinerungsalgorithmus auf die regularisierte Version des Modell-Problems angewendet. Der Parameter ϵ wurde zwischen $10^{-6}, \dots, 10^{-1}$ variiert. Durch Setzen des Schwellwertes $\delta = 10^{-9}, \dots, 10^{-2}$ wurde die Größe von $\mathcal{T}^{k_{end}}$ gesteuert. Der Verfeinerungsalgorithmus wurde für jede Kombination (ϵ, δ) jeweils von vorn gestartet, so daß es keine äußere Schleife gab, mittels derer Endergebnisse von z.B. groben Rechnungen mit Parameter δ als Anfangswerte für Rechnungen mit Parameter $\delta/2$ gesetzt werden.

Für die Diskretisierung haben wir eine Finite-Differenzen-Methode mit einem Verfahren 4. Ordnung, Interpolets 4. Ordnung ($N = K = 4$) und den beiden alternativen Algorithmen AFD bzw. AFD2 sowie die Kollokationsmethode mit Interpolets 6. Ordnung ($N = 6, K = 4$) gewählt. Die Dirichlet-Randbedingungen wurden in allen Fällen mit der in Abschnitt 6.2 beschriebenen Homogenisierungstechnik diskretisiert.

Ein wichtiges Problem ist die Berechnung der Normen des Fehlers. Für die jeweils gegebene Indexmenge \mathcal{T} wurden dabei die folgenden Approximationen der \mathcal{L}^2 - bzw. der \mathcal{L}^∞ -Norm benutzt

$$\|v\|_2 := \left(\int_{[0,1]^2} I_{\mathcal{T}}(v^2) d\mathbf{x} \right)^{1/2} \quad \text{sowie} \quad \|v\|_\infty := \|v\|_{\Omega_{\mathcal{T}}} .$$

Das Integral in $\|\cdot\|_2$ kann dabei exakt berechnet werden, da die Integrale der einzelnen Basis-Funktionen bekannt sind. Unserer Erfahrung nach sind diese Fehler-Norm-Approximationen recht genau. Alternativ wurde der Fehler auf einem vollen Gitter mit $2^{12} \times 2^{12}$ Gitterpunkten berechnet. Die Abweichungen waren dabei immer im Bereich von einigen Prozent, was für die Messung von Fehlern vollkommen ausreichend ist.

In Bild 8.3 ist der Fehler für die verschiedenen Diskretisierungstechniken und die diversen Regularisierungsparameter ϵ über die Zahl der Freiheitsgrade von $\mathcal{T}^{k_{end}}$ aufgetragen. Das in (8.15) beschriebene Verhalten des \mathcal{L}^2 -Fehlers beobachtet man nur für das Finite-Differenzen-Verfahren mit dem AFD Schema und auch dann nicht besonders gut, da der Fehler sehr unregelmäßig abfällt (Bild 8.3 oben links). Es gibt teilweise große Sprünge in den Konvergenzkurven. Der Grund dafür könnte sein, daß die Qualität der in AFD jeweils ausgewählten nicht-äquidistanten Differenzen-Sterne stark schwankt. Bei manchen Gittern müssen offenbar sehr ungünstige Sterne benutzt werden, siehe die Diskussion in Abschnitt 3.6.

Bei der Diskretisierung mit dem AFD2 Schema ist das Konvergenzverhalten zwar sehr viel gleichmäßiger, jedoch kann dieses Schema offenbar auch bei hohen Anzahlen von Freiheitsgraden bzw. benutzten Leveln keinen Nutzen aus der an sich hohen Glattheit von u_ϵ ziehen (Bild 8.3 mitte links). Die Konvergenzrate ist konstant 2. Die Ursache könnte hier sein, daß die effektive Konsistenzordnung nur $N - 2 = 2$ ist, siehe die Diskussion in Abschnitt 3.6.

Bei der Kollokationsmethode verhält sich der Fehler für kleine ϵ nicht wie $|\mathcal{T}|^{-6/2}$ bzw. $|\mathcal{T}|^{-6}$ sondern nur etwa wie $|\mathcal{T}|^{-2}$ bzw. $|\mathcal{T}|^{-3}$ (Bild 8.3 unten links). Möglicherweise liegt das an der reduzierten Konsistenzordnung von 4 bzw. einer nicht optimalen Verfeinerung.

In keinem Fall scheint es möglich zu sein, in der preasymptotischen Phase mit den anisotropen Wavelets eine höhere Konvergenzordnung als $N/2$ zu erreichen.

Im Großen und Ganzen treffen die obigen Aussagen auch auf den Maximumsfehler zu. Jedoch ist hier in der preasymptotischen Phase die Konvergenzrate anscheinend kleiner als 2. Immerhin ist die Maximumsnorm auch eine deutlich stärkere Norm als die \mathcal{L}^2 -Norm.

Zum Abschluß dieses Abschnittes soll noch kurz auf den Rechenaufwand eingegangen werden. In Bild 8.4 ist dazu beispielsweise das Verhältnis von $\sum_k^{k_{end}} |\mathcal{T}^k|$ und $|\mathcal{T}^{k_{end}}|$ für die einzelnen Läufe des adaptiven Verfeinerungsalgorithmus mit der Finite-Differenzen-Diskretisierung AFD2 aufgetragen. Die einzelnen Kurven beziehen sich jeweils auf einen festen ϵ -Wert. Jede Kurve zeigt, daß trotz kleiner werdendem Schwellwert δ und der damit verbundenen Zunahme von Freiheitsgraden in $\mathcal{T}^{k_{end}}$, das Verhältnis $\sum_k^{k_{end}} |\mathcal{T}^k| / |\mathcal{T}^{k_{end}}|$ ungefähr konstant bleibt. Der Rechenaufwand ist also proportional zur Anzahl von Freiheitsgraden in der finalen Lösung. Unterschiede gibt es nur zwischen den einzelnen Kurven. Jedoch ist hier zu beobachten, daß sich die Quotienten mit kleiner werdendem ϵ einander annähern. Als Faustformel kann man sagen, daß der Gesamtrechenaufwand etwa 10 mal höher ist als für die letzte und teuerste

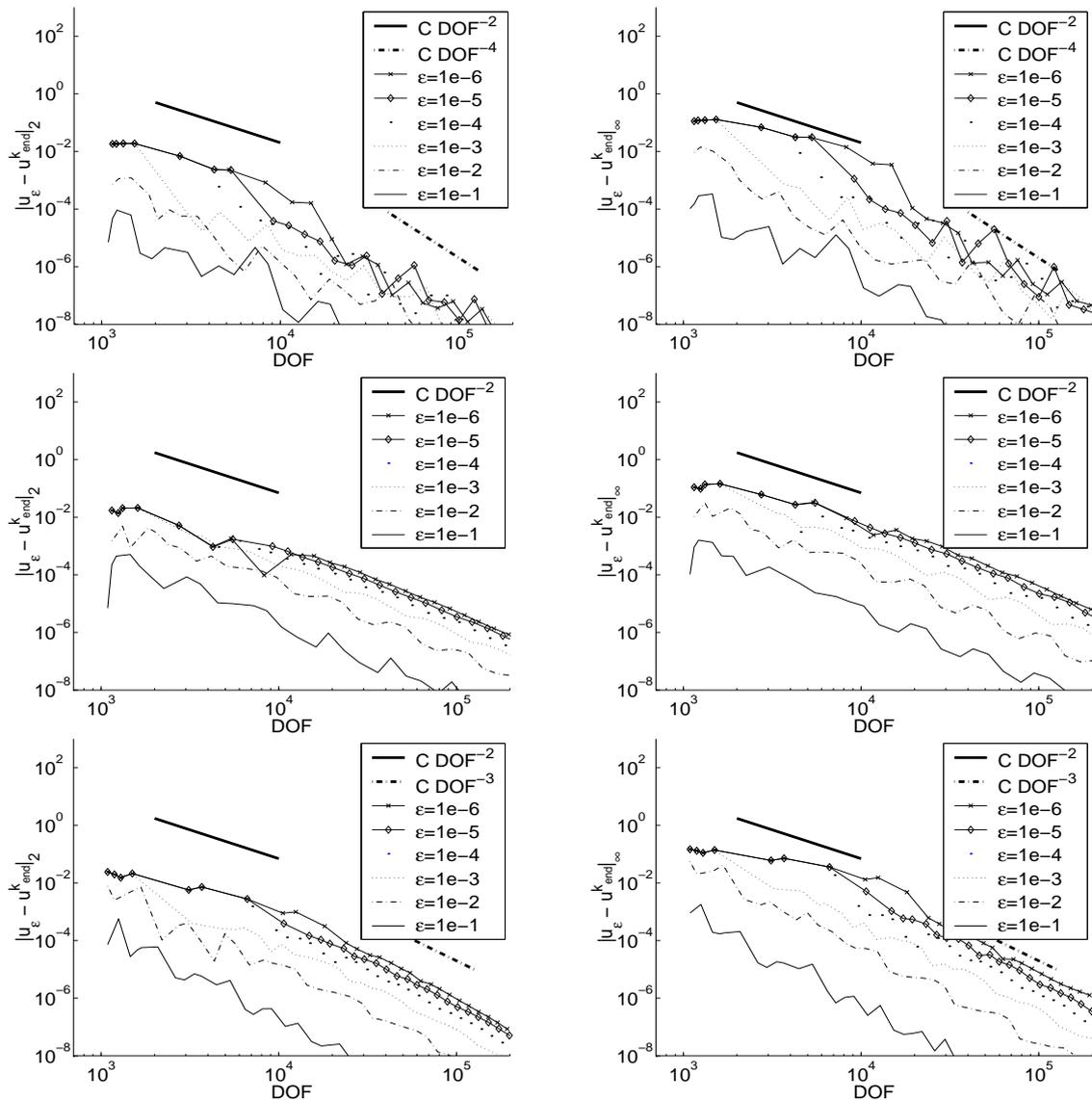


Abbildung 8.3: Konvergenz für die Finite-Differenzen-Diskretisierung mit $N = K = 4$ und Algorithmus AFD (oben), sowie mit $N = K = 4$ und Algorithmus AFD2 (Mitte) bzw. die Kollokationsmethode $N = 6, K = 4$ (unten). Links: $\|\cdot\|_2$ -Norm. Rechts: $\|\cdot\|_\infty$ -Norm.

Rechnung allein.

8.4 Numerische Experimente II

Als zweites Beispiel betrachten wir wieder das Dirichlet-Problem (8.14), diesmal aber mit glatter Lösung $u(\mathbf{x}) = \sin(2\pi x_1) \sin(2\pi x_2)$. Entsprechend unserer Untersuchungen in Kapitel 5 kann diese Funktion optimal mit regulären dünnen Gittern, d.h. $\mathcal{T}^L := \{\mathbf{l} : |\mathbf{l}|_1 \leq l_0 + L\}$ approximiert werden. Für diesen Fall wurde in Theorem 6.1 gezeigt, daß der Konsistenzfehler die Form $C(u)L^{d-1}2^{-\min(N,K)L}$ hat. Ein entsprechendes Verhalten des Konvergenzfehlers beobachtet man auch in Experimenten mit vorgegebenen Indexmengen \mathcal{T}^L . In Bild 8.5 ist

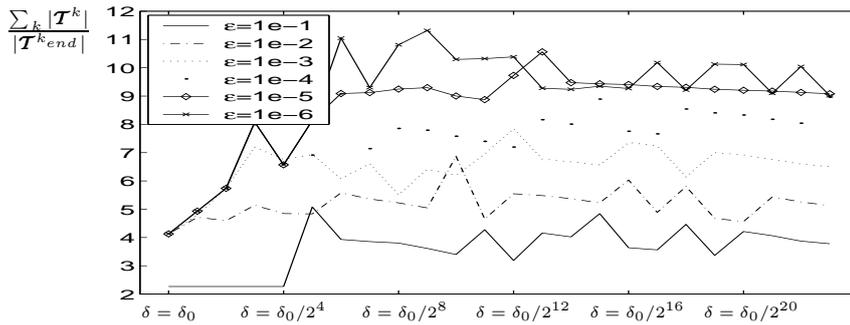


Abbildung 8.4: Das Verhältnis der Summe aller Freiheitsgrade zur Zahl der Freiheitsgrade beim Terminieren des adaptiven Verfeinerungsschemas. Der Schwellwert δ wurde ausgehend von einem δ_0 jeweils um den Faktor zwei verkleinert.

dazu für ein Finite-Differenzen-Schema mit $N = K = 4$ und die Kollokationsmethode mit $N = 6, K = N - 2 = 4$ der Fehler über das maximale Level L aufgetragen. Der Fehler fällt

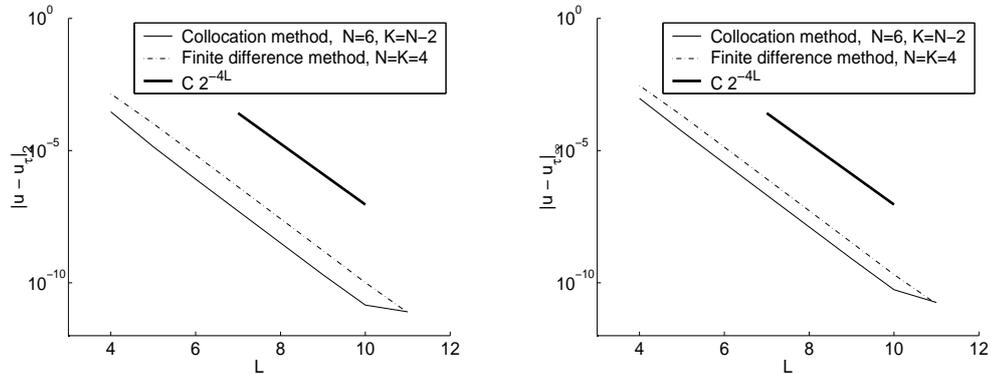


Abbildung 8.5: Konvergenzfehler für reguläre dünne Gitter für die Finite-Differenzen-Technik und die Kollokationsmethode. Links: $\|\cdot\|_2$ -Norm. Rechts: $\|\cdot\|_\infty$ -Norm.

für beide Verfahren etwa wie $|\mathcal{T}^L|^{-4}$ ab. Uns interessiert in diesem Abschnitt aber vielmehr, ob der adaptive Verfeinerungsalgorithmus in der Lage ist, Indexmengen mit einer ähnlichen Konvergenzrate des Fehlers zu generieren. Zur Beantwortung dieser Frage wurde wie im Abschnitt 8.3 vorgegangen: Wieder wurde das Finite-Differenzen-Schema 4. Ordnung, diesmal nur AFD, bzw. die Kollokationsmethode mit Interpolets, $N = 6$, verwendet. Durch Variation des Schwellwertes δ wurde die Zahl der Freiheitsgrade von $\mathcal{T}^{k_{end}}$ kontrolliert. Die Fehler wurden wieder in den diskreten Normen $\|\cdot\|_2$ bzw. $\|\cdot\|_\infty$ gemessen. In Bild 8.6 sind die resultierenden Konvergenzkurven zu sehen. Für beide Diskretisierungen wird die optimale Konvergenzrate von etwa 4 erreicht. Jedoch beobachtet man bei der Finite-Differenzen-Diskretisierung ein recht unregelmäßiges Konvergenzverhalten. Wie auch bei den ersten Experimenten könnte der Grund dafür sein, daß die Qualität der in AFD jeweils ausgewählten nicht-äquidistanten Differenzen-Sterne stark schwankt.

(Zusammenfassung) Die Experimente in den vorangegangenen beiden Abschnitten haben gezeigt, daß die einfache Strategie, mit einem Schwellwert signifikante Indizes zu bestimmen und außerdem deren Nachbarn einzufügen, durchaus sinnvolle Verfeinerungen produziert. Für ein glattes Problem wurden mit der Kollokationsmethode und der Finite-Differenzen-Technik

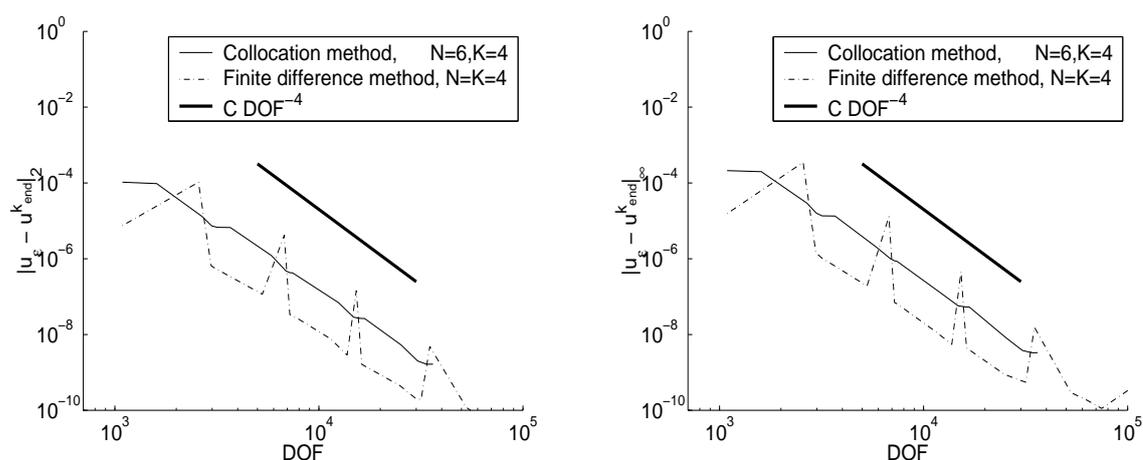


Abbildung 8.6: Konvergenzfehler bei adaptiver Verfeinerung für die Finite-Differenzen-Technik und die Kollokationsmethode. Links: $\|\cdot\|_2$ -Norm. Rechts: $\|\cdot\|_\infty$ -Norm.

AFD optimale Konvergenzraten erreicht. Beim Problem mit quasi-singulärer Lösung wurde nur für die Finite-Differenzen-Technik mit dem AFD Schema das zu erwartende Konvergenzverhalten erreicht. In einer preasymptotischen Phase war die Konvergenzrate etwa $N/2$, wie bei der Best- n -Term-Approximation der echt singulären Funktion. Ab einer gewissen Zahl von Freiheitsgraden, konnte das Verfahren dann aber die Glattheit der regularisierten Lösung ausnutzen und höhere Konvergenzraten erzielen. Bei der Kollokationsmethode bzw. dem AFD2 Schema führt möglicherweise die reduzierte Konsistenzordnung zu schlechteren Konvergenzraten.

Kapitel 9

Erhaltungsgleichungen

Ein für die Strömungsmechanik besonders wichtiger Typ von Gleichungen sind sogenannte Erhaltungsgleichungen [Lax71]. Ein einfaches Beispiel ist

$$\begin{aligned}\partial_t u + \nabla \cdot (\mathbf{a}u) &= 0 \quad \text{für } (\mathbf{x}, t) \in \Omega \times [0, T] \\ u(\mathbf{x}, 0) &= u^0(\mathbf{x}) .\end{aligned}\tag{9.1}$$

Diese Gleichungen beschreiben das Verhalten eines Skalars $u \in \mathbb{R}$ (z.B. Konzentrationsverteilung) unter dem Einfluß eines transportierenden Geschwindigkeitsfeldes \mathbf{a} , das wir als divergenzfrei annehmen

$$\nabla \cdot \mathbf{a} = 0 .$$

Für den Fall periodischer oder homogener Dirichlet-Randbedingungen kann man dann die folgenden Aussagen herleiten

$$\begin{aligned}\partial_t \int_{\Omega} u d\mathbf{x} &= 0 \quad (\text{Masseerhaltung}) \\ \partial_t \int_{\Omega} u^2 d\mathbf{x} &= 0 \quad (\text{Energieerhaltung}) ,\end{aligned}$$

welche den Namen "Erhaltungsgleichungen" erklären. Gleichung (9.1) stellt eine linearisierte Form des Impuls-Satzes in den Navier-Stokes-Gleichungen für den Fall $\nu \rightarrow 0$ dar. Gerade für den Fall sehr kleiner Viskosität ν ist dabei die Qualität der Diskretisierung von Zeitableitung und konvektivem Term von besonderer Bedeutung für die Genauigkeit der numerischen Lösung. Die Eigenschaften dieser Diskretisierungen können einfach und zuverlässig an Hand des Modellproblems (9.1) untersucht werden. So ist z.B. von besonderem Interesse, ob aus der Diskretisierung von Zeitableitung und konvektivem Term künstliche Masse- oder Energie-Defekte resultieren. Solche Defekte spielen bei Langzeitsimulationen von z.B. turbulenten Strömungen eine große Rolle. Wegen des chaotischen Charakters der Lösung kann man nicht erwarten, daß der Abstand zwischen einer numerischen Lösung und der exakten Lösung mit wachsender Zeit klein bleibt [JRB95]. Jedoch sollte die numerische Lösung das gleiche statistische Verhalten zeigen, wie die exakte Lösung. Dafür müssen aber im statistischen Sinne die Masse- und Energie-Bilanzen übereinstimmen.

Die Diskretisierungstechniken für zeitabhängige partielle Differentialgleichungen lassen sich im wesentlichen in zwei Gruppen einteilen: Raum-Zeit-Verfahren [GZ99, Van93] und Zeitschritt-Verfahren [CHQZ88, Gri72]. Bei erstem Typ wird der komplette Raum-Zeit-Zylinder

gleichzeitig diskretisiert. Auf diese Weise können z.B. adaptive Diskretisierungen die möglicherweise vorhandene Glattheit der Lösung in Raum *und* Zeit ausnutzen. Als Nachteil ergeben sich jedoch im allgemeinen hohe Speicheranforderungen, da letztlich ein höherdimensionales Gebiet diskretisiert werden muß. Ein weiterer Nachteil ist die mangelnde theoretische Fundierung. In Experimenten hat sich auch gezeigt, daß die bei der Diskretisierung entstehenden linearen Gleichungssysteme selbst für einfache Probleme wie die Wärmeleitungsgleichung sehr schlecht konditioniert sein können [Zum01b]. Aus diesem Grund werden in dieser Arbeit ausschließlich Zeitschritt-Verfahren betrachtet, die weit entwickelt und gut verstanden sind, sowie einen erheblich kleineren Speicherbedarf haben.

9.1 Zeitschritt-Verfahren und Verfeinerungsstrategien

Die wesentliche Idee ist, $u(\mathbf{x}, t)$ nur an bestimmten Zeitpunkten $t^n = n\Delta$, $n \in \mathbb{N}$, zu berechnen: $u^n \approx u(\cdot, t^n)$. Gleichung (9.1) wird dann im Ort z.B. mittels der adaptiven Multiskalenbasierten Finite-Differenzen-Technik diskretisiert und in der Zeit mittels

$$\frac{u^{n+1} - \sum_{r=0}^m \alpha_r u^{n-r}}{\Delta t} + K(u^{n+1}, u^n, \dots) = 0. \quad (9.2)$$

Dabei stellen die Werte $\{\alpha_r\}$ Verfahrensparameter dar, die zusammen mit $K(\cdot, \cdot)$ die Konsistenzordnung bestimmen, [Gri72, HW91]. $K(\cdot, \cdot)$ ist eine Diskretisierung des konvektiven Terms. Wir nehmen der Einfachheit halber an, daß $K(\cdot, \cdot)$ *nicht* von u^{n+1} abhängt, so daß u^{n+1} explizit aus u^0, \dots, u^n berechnet werden kann. Dabei ergibt sich die folgende CFL-Bedingung an die Zeitschrittweite Δt

$$\Delta t \lesssim \frac{1}{\|\mathbf{a}\|_\infty 2^L}, \quad (9.3)$$

die erfüllt sein muß, um die Stabilität des Verfahrens zu gewährleisten [CHQZ88]. Aus diesem Grund ist es angebracht, das maximale Verfeinerungslevel L der Ortsdiskretisierung zu beschränken, um nicht allzukleine Zeitschritte benutzen zu müssen. Wie man in den numerischen Experimenten im nächsten Abschnitt sehen wird, steht dies mitunter im Widerspruch zum Ziel, eine möglichst effektive adaptive Ortsdiskretisierung zu erreichen. Ein möglicher Ausweg wäre eine implizite Diskretisierung des konvektiven Terms, bei der $K(\cdot, \cdot)$ von u^{n+1} abhängen darf. Allerdings sind dann recht unangenehme Gleichungssysteme zu lösen. Eine weitere Möglichkeit wären lokale Zeitschrittverfahren, bei denen die Unbekannten eines jeden Verfeinerungslevels \mathbf{l} mit einer Zeitschrittweite $O(\frac{1}{\|\mathbf{a}\|_\infty 2^{|\mathbf{l}|_\infty}})$ evolviert werden. Diese Verfahren sind allerdings algorithmisch komplizierter.

Eine der wichtigsten Fragen für adaptive Methoden ist die geeignete Wahl der adaptiven Basis. Ein typisches Phänomen bei Erhaltungsgleichungen ist die Formation von Schocks in der Lösung nach endlicher Zeit, die auch noch weiterwandern können. Somit ist klar, daß eine gute adaptive Basis selbst auch zeitabhängig sein muß, um der Bewegung solcher Unstetigkeitsstellen folgen zu können. Die Idee ist nun, im Zeitschrittverfahren neben u^n auch die adaptive Indexmenge \mathcal{T} von Zeitschritt zu Zeitschritt zu evolviert werden. Die Grobstruktur einer entsprechenden Strategie ist:

Adaptives Zeitschrittverfahren (AZV)

1. Zum Zeitpunkt t^n sei \mathcal{T}^n und die Lösung $u^n \in \text{span } \Psi_{\mathcal{T}^n}$ gegeben.
2. Bestimme $\hat{\mathcal{T}} \supset \mathcal{T}^n$, so daß die (noch nicht bekannte) Lösung u^{n+1} zum neuen Zeitpunkt t^{n+1} *garantiert* gut mittels $\Psi_{\hat{\mathcal{T}}}$ dargestellt werden kann.
3. Berechne $\hat{u}^{n+1} \in \text{span} \Psi_{\hat{\mathcal{T}}^{n+1}}$ aus u^n mittels (9.2).
4. Bestimme die Menge der *tatsächlich* wichtigen Indizes $\mathcal{T}^{n+1} \subset \hat{\mathcal{T}}^{n+1}$ und setze $u^{n+1} := \sum_{(\mathbf{l}, \mathbf{t}) \in \mathcal{T}^{n+1}} \hat{u}_{(\mathbf{l}, \mathbf{t})}^{n+1} \psi_{(\mathbf{l}, \mathbf{t})}$.
5. $n = n + 1$, goto 2.

Im vierten Schritt soll der Fehler $\|u^{n+1} - \hat{u}^{n+1}\|$ hinreichend klein bleiben. Üblicherweise wird dafür ein Abschneide-Verfahren mit Schwellwert δ benutzt. Ggf. können die Koeffizienten dabei auch mit einem levelabhängigen Faktor multipliziert werden.

Für die Wahl von $\hat{\mathcal{T}} \supset \mathcal{T}^n$ bzw. $\mathcal{T}^{n+1} \subset \hat{\mathcal{T}}^{n+1}$ sind von Harten [Har94, Har95] einfache Heuristiken angegeben worden. Für die Bestimmung von $\hat{\mathcal{T}}$ nimmt man z.B. an, daß die Lösung mit endlicher Geschwindigkeit transportiert wird und die CFL-Bedingung erfüllt ist. Bei reinem Transport ist es somit ausreichend, für einen Index $(\mathbf{l}, \mathbf{t}) \in \mathcal{T}^n$ auch die Nachbarn

$$\{(\mathbf{l}, \mathbf{s}) : |\mathbf{s} - \mathbf{t}| < R_{ort}\} \quad (9.4)$$

auf gleichem Level \mathbf{l} in $\hat{\mathcal{T}}$ aufzunehmen. R_{ort} ist dabei ein vorgegebener Parameter. Weiterhin wird angenommen, daß sich Schocks nicht sprunghaft herausbilden so daß es ausreicht, Nachbarn auf bestimmten höheren Levels $\mathbf{k} \geq \mathbf{l}$ in $\hat{\mathcal{T}}$ zu halten. Während in [Har95] nur isotrop verfeinerte Multiskalenfunktionen betrachtet wurden, wo die Level linear angeordnet sind, gibt es in unserem Fall keine kanonische Wahl für die Level $\mathbf{k} \geq \mathbf{l}$ auf denen Nachbarn eingefügt werden sollten. Eine erste einfache Idee wäre hier:

$$\{(\mathbf{k}, \mathbf{s}) : \mathbf{k} \geq \mathbf{l}, |\mathbf{k} - \mathbf{l}|_\infty \leq R_{level} \text{ und } \max_i |2^{-(k_i - l_i)} s_i - t_i| < R_{ort}\}, \quad (9.5)$$

wobei R_{level} einen geeignet zu wählenden Parameter darstellt. Eine andere Möglichkeit wäre, in (9.5) die Bedingung $|\mathbf{k} - \mathbf{l}|_\infty \leq R_{level}$ durch $|\mathbf{k} - \mathbf{l}|_1 \leq R_{level}$ zu ersetzen. Je nach Größe von R_{level} hat dies enorme Auswirkungen auf die Anzahl von Freiheitsgraden in $\hat{\mathcal{T}}$: $|\cdot|_\infty$ entspricht einem Verfeinern analog zu einem vollen Gitter wobei $O(2^{d \cdot R_{level}})$ neue Indizes eingefügt werden, wogegen $|\cdot|_1$ einem Verfeinern analog zum dünnen Gitter entspricht, wobei nur $O(2^{R_{level}} R_{level}^{d-1})$ Indizes eingefügt werden. In unseren Experimenten hat sich jedoch gezeigt, daß auch für $R_{level} = 1$ bereits gute Ergebnisse erzielt werden, so daß die Frage, ob $|\cdot|_\infty$ oder $|\cdot|_1$, letztlich nicht entscheidend ist.

9.2 Numerische Experimente

Als Beispiel betrachten wir den Molenkamp-Test [VK93], bei dem eine Anfangsverteilung u^0 auf $\Omega =]-1, 1]^2$ einer kreisförmigen Konvektion ausgesetzt wird, d.h.

$$\mathbf{a}(x, y) := \begin{pmatrix} y \\ -x \end{pmatrix}.$$

Die Anfangsverteilung ist in unseren Experimenten durch

$$u^0(\mathbf{x}) := \max\left(0, 4e^{-15\|\mathbf{x}-\mathbf{x}_0\|^2} - 3\right), \quad \mathbf{x}_0 = (1/4, 1/4)$$

gegeben, siehe Bild 9.1. Diese Verteilung ist lediglich stetig also nicht sehr glatt. Das stellt

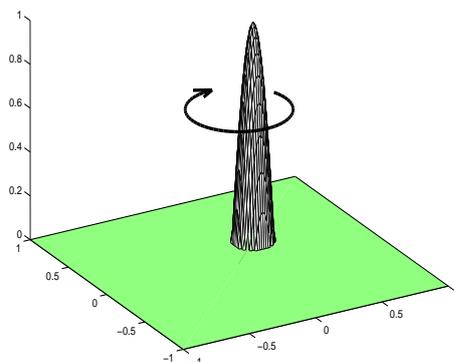


Abbildung 9.1: Anfangsbedingung für den Molenkamp-Test.

eine zusätzliche Schwierigkeit dar, da numerische Verfahren dazu neigen, in der Nähe von Sprüngen und Knicken unphysikalische Oszillationen zu generieren. Tatsächlich ist man in der Numerik von Erhaltungsgleichungen fast ausschließlich damit beschäftigt, Verfahren zu entwickeln, die einerseits solche Oszillationen weitgehend vermeiden und andererseits die Sprünge und Knicke möglichst gut erhalten. Eine Übersicht über ENO und WENO Schemen, bei denen das sehr gut gelingt, ist in [Shu99] zu finden. Das für die Experimente benutzte Verfahren ist durch die folgenden Parameter charakterisiert:

finite Differenzen	: AFD2, konservatives WENO 5. Ordnung
Interpolets	: $N = 4$
maximales Level	: $L = 9$ bzw. 8
R_{level}	: 1 (fest)
R_{ort}	: 1 (fest)
Schwellwert	: $\delta = 3e-2, \dots, 3e-4$
Zeitdiskretisierung	: Adams-Bashforth 3. Ordnung (AB3)
Zeitschrittweite	: $\Delta t = 0.0005$ ($CFL = 0.1$)

Durch die Kombination von Interpolets der Ordnung $N = 4$ und dem einfachen AFD2-Schema für die Finite-Differenzen-Diskretisierung ist die *formale* Konsistenzordnung nicht 5, wie beim ursprünglichen WENO Schema, sondern nur noch 3. Da allerdings bei unserem Test die Anfangsverteilung bzw. Lösung nicht sehr glatt ist, ist die Konsistenzordnung des WENO Schemas ohnehin kleiner als 5, wie man an den Ergebnissen einer nichtadaptiven Vergleichsrechnung sehen kann. Dazu wurde ein analoges Finite-Differenzen-Schema (AB3, WENO5) auf kartesischen Gittern mit Maschenweite $h = 2^{-5}, \dots, 2^{-9}$ implementiert.

Zur Berechnung von Fehlern, Masse und Energie sind die diskreten Lösungen jeweils auf ein Gitter mit Maschenweite 2^{-L} interpoliert worden, wodurch eine hohe Genauigkeit dieser Werte erzielt wird.

Zunächst vergleichen wir die Effizienz der beiden Verfahren bzgl. Fehlern in der L_1 - und L_∞ -Norm. In Bild 9.2 sind die Fehler zum Zeitpunkt $t = 2\pi$ (entspricht einem Umlauf) gegen

die Anzahl von Freiheitsgraden aufgetragen. Für die adaptive Rechnung wurde eine über die Zeit gemittelte Zahl von Freiheitsgraden benutzt, da \mathcal{T}^n und damit $|\mathcal{T}^n|$ nicht konstant ist. Allerdings sind die Schwankungen recht gering, siehe Bild 9.3 (links). Ein Beispiel für die resultierenden adaptiven Gitter ist in Bild 9.3 (rechts) angegeben. Erwartungsgemäß sind die Gitterpunkte stark in der Nähe des Knickes konzentriert.

Aus Bild 9.2 geht hervor, daß das nichtadaptive Verfahren nur mit den Raten von etwa $\sim DOF^{-1/2} = 2^{-L}$ (L_∞) und $\sim DOF^{-1} = 2^{-2L}$ (L_1) konvergiert, wogegen das adaptive Verfahren die erheblich besseren Raten von etwa $\sim DOF^{-1}$ bzw. $\sim DOF^{-2}$ erreicht. Allerdings ist dabei der kleinst-mögliche Fehler durch das maximale Level L bestimmt. Ab einer bestimmten Zahl von Freiheitsgraden wird eine Sättigung erreicht und zusätzliche Freiheitsgrade bringen keinen substantiellen Gewinn mehr. Man müßte eigentlich L unbeschränkt lassen, um die bestmögliche Effizienz zu erzielen. Jedoch sind dann wegen der CFL-Bedingung sehr kleine Zeitschrittweiten erforderlich, was natürlich zu hohen Kosten führt.

Neben den L_1/L_∞ -Fehlern ist das Auftreten und die Größe von unphysikalischen Oszillationen von Bedeutung. Diese sind für die beiden Verfahren in Bild 9.4 dargestellt. Hier ist das adaptive Verfahren im Nachteil. Die auftretenden Oszillationen sind um einen Faktor von ca. 4 größer als im nichtadaptiven Fall. Die Ursache dafür ist die Zusammensetzung der Multiskalen-basierten Finite-Differenzen-Methode aus einem Finite-Differenzen-Schema und einem Interpolationsschema, vergleiche Abschnitt 6.3. Während das WENO-Schema Oszillationen durch eine geeignete Wahl des Differenzen-Sterns vermeidet, ist der Interpolations-Stern durch die Interpolets fest vorgegeben. Die Größe der Oszillationen war bei beiden Verfahren für $t \in [0, 2\pi]$ ungefähr konstant. In Bild 9.5 (links) sind dazu die maximalen Unterschwinger für $t \in [0, 2\pi]$ aufgetragen. Ein Vergleich mit den Werten aus Bild 9.2 zeigt, daß die Oszillationen deutlich kleiner sind als der L_∞ -Fehler, so daß dieser Fehler im wesentlichen von der numerischen Diffusion dominiert ist, wie das für Finite-Differenzen-Techniken üblich ist. In Bild 9.5 sind dazu Kontur-Linien der Niveaus $u = \{5e-3, 1e-1, 0.95\}$ für die adaptive ($\delta = 3e-4$, $L = 9$) und die exakte Lösung dargestellt. Die Kontur-Linien für die exakte Lösung (gestrichelte Linien) zeigen wegen der steileren Gradienten und der einfachen stückweise linearen Interpolation im Kontur-Algorithmus deutliche Aliasing-Fehler, die uns aber hier nicht weiter stören.

Am kleinen Niveau-Wert $5e-3$ sieht man, daß die numerische Lösung in der Nähe des Knickes etwas auseinandergefallen ist. Dieser Effekt ist besonders ausgeprägt in Richtung des Transports. Weiterhin ist zu entnehmen, daß der Phasenfehler, bei dem die numerische Lösung zwar weitgehend die richtige Form und Größe hat, aber der exakten Lösung etwas hinterherhinkt [CHQZ88], sehr klein ist. Für alle anderen Niveau-Werte sind die Abweichungen so klein, daß sie kaum noch wahrzunehmen sind.

In Bild 9.6 ist der zeitliche Verlauf von Masse und Energie für die adaptive Rechnung dargestellt. Während beim nichtadaptiven Verfahren die Masse nahezu exakt erhalten wird, gibt es bei den adaptiven Rechnungen kleine Schwankungen. Im statistischen Mittel kann aber wohl auch hier von einer Masseerhaltung ausgegangen werden. Prinzipiell wäre es auch für die adaptiven Rechnungen möglich, die Masse exakt zu erhalten. Bei periodischen oder homogenen Dirichlet-Rändern ist dazu einfach der numerische konvektive Term auf den Raum der Funktionen mit verschwindendem Mittelwert zu projizieren:

$$\nabla \cdot (\mathbf{a}u)_{num'} := \nabla \cdot (\mathbf{a}u)_{num} - \frac{1}{|\Omega|} \int_{\Omega} \nabla \cdot (\mathbf{a}u)_{num} dx .$$

Diese Projektion entspricht der \mathcal{L}^2 -orthogonalen Projektion, so daß dadurch sogar der Dis-

diskretisierungsfehler in der \mathcal{L}^2 -Norm verkleinert wird.

Bezüglich Energieerhaltung verhalten sich das adaptive und das nichtadaptive Verfahren ungefähr gleich. Die im WENO Schema enthaltenen Mechanismen zur Oszillationsdämpfung bewirken eine Art numerische (Hyper-)Diffusion, die zu Beginn der Rechnungen besonders stark ist, wenn die Lösungen noch nicht geglättet wurden. Anschließend ist die Diffusion dann aber sehr klein. Zum Vergleich wurde auf die rechte Seite von (9.1) explizit ein Zähigkeits-term $\nu \Delta u$, mit $\nu = 1e-6$ und $1e-7$, hinzugefügt. Zumindest für $\nu = 1e-6$ führt das zu einer deutlich stärkeren Dissipation von Energie als durch die numerische Diffusion.

(Zusammenfassung) Das adaptive Multiskalen-basierte Verfahren führt bezüglich $\mathcal{L}^1/\mathcal{L}^\infty$ -Fehlern zu deutlich besseren Konvergenz-Raten der Form $DOF^{-\alpha}$ als nichtadaptive Verfahren. Die kleinsten erzielbaren Fehler sind jedoch durch das maximale Level L bestimmt. Unphysikalische Oszillationen sind für das adaptive Verfahren größer als beim nichtadaptiven, jedoch von der gleichen Größenordnung. Bei der adaptiven Diskretisierung geht die Eigenschaft der Masseerhaltung des WENO Schemas verloren. Allerdings sind die Abweichungen sehr klein – insbesondere im statistischen Mittel. Die numerische Diffusion ist wie beim nichtadaptiven Verfahren sehr gering, so daß die Multiskalen-basierte Finite-Differenzen-Methode für die Simulation von Strömungen mit hohen Reynoldszahlen geeignet ist.

Diese guten bis sehr guten Ergebnisse wurden mit der einfachsten Verfeinerungsstrategie (9.5) und $R_{level} = R_{ort} = 1$ erzielt, die sich somit als sehr robust erwiesen hat.

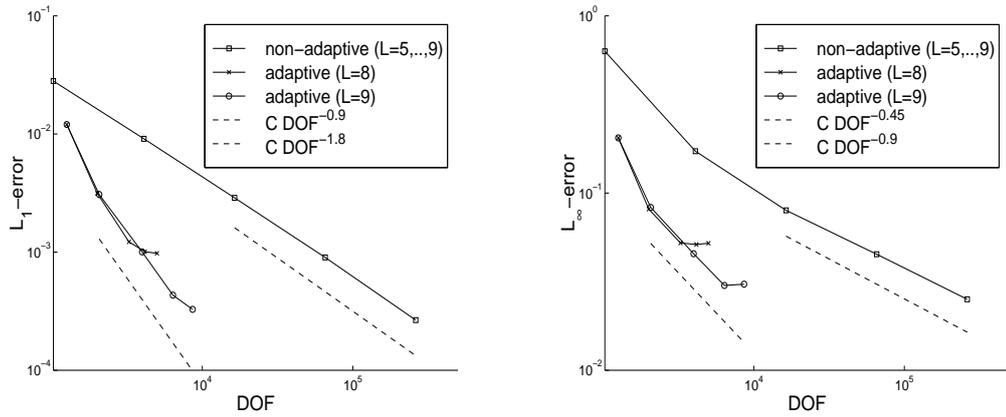


Abbildung 9.2: Konvergenz der Fehler zum Zeitpunkt $t = 2\pi$ (ein Umlauf) in Abhängigkeit von der Zahl der Freiheitsgrade für den Molenkamp-Test.

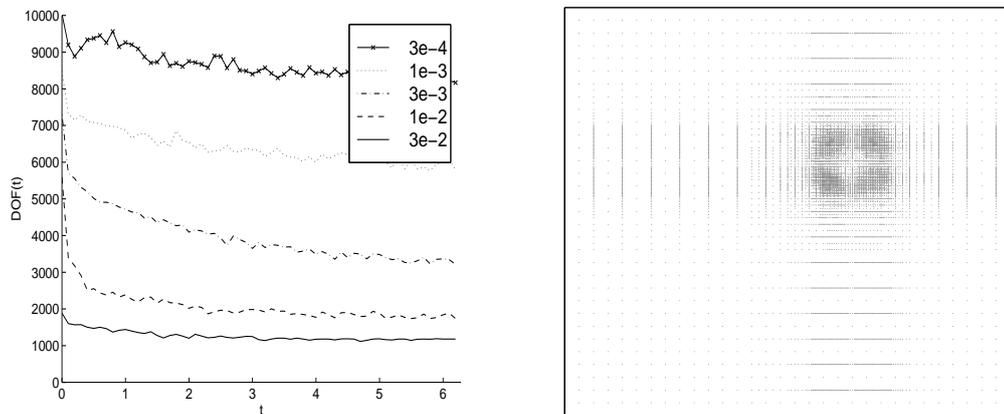


Abbildung 9.3: Links: Zeitlicher Verlauf der Anzahl der Freiheitsgrade. Rechts: adaptives Gitter für den Zeitpunkt $t = 2\pi$ ($\delta = 3e-4$, $L = 9$).

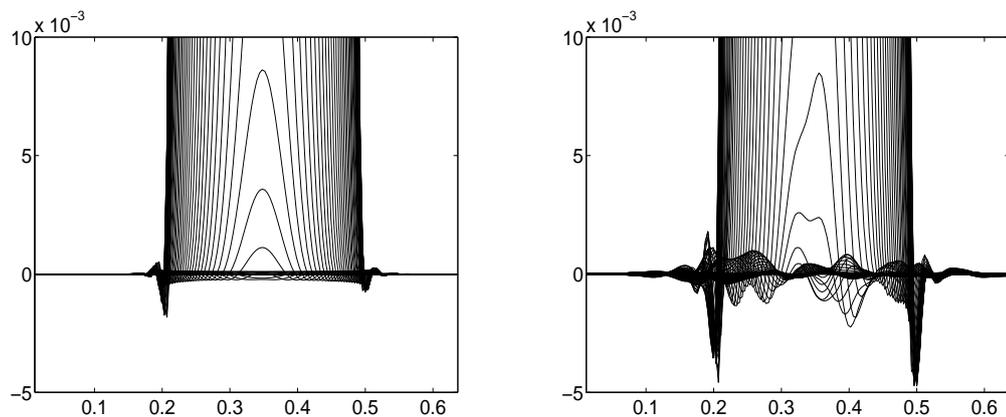


Abbildung 9.4: Oszillationen in den numerischen Lösungen für $\delta = 3e-4$ und $L = 9$ zum Zeitpunkt $t = 2\pi$. Vergrößerung der Frontal-Ansicht. Links: nichtadaptiv. Rechts: adaptiv.

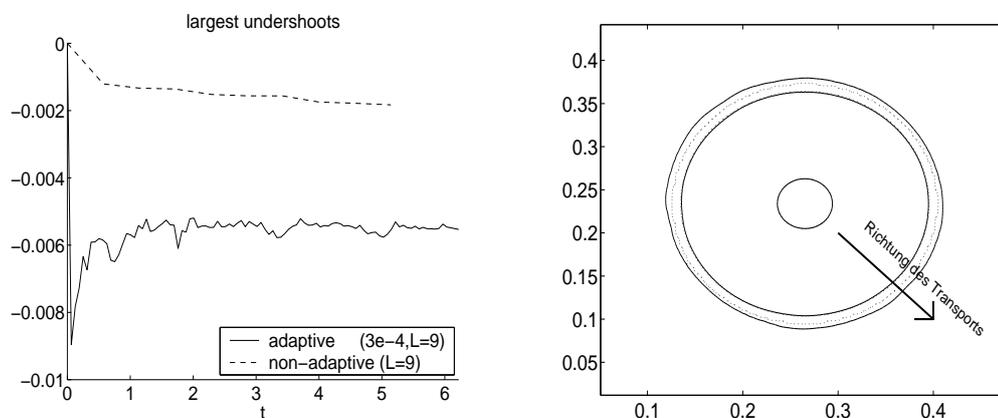


Abbildung 9.5: Links: Zeitlicher Verlauf der maximalen Unterschinger für adaptive und nichtadaptive Rechnungen. Rechts: Kontur-Linien ($5e-3, 1e-1, 0.95$) für die adaptive ($\delta = 3e-4, L = 9$) (durchgezogene Linien) und die exakte Lösung (gestrichelte Linien).

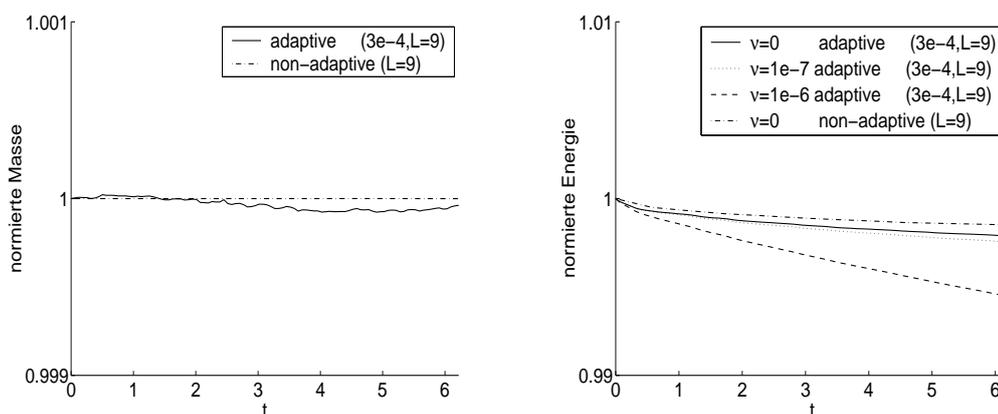


Abbildung 9.6: Erhaltung von Masse und Energie. Links: Masse bei reiner Konvektion (normiert auf 1). Rechts: Vergleich Energie (normiert auf 1) bei reiner Konvektion und zusätzlicher kleiner Diffusion $\nu = 1e-7$.

Kapitel 10

Navier-Stokes Gleichungen

In diesem Kapitel benutzen wir das Multiskalen-basierte Verfahren für die Lösung der Navier-Stokes-Gleichungen. Diese Gleichungen beschreiben die Bewegung von inkompressiblen Fluiden in Abhängigkeit von Rand- und Anfangsbedingungen sowie den physikalischen Eigenschaften der Fluide. In der Formulierung mittels der sogenannten primitiven Variablen, Geschwindigkeit $\mathbf{u} = (u, v, w)^T$ und Druck p , sind die Navier-Stokes-Gleichungen durch

$$\partial_t \mathbf{u} + \nabla \cdot (\mathbf{u}\mathbf{u}^T) = \mathbf{f} - \nabla p + \nu \Delta \mathbf{u} \quad \text{in } \Omega \subset \mathbb{R}^3 \quad (10.1)$$

$$\nabla \cdot \mathbf{u} = 0 \quad (10.2)$$

gegeben. Dazu kommen noch Anfangsbedingungen $\mathbf{u}(\mathbf{x}, 0) = \mathbf{u}^0(\mathbf{x})$ und Randbedingungen. Der Parameter ν ist die kinematische Viskosität des Fluids und \mathbf{f} ist eine vorgegebene äußere Volumenkraft. Die erste Gleichung (10.1) beschreibt die Impulserhaltung und die zweite Gleichung (10.2) die Masseerhaltung.

Der Impulssatz (10.1) hat im wesentlichen wieder die Struktur des Modell-Problems (9.1). Allerdings ist die rechte Seite nicht Null und die Geschwindigkeit \mathbf{u} tritt sowohl als transportierende Größe als auch als transportierte Größe im konvektiven Term auf. Trotzdem können aber für die drei Komponenten des Impulssatzes wieder die gleichen Diskretisierungstechniken wie für skalare Erhaltungsgleichungen benutzt werden, d.h. ein adaptives Zeitschrittverfahren mit expliziter Behandlung und ENO/WENO-Diskretisierung des konvektiven Terms.

Für die Numerik besteht die zusätzliche Schwierigkeit bei den Navier-Stokes-Gleichungen nun darin, daß die Differentialgleichungen (10.1) und (10.2) gekoppelt sind.

Speziell für Galerkin-Diskretisierungen hat man die Möglichkeit diese Kopplung zu eliminieren, indem divergenzfreie Ansatzfunktionen für \mathbf{u} benutzt werden. In diesem Fall ist (10.2) automatisch erfüllt. Auch der Druckgradient in (10.1) fällt nach partieller Integration bei geeigneten Randbedingungen weg. Es ist jedoch nicht unmittelbar klar, ob diese partielle Integration auch auf die in dieser Arbeit untersuchten Finite-Differenzen-Diskretisierungen übertragen werden kann. Falls nicht, hätte man durch divergenzfreie Multiskalenfunktionen zwar formal Gleichung (10.2) eliminiert, jedoch die entsprechende Variable p fällt nicht weg. Ein weiterer Nachteil ist, daß die Konstruktion divergenzfreier Tensorprodukt-Multiskalenfunktionen wohl nur auf einfachen Gebieten wie $[0, 1]^d$ möglich ist [Urb95].

Eine andere Lösung für das Kopplungsproblem ergibt sich aus der Formulierung der Navier-Stokes-Gleichungen mittels Geschwindigkeit und Rotation $\boldsymbol{\omega} := \nabla \times \mathbf{u}$. Falls auf $\partial\Omega$ für \mathbf{u} homogene Dirichlet-Randwerte vorgegeben sind, lauten die Gleichungen [KS99]:

$$\partial_t \boldsymbol{\omega} + \nabla \times (\boldsymbol{\omega} \times \mathbf{u}) = \nabla \times \mathbf{f} - \nu \nabla \times \nabla \times \boldsymbol{\omega} \quad \text{in } \Omega \quad (10.3)$$

$$\begin{aligned}
\Delta \mathbf{u} &= -\nabla \times \boldsymbol{\omega} & \text{in } \Omega \\
\boldsymbol{\omega} &= \nabla \times \mathbf{u} & \text{auf } \partial\Omega \\
\mathbf{u} &= 0 & \text{auf } \partial\Omega \\
\boldsymbol{\omega}(\mathbf{x}, 0) &= \nabla \times \mathbf{u}(\mathbf{x}, 0) & \text{in } \partial\Omega .
\end{aligned} \tag{10.4}$$

Hier kann das Analogon des Impulssatzes, d.h. Gleichung (10.3), wieder mit einem Zeitschritt-Verfahren diskretisiert werden. Die für die Auswertung des konvektiven Terms notwendige Geschwindigkeit \mathbf{u} kann mit Hilfe der drei Poisson-Gleichungen (10.4) berechnet werden. Die \mathbf{u} - $\boldsymbol{\omega}$ -Formulierung der Navier-Stokes-Gleichungen ist damit numerisch relativ teuer. Im Vergleich dazu muß bei den im nächsten Abschnitt vorgestellten Druck-Korrektur-Verfahren je Zeitschritt nur eine Poisson-Gleichung gelöst werden.

Eine dritte alternative Formulierung der Navier-Stokes-Gleichungen ist in [Ose89] angegeben worden. Dabei werden \mathbf{u} und p durch zwei neue Variablen \mathbf{m} und χ ersetzt, die über

$$\mathbf{m} = \mathbf{u} + \nabla \chi \tag{10.5}$$

mit der Geschwindigkeit zusammenhängen. Die neue Größe \mathbf{m} gehorcht der Erhaltungsgleichung

$$\partial_t \mathbf{m} + \nabla \cdot (\mathbf{u} \mathbf{u}^T) = \mathbf{f} + \nu \Delta \mathbf{m} .$$

Die Randbedingungen für \mathbf{m} und χ sind nicht eindeutig festgelegt, jedoch muß (10.5) auf dem Rand gelten [BCM99]. Die für die Auswertung des konvektiven Terms notwendige Geschwindigkeit \mathbf{u} kann mittels einer Hodge-Projektion aus \mathbf{m} berechnet werden: Da $\nabla \cdot \mathbf{u} = 0$ gelten soll, erfüllt χ die Poisson-Gleichung

$$\nabla \cdot \mathbf{m} = 0 + \nabla \cdot \nabla \chi .$$

Wenn χ bekannt ist, kann auch \mathbf{u} berechnet werden. Somit ist wie bei Druck-Korrektur-Verfahren jeweils nur eine Poisson-Gleichung je Zeitschritt zu lösen. Es ist deshalb nicht verwunderlich, daß in der Literatur bekannte numerischen Verfahren [Cor95, Cor96, BCM99], die auf dieser dritten Formulierung der Navier-Stokes-Gleichungen basieren, den Druck-Korrektur-Verfahren sehr ähnlich sind.

10.1 Druck-Korrektur-Verfahren

Bei Druck-Korrektur-Verfahren [Cho68, Tem69, BCG89], die auch als Projektionsmethoden bezeichnet werden, werden die Navier-Stokes-Gleichungen in der \mathbf{u} - p -Formulierung mit einem Zeitschritt-Verfahren diskretisiert. In jedem Zeitschritt werden zwei entkoppelte Teilprobleme gelöst, die im wesentlichen die folgende Form haben:

1. Im ersten Teilschritt wird mittels des Impulssatzes eine Hilfgeschwindigkeit \mathbf{u}^* berechnet, die nicht divergenzfrei zu sein braucht:

$$\frac{\mathbf{u}^* - \mathbf{u}^n}{\Delta t} + K(\mathbf{u}^n, \mathbf{u}^{n-1}, \dots) = -\nabla p^n + \nu \Delta \mathbf{u}^* .$$

Wir haben hier angenommen, daß der Druck mit einem expliziten Euler-Schema und der diffusive Term mit einem impliziten Euler-Schema diskretisiert wird. Andere Schemen höherer Ordnung sind möglich, jedoch muß der Druck immer explizit behandelt werden. Hinter $K(\mathbf{u}^n, \mathbf{u}^{n-1}, \dots)$ verbirgt sich eine (explizite) Diskretisierung des konvektiven Terms.

2. In einem zweiten Schritt wird dann aus \mathbf{u}^* das eigentlich gesuchte divergenzfreie Geschwindigkeitsfeld \mathbf{u}^{n+1} berechnet. Dazu wird das Sattelpunktproblem

$$\frac{\mathbf{u}^{n+1} - \mathbf{u}^*}{\Delta t} = -\nabla(p^{n+1} - p^n) \quad (10.6)$$

$$\nabla \cdot \mathbf{u}^{n+1} = 0 \quad (10.7)$$

gelöst. Wendet man auf (10.6) den Divergenz-Operator an und nutzt (10.7), dann folgt die Druck-Poisson-Gleichung

$$\frac{1}{\Delta t} \nabla \cdot \mathbf{u}^* = \nabla \cdot \nabla(p^{n+1} - p^n) . \quad (10.8)$$

Die Geschwindigkeit \mathbf{u}^{n+1} ergibt sich dann aus (10.6). Beim Lösen des Sattelpunktproblems ist darauf zu achten, daß der resultierende Schur-Komplement-Operator $\nabla \cdot \nabla$ stabil diskretisiert wird. Darauf wird detailliert in Abschnitt 10.2 eingegangen.

Gegenstand heftiger Diskussionen [GS87] sind bei Druck-Korrektur-Verfahren die Randbedingungen für \mathbf{u}^* und p an Rändern Γ_D mit Dirichlet-Randwerten für \mathbf{u} . Die übliche Wahl ist hier $\mathbf{u}^*|_{\Gamma_D} = \mathbf{u}^{n+1}|_{\Gamma_D}$ sowie als Folgerung daraus $\partial_{\mathbf{n}} p^{n+1}|_{\Gamma_D} = 0$ (multipliziere (10.6) mit \mathbf{n} und werte auf Γ_D aus). Im allgemeinen sind die homogenen Neumann-Randwerte für p jedoch unphysikalisch, so daß an Dirichlet-Rändern Grenzschicht-artige Fehler für p resultieren [EL95, Pro97, Wet]. Diese haben die maximale Größe von $(\nu \Delta t)^{1/2}$. Sie sind also für hohe Reynoldszahlen und kleine Zeitschrittweiten relativ klein. Um das Problem der Fehlergrenzschichten zu beseitigen, sind modifizierte Druck-Korrektur-Verfahren entwickelt worden [Pro97, BCM99], bei denen \mathbf{u}^{n+1} ebenfalls mittels eines Sattelpunktproblems analog (10.6) und (10.7) berechnet wird. In diesem tritt aber der eigentliche Druck p^{n+1} selbst nicht mehr auf:

$$\frac{\mathbf{u}^{n+1} - \mathbf{u}^*}{\Delta t} = -\nabla \chi^{n+1} , \quad \nabla \mathbf{u}^{n+1} = 0 , \quad \partial_{\mathbf{n}} \chi^{n+1}|_{\Gamma_D} = 0 .$$

Der Druckgradient für den nächsten Zeitschritt wird dann z.B. mittels $\nabla p^{n+1} = \nabla p^n - \nu \Delta \nabla \chi^{n+1}$ aufdatiert [BCM99]. Die Hilfsgröße χ nimmt somit homogene Neumann-Randwerte an, der eigentliche Druck p aber im allgemeinen nicht.

Eine andere Variante, die Fehlergrenzschichten zu vermeiden, besteht darin, andere Randbedingungen für p^{n+1} in (10.8) zu benutzen. Diskutiert wurden inhomogene Neumann-Randwerte, z.B. der Form [KIO91, OID86]

$$\partial_{\mathbf{n}} p^{n+1} = -\nu \mathbf{n} \nabla \times \nabla \times (2\mathbf{u}^n - \mathbf{u}^n) \approx -\nu \mathbf{n} \nabla \times \boldsymbol{\omega}^{n+1} ,$$

oder auch Dirichlet-Werte, die aus Tangential-Ableitungen von p am Rand ermittelt werden können [GS87, Sch98a].

Eine genaue Diskussion, welche der verschiedenen Varianten am geeignetsten für adaptive Diskretisierungen auf dünnen Gittern ist, würde jedoch über den Rahmen dieser Arbeit hinausgehen. Aus diesem Grund beschränken wir uns im restlichen Teil dieses Kapitels auf periodische Randbedingungen. In diesem Fall gibt es die oben erwähnten Schwierigkeiten an den Rändern nicht.

Alles in allem hat der komplette Algorithmus für die adaptive Lösung der Navier-Stokes-Gleichungen letztlich wieder die Form des adaptiven Zeitschritt-Verfahrens AZV aus Kapitel 9. Allerdings wird nun nicht nur die skalare Lösung $u^n \in \text{span} \Psi_{\mathcal{T}^n}$ evolviert, sondern

alle drei Geschwindigkeitskomponenten u^n, v^n, w^n und der Druck p^n . Bei der Berechnung der tatsächlich wichtigen Indizes in Schritt 5 von AZV bietet es sich an, die Multiskalenkoeffizienten aller Unbekannten u^n, \dots, p^n zusammengenommen im Abschneide-Verfahren zu berücksichtigen, d.h. es wird ein Kriterium der Form

$$|u_{(\mathbf{l}, \mathbf{t})}^n| + |v_{(\mathbf{l}, \mathbf{t})}^n| + |w_{(\mathbf{l}, \mathbf{t})}^n| + |p_{(\mathbf{l}, \mathbf{t})}^n| > \delta$$

benutzt. Eine genaue Untersuchung, ob beim Abschneiden eventuell Level-abhängige Gewichtungsfaktoren benutzt werden sollten, wie diese aussehen könnten und ob es dabei eventuell Unterschiede zwischen u^n, \dots, w^n und p^n gibt, würde ebenfalls über den Rahmen dieser Arbeit hinausgehen. Unsere Erfahrungen mit obigem einfachen Schema sind jedoch gut.

10.2 Stabile Diskretisierung der Druck-Poisson-Gleichung

Bei Druck-Korrektur-Verfahren ist auf eine stabile Diskretisierung des Sattelpunktproblems (10.6) und (10.7) zu achten. Dabei treten in (10.6) und (10.7) nicht die kontinuierlichen Gradienten- bzw. Divergenzoperatoren auf, sondern diskrete Varianten, die hier mit ∇_h bzw. $\nabla_h \cdot$ bezeichnet sein sollen. Aus (10.6) und (10.7) wird somit

$$\frac{\mathbf{u}^{n+1} - \mathbf{u}^*}{\Delta t} = -\nabla_h(p^{n+1} - p^n), \quad \nabla_h \cdot \mathbf{u}^{n+1} = 0.$$

Daraus ergibt sich die diskrete Druck-Poisson-Gleichung

$$\nabla_h \cdot \nabla_h(p^{n+1} - p^n) = \frac{1}{\Delta t} \nabla_h \cdot \mathbf{u}^*. \quad (10.9)$$

Der darin enthaltene Laplace-Operator entspricht nicht einer üblichen Diskretisierung, bei der direkt die zweiten partiellen Ableitungen $\partial_{x_i x_i}$ diskretisiert werden, sondern dem Hintereinanderanwenden von $\nabla_h \cdot$ und ∇_h . Die Auswertung von $\nabla_h \cdot \nabla_h$ ist somit etwa doppelt so teuer, wie die Auswertung von Δ_h mittels direkter Diskretisierung von $\partial_{x_i x_i}$. Jedoch kann nur mit dem zusammengesetzten Operator $\nabla_h \cdot \nabla_h$ gewährleistet werden, daß nach der Geschwindigkeitskorrektur, $\mathbf{u}^{n+1} = \mathbf{u}^* - \Delta t \nabla_h(p^{n+1} - p^n)$, die neue Geschwindigkeit \mathbf{u}^{n+1} diskret divergenzfrei ist, d.h. $\nabla_h \cdot \mathbf{u}^{n+1} = 0$.

Das eigentliche Problem von $\nabla_h \cdot \nabla_h$ sind aber nicht so sehr die Kosten für die Auswertung als vielmehr die Stabilitätseigenschaften dieses Operators. Um dies etwas näher zu erläutern schränken wir uns wie in Kapitel 7 auf den nichtadaptiven Fall ein, d.h. $\mathcal{T} = \{(\mathbf{l}, \mathbf{t}) : |\mathbf{l}|_\infty \leq L\}$. Außerdem benutzen wir die dortige Notation für die Matrix-Darstellung der diskreten Operatoren. Mit T_h sei wie in (7.3) die Matrix für die inverse Multiskalentransformation bezüglich \mathcal{L}^2 -normierter Interpolets bezeichnet. Die Matrix $G \in M(S(L), S(L))$ sei die Matrix für das univariate Finite-Differenzen- oder Kollokationsschema zur Berechnung der ersten Ableitung. G könnte beispielsweise sein:

$$\begin{aligned} G &= \langle \tilde{\Phi}^L, \partial_x \Phi^L \rangle \quad (\text{Kollokationsmethode}) \\ \text{oder } G &= 2^L [-1 \ 0 \ 1]/2 \quad (\text{zentrale Finite-Differenz zweiter Ordnung}). \end{aligned}$$

Im zweidimensionalen Fall hat somit die mit \mathbf{R} , siehe Kapitel 7, umskalierte Koeffizientenmatrix von $\nabla_h \cdot \nabla_h$ die Form

$$T_h^{-1} G T_h T_h^{-1} G T_h \otimes I + I \otimes T_h^{-1} G T_h T_h^{-1} G T_h = T_h^{-1} G^2 T_h \otimes I + I \otimes T_h^{-1} G^2 T_h. \quad (10.10)$$

D.h. die Matrix B aus (7.2) ist in unserem Fall gleich $B = G^2$. Eine der wichtigsten Voraussetzungen für den Nachweis der Vorkonditionierungseigenschaft war die spektrale Äquivalenz von B und $\langle \partial_x \Phi^L, \partial_x \Phi^L \rangle$, Bedingung (7.4). Für den shiftinvarianten Fall war dies gleichbedeutend damit, daß sich die Eigenwerte λ_k von B für $0 \leq k \leq 2^{L-1}$ wie $|\lambda_k| \sim k^2$ verhalten.

Lemma 10.1 (Instabilität von zentralen Diskretisierungen) *Für die Kollokationsdiskretisierung oder shiftinvariante schiefsymmetrische Finite-Differenzen-Schemen ist G^2 nicht spektral äquivalent zu $\langle \partial_x \Phi^L, \partial_x \Phi^L \rangle$ und es gilt $\lambda_{2^{L-1}} = 0$. Der entsprechende Eigenvektor von G ist der even-odd-Mode $(1, -1, 1, -1, \dots, 1, -1)$.*

Beweis: Die Voraussetzungen an das Finite-Differenzen-Schema besagen, daß G eine zirkulante Toeplitz-Matrix $[\dots, a_{-1}, a_0, a_1, \dots]$ ist. Dabei sind $\{a_s\}$ die entsprechenden Koeffizienten des Differenzen-Sterns. Für die Kollokationsmethode hat G trivialerweise diese Form. Die Koeffizienten $\{a_s\}$ erfüllen in beiden Fällen die Bedingung $a_s = -a_{-s}$, so daß G schiefsymmetrisch ist. Die Eigenwerte von G können über das Symbol $a(z) = \sum_s a_s z^s$ berechnet werden. Der 2^{L-1} -te Eigenwert ist

$$a(e^{i2\pi 2^{L-1}/2^L}) = a(-1) = \sum_{s>0} a_s (-1)^s + a_{-s} (-1)^{-s} = 0.$$

Damit ist auch $\lambda_{2^{L-1}} = a^2(-1) = 0$. Der entsprechende Eigenvektor ist der 2^{L-1} -te Fourier-Mode mit den Komponenten $e^{i2\pi s 2^{L-1}/2^L} = e^{i\pi s} = (-1)^s$, $0 \leq s < 2^L$. \square

In Bild 10.1 sind die Eigenwerte $|\lambda_k|$, $0 \leq k \leq 2^{L-1}$, für die Kollokationsmethode bzw. für das Finite-Differenzen-Schema $\frac{1}{2}[-1 \ 0 \ 1]$ für $L = 6$ aufgetragen. Wie üblich verwendet man für die x -Achse die Skalierung $2\pi/2^L$ und für die y -Achse die Skalierung 4^{-L} .

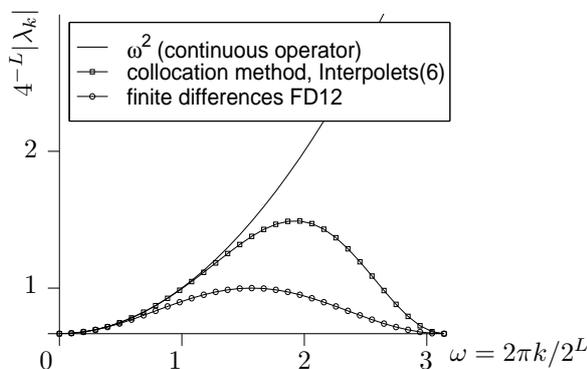


Abbildung 10.1: Skalierte Eigenwerte $4^{-L}|\lambda_k|$ von G^2 für eine Kollokations- bzw. Finite-Differenzen-Diskretisierung von ∂_x . Für $k \approx 2^{L-1}$ sind sie nicht $\sim \omega^2$.

Aus Bild 10.1 bzw. aus Lemma 10.1 ergeben sich zwei Folgerungen:

Bemerkung 10.1 (Parasitäre Oszillationen) Da der hochfrequente even-odd-Mode zum Kern von G gehört, kann es vorkommen, daß die Lösung von (10.9) mit Schachbrett-artigen Oszillationen verschmutzt ist. Im nichtadaptiven Fall können diese theoretisch weggefiltert werden, da die parasitären Eigenvektoren exakt bekannt sind. Im adaptiven Fall können ebenfalls solche Oszillationen auftreten (ohne Bild oder Beweis) jedoch ist hier nicht klar, wie die parasitären bzw. nahezu parasitären Eigenvektoren genau aussehen und wie demnach gefiltert werden müßte.

Bemerkung 10.2 (Vorkonditionieren) Selbst wenn man in der Lage wäre, die parasitären Oszillationen zu beseitigen, würde trotzdem nicht $|\lambda_k| \sim k^2$ gelten. D.h. die Stabilitätskonstanten in (7.4) wären weiterhin beliebig schlecht. Daraus folgt, daß die in Kapitel 7 vorgestellten Vorkonditionierer nicht funktionieren würden.

Wir werden nun stabile Diskretisierungen für die Druck-Poisson-Gleichung vorstellen, deren entscheidende Idee Finite-Differenzen-Verfahren für die Navier-Stokes-Gleichungen auf versetzten Gitter [HW65] entstammt. Im eindimensionalen Fall sind bei so einem Gitter die Aufsetzpunkte für die Geschwindigkeit in den Gitterlinien $x_i = i2^{-L}$ angeordnet und die Aufsetzpunkte $x_{i+1/2}$ für die Druckwerte p_i versetzt genau dazwischen. Der Druckgradient muß in x_i ausgewertet werden, wogegen die Divergenz in $x_{i+1/2}$ ausgewertet wird. Der diskrete Laplace-Operator für die Druck-Poisson-Gleichung entsteht somit durch Hintereinanderschalten des Vorwärts-Differenzen-Sterns $2^L[0 \ -1 \ 1]$ für den Druck-Gradienten und des Rückwärts-Differenzen-Sterns $2^L[-1 \ 1 \ 0]$ für die Divergenz: $\nabla_h \cdot \nabla_h p_i = [(p_{i+1} - p_i)2^L - (p_i - p_{i-1})2^L]2^L = 4^L[p_{i+1} - 2p_i + p_{i-1}]$. Dies ist die übliche stabile Diskretisierung des Laplace-Operators. Durch die geschickte Anordnung der Unbekannten ist die Konsistenzordnung der Vorwärts- bzw. Rückwärts-Differenzen-Sterne in den Aufsetzpunkten x_i bzw. $x_{i+1/2}$ gleich zwei.

Leider ist es mit den adaptiven Multiskalendiskretisierungen nicht ohne Weiteres möglich, die versetzte Anordnung von Geschwindigkeits- und Druck-Werten nachzuahmen, da die Ansatzfunktionen, die die Aufsetzpunkte bestimmen, für \mathbf{u} und p gleich sein sollen. Problemlos möglich ist es aber, das Prinzip "Vorwärts-/Rückwärts-Differenzen" zu übernehmen. Die Stabilisierungstechnik besteht darin, für den Druck-Gradienten eine upwind-artige Finite-Differenzen-Diskretisierung ∇_h^+ zu benutzen und für die Divergenz eine downwind-artige Diskretisierung ∇_h^- . Die Druck-Poisson-Gleichung wird damit $\nabla_h^- \cdot \nabla_h^+(p^{n+1} - p^n) = \frac{1}{\Delta t} \nabla_h^- \cdot \mathbf{u}^*$. Durch dieses Schema bleibt die div-grad-Struktur des Laplace-Operators erhalten und die aufdatierte Geschwindigkeit \mathbf{u}^{n+1} ist diskret divergenzfrei $\nabla_H^- \cdot \mathbf{u}^{n+1} = 0$. Die Konstruktion von geeigneten upwind- bzw. downwind-Differenzen-Schemen ist Gegenstand des nachfolgenden Lemmas.

Lemma 10.2 (Konsistenz von upwind-/downwind-Finite-Differenzen-Schemen)

Aus einem beliebigen Kollokations- oder Finite-Differenzen-Schema zur Berechnung von ∂_x mit Konsistenzordnung K kann durch Addition eines geeignet gewichteten Finite-Differenzen-Schemas zur Berechnung der J -ten Ableitung, $J \geq K + 1$, ein Finite-Differenzen-Schema für ∂_x konstruiert werden, das ebenfalls die Konsistenzordnung K hat.

Konkret: Sei G die $S(L) \times S(L)$ -Matrix des Kollokations- oder Finite-Differenzen-Schemas für ∂_x und F die $S(L) \times S(L)$ -Matrix des Finite-Differenzen-Schemas für die J -te Ableitung, dann ist für beliebiges c

$$G + c2^{L(1-J)}F$$

die Matrix eines Finite-Differenzen-Schemas für ∂_x mit Konsistenzordnung $\geq K$.

Speziell für den Fall daß J gerade und $c > 0$ ist, entspricht $G + c2^{L(1-J)}F$ einem upwind-Schema und $G - c2^{L(1-J)}F$ einem downwind-Schema. Beispielsweise kann der upwind-Stern $[0 \ -1 \ 1]$ aus den zentralen Schemen $[-1 \ 0 \ 1]/2$ und $[1 \ -2 \ 1]$ konstruiert werden:

$$[0 \ -1 \ 1] = [-1 \ 0 \ 1]/2 + [1 \ -2 \ 1]/2.$$

Beweis: (Lemma 10.2). Eine Matrix E ist genau dann die Darstellung eines Finite-Differenzen- oder Kollokationsschemas für die Berechnung der r -ten Ableitung mit Konsistenzordnung P ,

falls für $s \in S(L)$ gilt [Kos00]

$$\sum_t E_{st}(s-t)^i = \begin{cases} 0 & \text{für } i = \{0, \dots, r+P-1\} \setminus r \\ 2^{rL} & \text{für } i = r \\ O(2^{rL}) & \text{für } i \geq r+P \end{cases}. \quad (10.11)$$

Nach Voraussetzung erfüllen G und F diese Bedingung mit $r = 1$, $P = K$ bzw. $r = J \geq K+1$ und $P \geq 0$. Aus diesem Grund erfüllt die zusammengesetzte Matrix $G + c2^{L(1-J)}F$ diese Bedingung mit $r = 1$ und $P \geq K$. \square

Wir untersuchen nun den Stabilisierungseffekt durch die upwind-/downwind-Diskretisierung. Die Koeffizientenmatrix für den resultierenden Schur-Komplement-Operator $\nabla_h^- \cdot \nabla_h^+$ kann als

$$T_h^{-1} \underbrace{(G - c2^{L(1-J)}F)(G + c2^{L(1-J)}F)}_{=:B} T_h \otimes I + I \otimes T_h^{-1}(G - c2^{L(1-J)}F)(G + c2^{L(1-J)}F)T_h$$

geschrieben werden, vergleiche (10.10). Speziell für periodische Randbedingungen sind G und F typischerweise zirkulante Toeplitz-Matrizen, so daß $GF = FG$ und damit

$$B = G^2 - c^2 2^{2L(1-J)} F^2. \quad (10.12)$$

Außerdem haben G und F eine gemeinsame Basis von Eigenvektoren, die Fourier-Moden, so daß die Eigenwerte λ_k von B einfach aus den Eigenwerten μ_k von G und η_k von F berechnet werden können:

$$\lambda_k = \mu_k^2 - c^2 2^{2L(1-J)} \eta_k^2.$$

Falls G eine schiefsymmetrische Matrix ist, wie dies für die Kollokationsmethode oder die üblichen zentralen Finite-Differenzen-Schemen der Fall ist, dann ist μ_k rein imaginär und μ_k^2 eine reelle Zahl $\mu_k^2 \leq 0$. Falls F eine symmetrische Matrix ist, wie dies für die üblichen zentralen Finite-Differenzen-Schemen für *gerade* Ableitungen der Fall ist, dann ist η_k rein reell und $\eta_k^2 \geq 0$. Der Trick ist nun, daß für den kritischen Bereich $k \approx 2^{L-1}$, für den $|\mu_k^2|$ *nicht* $\sim k^2$ ist (siehe Bild 10.1), $|\eta_k| \sim k^J \sim 2^{JL}$ sein kann, so daß insgesamt $\lambda_k \sim k^2$.

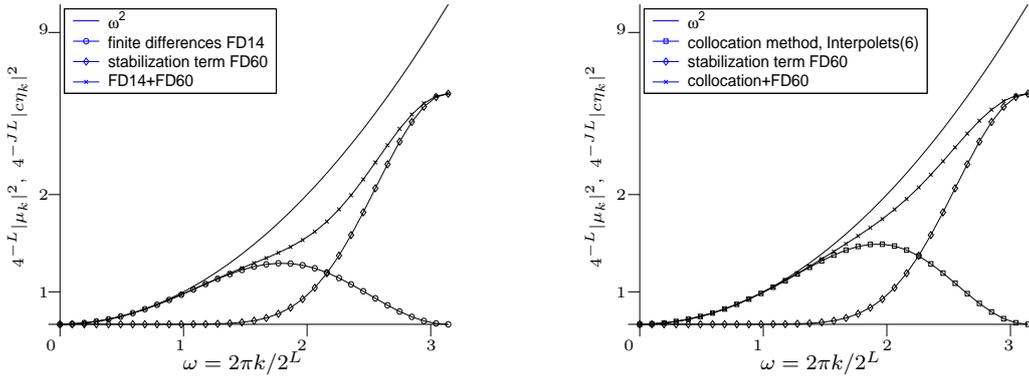
Um dies zu veranschaulichen, haben wir in Bild 10.2 die skalierten Eigenwerte $4^{-L}|\lambda_k|$, $4^{-L}|\mu_k|^2$ und $|c2^{-JL}\eta_k|^2$ für zwei besonders interessante Verfahren aufgetragen: die Kollokationsmethode mit Interpolants der Ordnung $N = 6$, d.h. Konsistenzordnung $K = 5$, und einer Finite-Differenzen-Stabilisierung mit $J = 6$

$$F = 2^{6L}[1 \quad -6 \quad 15 \quad -20 \quad 15 \quad -6 \quad 1]/720, \quad c = 30$$

sowie das zentrale Finite-Differenzen-Schema $G = 2^L[1 \quad -8 \quad 0 \quad 8 \quad -1]/12$ mit Konsistenzordnung $K = 4$ und der gleichen Stabilisierung wie bei der Kollokationsmethode. Offensichtlich gilt für beide Verfahren im Bereich $0 \leq \omega := 2\pi k/2^L \leq 1.5$ noch sehr gut $|\mu_k|^2 \sim k^2$ bzw. gleichbedeutend $4^{-L}|\mu_k|^2 \sim \omega^2$. Diese Eigenschaft geht aber für $\omega \rightarrow \pi$ mehr und mehr verloren. In genau diesem Bereich fängt aber der Stabilisierungsterm $2^{-JL}\eta_k$ an zu wirken und führt somit zu $|\lambda_k| \sim k^2$ für den kompletten ω -Bereich.

Diese Argumentation kann man auch in eine exakte mathematische Form bringen.

Lemma 10.3 (Stabilisierung von Div-Grad-Operatoren) *Die Matrizen G und F seien zirkulante Toeplitz-Matrizen mit reellen Generatoren $\{a_s\}$ bzw. $\{b_s\}$. G sei schiefsymmetrisch und F sei symmetrisch, d.h. $a_s = -a_{-s}$ und $b_s = b_{-s}$. Die zugehörigen Symbole erfüllen*

Abbildung 10.2: Skalierte Eigenwerte von G^2 , $(cF)^2$ und B .

für ein $\omega_0 < \pi$ die folgenden Bedingungen

$$\begin{aligned} |a(e^{i\omega})| &\sim \omega \quad \text{und} \quad |b(e^{i\omega})| \sim \omega^J \quad \text{für } 0 \leq \omega \leq \omega_0 \\ |a(e^{i\omega})| &\lesssim 1 \quad \text{und} \quad |b(e^{i\omega})| \sim 1 \quad \text{für } \omega_0 \leq \omega \leq \pi. \end{aligned} \quad (10.13)$$

Dann gilt für die Eigenwerte λ_k von B aus (10.12) $|\lambda_k| \sim k^2$ für $0 \leq k \leq 2^{L-1}$. D.h. B ist spektral äquivalent zur Steifigkeitsmatrix $\langle \partial_x \Phi^L, \partial_x \Phi^L \rangle$. Die Äquivalenzkonstanten sind unabhängig von L .

Beweis: Für zirkulante Toeplitz-Matrizen sind die Eigenwerte direkt aus dem Symbol der Generatoren berechenbar. In unserem Fall ist $\mu_k = 2^L a(e^{-i2\pi k/2^L})$ und $\eta_k = 2^{JL} b(e^{-i2\pi k/2^L})$. Also

$$\lambda_k = 4^L [a^2(e^{-i2\pi k/2^L}) - c^2 b^2(e^{-i2\pi k/2^L})].$$

Da G schief-symmetrisch ist und F symmetrisch, ist $a^2(e^{-i2\pi k/2^L}) = a^2(e^{i2\pi k/2^L}) \leq 0$ bzw. $b^2(e^{-i2\pi k/2^L}) = b^2(e^{i2\pi k/2^L}) \geq 0$. Speziell für $k = 0$ ist wegen der Konsistenzbedingung an G und F : $\mu_0 = a(1) = \sum_s a_s = 0$ bzw. $\eta_0 = b(1) = \sum_s b_s = 0$. Für $1 \leq k \leq \omega_0 2^L / (2\pi)$ erhalten wir mit (10.13)

$$k^2 \lesssim 4^L \left(\frac{k^2}{4^L} + c^2 \frac{1}{4^{JL}} \right) \lesssim |\lambda_k| \lesssim 4^L \left(\frac{k^2}{4^L} + c^2 \frac{\omega_0^2 4^{JL}}{4^{JL}} \right) \lesssim k^2.$$

Analog erhalten wir für $\omega_0 2^L / (2\pi) \leq k \leq 2^{L-1}$ die Aussage $|\lambda_k| \sim k^2 \sim 4^L$. \square

Beenden wollen wir diesen Abschnitt mit einigen Bemerkungen zur praktischen Realisierung der Stabilisierungstechnik.

Bemerkung 10.3 (Wahl von c) Der Parameter c wird so gewählt, daß für die Diagramme aus Bild 10.2 augenscheinlich $|\lambda_k| \sim k^2$ mit guten Äquivalenzkonstanten gilt. Eine Optimierung bezüglich der Konditionszahlen der resultierenden Operatormatrizen aus Theorem 7.1 wäre möglich, ist aber noch nicht implementiert.

Bemerkung 10.4 (Numerische Kosten) Für das oben erwähnte Finite-Differenzen-Schema mit Konsistenzordnung 4 besteht die Stabilisierungstechnik lediglich darin, daß in AFD bei der Berechnung von $\mathbf{v}^{S^e(l)}$ der Stern $2^l [1 \ -8 \ 0 \ 8 \ -1] / 12$ durch den etwas breiteren Stern

$$2^l [0 \ 1 \ -8 \ 0 \ 8 \ -1 \ 0] / 12 \pm c^2 2^l [1 \ -6 \ 15 \ -20 \ 15 \ -6 \ 1] / 720$$

ersetzt wird. Somit entstehen zusätzliche Kosten nur durch die größere Breite des neuen Sterns. Bei der Kollokationsmethode ist hingegen neben APG nun auch AFD für die Berechnung des Stabilisierungsterms auszuführen, so daß hier die Kosten spürbar steigen.

Bemerkung 10.5 (Ähnliche Arbeiten) Für ein spezielles Finite-Differenzen-Schema auf uniformen, nichtversetzten Gittern wurde in [Str97] eine sehr ähnliche Stabilisierungstechnik vorgestellt. Auch dort wird ein zentraler Differenzen-Stern durch explizites Hinzufügen eines Sterns für eine Ableitung höherer Ordnung stabilisiert. Allerdings entspricht dieser Term nicht wie hier einer geraden Ableitung, sondern einer ungeraden Ableitung, die aber selbst wiederum mit einem upwind- oder downwind-Stern diskretisiert ist.

Bemerkung 10.6 (Laplace-Operatoren mit variablen Koeffizienten) Die hier untersuchten Div-Grad-Operatoren treten auch bei der Diskretisierung von elliptischen Operatoren mit variablen Koeffizienten auf, z.B. bei $\nabla \cdot a(\mathbf{x})\nabla u(\mathbf{x}) = f(\mathbf{x})$. Solche Operatoren können in der Form $\nabla^- \cdot (a \cdot \nabla^+)$ diskretisiert werden. Sofern die Koeffizientenfunktion a dies erlaubt, erhält man auf diese Weise auch wieder stabile diskrete Operatoren.

10.3 Numerische Experimente: 2D Scherströmung

Eine der wichtigsten in der Realität auftretenden Strömungskonfigurationen ist die turbulente Scherschicht. Dabei grenzen zwei Fluid-Bereiche unterschiedlicher Geschwindigkeit aneinander und es kommt zu Verwirbelungen an der Grenzfläche. Für unsere numerischen Experimente betrachten wir eine sich zeitlich entwickelnde Scherschicht. Die Anfangsbedingung ist ein laminares Geschwindigkeitsfeld mit einem vorgegebenen Profil $u(y)$ für die u -Komponente und $v = 0$, siehe Bild 10.3. Diese Anfangsbedingung stellt eine exakte stationäre Lösung

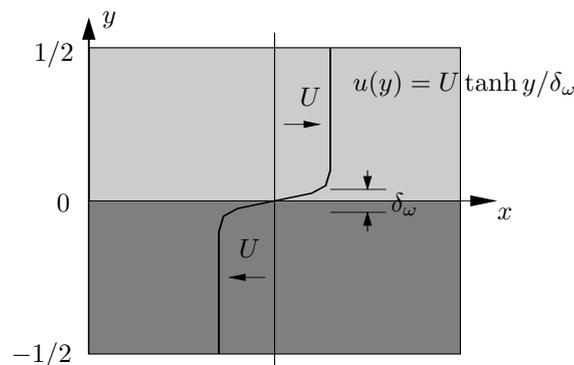


Abbildung 10.3: Geschwindigkeitsprofil der Anfangsbedingung für die Scherschicht. Die Rotationsgrenzschichtdicke ist als $\delta_\omega = 2U / \max_y |\partial_y u(y)|$ definiert.

der Navier-Stokes-Gleichungen für den Fall $\nu = 0$ dar, d.h. für reibungsfreie Fluide. Jedoch ist diese Lösung nicht stabil [Mic64]. Kleine Störungen werden durch Kelvin-Helmholtz-Instabilitäten verstärkt und es kommt zum Aufrollen von Wirbeln. Diesen Effekt beobachtet man auch bei reibungsbehafteten Fluiden. In Bild 10.7 ist die zeitliche Entwicklung eines solchen Prozesses an Hand der Rotation $\omega = \nabla \times \mathbf{u}$ dargestellt. Die maximale Anfangsgeschwindigkeit ist hier $U = 0.015$ und $\nu = 1.26e-6$. Um die Instabilitäten anzufachen, wurde das laminaire Geschwindigkeitsprofil unmittelbar in der Grenzschicht mit einem weißen Rauschen überlagert.

Neben dem Entstehen und Aufrollen von Kelvin-Helmholtz-Wirbeln beobachtet man auch,

daß die Wirbel im Laufe der Zeit verschmelzen. Dieser Effekt, bei dem aus vielen kleinen Wirbeln einige wenige große Wirbel entstehen, ist charakteristisch für zweidimensionale Strömungen, vergleiche Kapitel 4.6 [Cho98]. Dabei handelt es sich letztlich um einen Energie-Transfer von feinen zu groben Strukturen, d.h. von hohen zu niedrigen Leveln im Multiskalensinne [Nov76]. Zweidimensionale Turbulenz verhält sich damit fundamental anders, als dreidimensionale Turbulenz, bei der es zu einem Energie-Transfer von niedrigen zu hohen Leveln kommt (Zerfall von groben Strukturen zu feineren Strukturen). Die Experimente im nächsten Abschnitt sind ein sehr schönes Beispiel dafür. Im aktuellen Abschnitt soll zunächst aber untersucht werden, wie sich das adaptive Verfahren für die zweidimensionale Scherschicht verhält. Da für dieses Problem keine exakten Lösungen bekannt sind, vergleichen wir die Ergebnisse mit den Ergebnissen eines nichtadaptiven Fourier-Spektral-Codes mit sehr hoher Auflösung. Die letzteren Ergebnisse sind demnach als nahezu exakte Lösungen anzunehmen. Die beiden Verfahren sind durch die folgenden Parameter gekennzeichnet:

	adaptives Multiskalenverfahren	Spektral-code
Zeitdiskretisierung:	AB3	AB3
Zeitschrittweite	0.005	0.005
Ortsdiskretisierung	Interpolets $N = 6$, WENO5	512×1024 Fourier-Moden
Adaptivität	$R_{Level} = R_{Ort} = 1$, $\delta = 8e-4$ maximales Level $L = 10$	

Um periodische Randbedingungen verwenden zu können, wurden jeweils zwei ungefähr spiegelverkehrte Scherschichten im Gebiet $]0, 1[\times] - 1/2, 3/2[$ simuliert. Die Anfangsbedingung entsteht durch Spiegeln an der Achse $y = 1/2$, vergleiche Bild 10.3. Das maximale Level wurde auf 10 begrenzt, damit die aus der CFL-Bedingung resultierenden Zeitschrittweiten nicht allzu klein werden.

In Bild 10.8 (links, Mitte) vergleichen wir zunächst qualitativ die Resultate der beiden Rechnungen mittels Niveau-Linien der Rotation. Offensichtlich stimmen die adaptiven und die nichtadaptiven Ergebnisse gut über ein. Jedoch treten bei der adaptiven Rechnungen wieder Oszillationen auf, die dazu führen, daß die Niveau-Linien im Bereich der Filamente zwischen zwei Wirbeln unterbrochen sein können (siehe dritte Bild-Zeile). Auf der rechten Seite von Bild 10.8 sind die adaptiven Gitter $\Omega_{\mathcal{T}}$ für die unterschiedlichen Zeitpunkte dargestellt. Deutlich zu sehen ist die Konzentration der Gitterpunkte im Bereich der Wirbel und das Ausdünnen der Gitter mit zunehmender Zeit, da die Lösung immer glatter wird.

Der zeitliche Verlauf der Zahl der Freiheitsgrade ist in 10.4 dargestellt. Die angegebenen Zahlen beziehen sich auf das komplette periodisierte Problem. Für eine Scherschicht wäre ungefähr jeweils die Hälfte der Freiheitsgrade ausreichend. Da die Anfangsbedingung durch das aufgebrachte Rauschen nicht sehr glatt ist, sind zu Beginn viele Freiheitsgrade notwendig, um \mathbf{u} gut darstellen zu können. Jedoch wird \mathbf{u} durch die Diffusion schnell geglättet, so daß $|\mathcal{T}|$ schnell abfallen kann ($0 < t < 4$). Dieser Prozess wird durch die Kelvin-Helmholtz-Instabilitäten und die Bildung komplexerer Strukturen, wie Wirbel und Filamente, gestoppt und die Zahl der Freiheitsgrade steigt wieder ($4 < t < 12$). Durch das Verschmelzen der Wirbel zu wenigen großen Wirbeln vereinfachen sich die Strukturen jedoch, so daß $|\mathcal{T}|$ wieder abfallen kann.

Als letztes vergleichen wir die beiden statistischen Größen Energie $E(t)$ und Enstrophie $Z(t)$

$$E(t) := \int_{\Omega} |\mathbf{u}(\mathbf{x}, t)|^2 d\mathbf{x} \quad \text{bzw.} \quad Z(t) := \frac{1}{2} \int_{\Omega} |\boldsymbol{\omega}(\mathbf{x}, t)|^2 d\mathbf{x} .$$

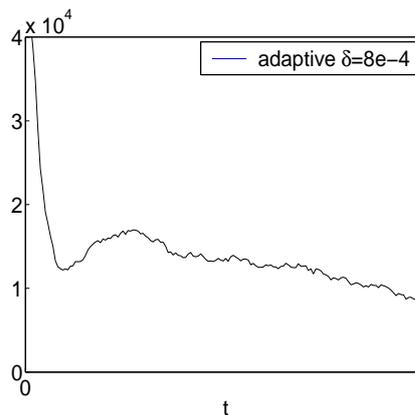


Abbildung 10.4: *Zeitliche Entwicklung der Zahl der Freiheitsgrade für die adaptive Rechnung.*

Diese Größen sind von besonderem Interesse für Langzeitsimulationen turbulenter Strömungen. In diesem Fall sind punktweise Vergleiche der numerischen Lösung mit der exakten Lösung nicht mehr möglich und auch nicht sinnvoll. Vielmehr ist man daran interessiert, daß die numerische Lösung die gleichen statistischen Eigenschaften hat, wie die exakte Lösung. So soll z.B. die Energie in beiden Fällen mit der gleichen Rate dissipiert werden. Analog für die Enstrophie. Der Vergleich der adaptiven Rechnung mit den Ergebnissen des Spektral-Codes ist hier besonders interessant, da Fourier-Spektral-Verfahren konservativ sind und sie somit die Energiedissipation nahezu exakt berechnen. In Bild 10.5 ist die zeitliche Entwicklung von Energie und Enstrophie für beide Rechnungen abgebildet. Offensichtlich stimmen die Ergebnisse sehr gut überein, so daß das adaptive Verfahren wirklich für die Simulation turbulenter Strömungen geeignet ist.

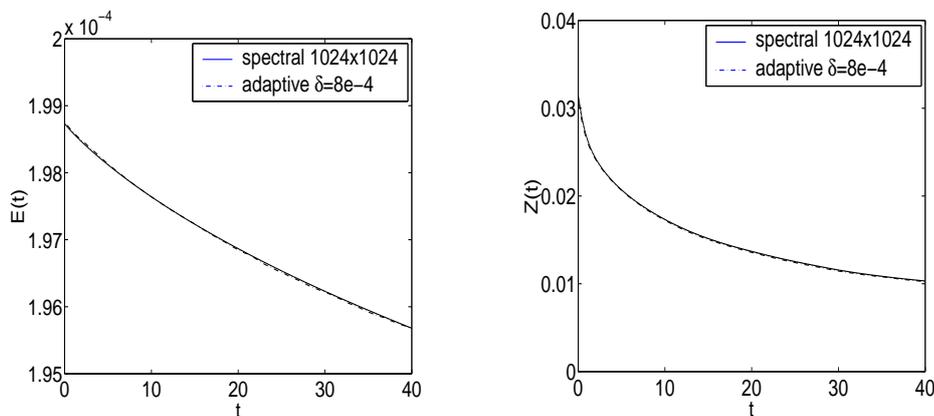


Abbildung 10.5: *Zeitliche Entwicklung von Energie und Enstrophie.*

10.4 Numerische Experimente: 3D Scherströmung

Wie wir im vorangegangenen Abschnitt gesehen hatten, kommt es bei zweidimensionalen Strömungen zum sukzessiven Verschmelzen von Wirbeln mit gleichem Drehsinn. Bei dreidimensionalen Strömungen gibt es diesen Effekt nicht mehr. Die zusätzliche dritte Dimension

erlaubt es, Wirbelröhren zu beugen oder zu falten und stark in die Länge zu ziehen, bis sie auseinanderreißen. Im gleichen Maße, wie die Wirbelröhren gestreckt werden, nimmt ihr Querschnitt ab, so daß auf diese Weise feinskaligere Strukturen entstehen. Somit wird Energie auf feinere Skalen transportiert. Diese Energiekaskade schreitet rekursiv solange voran, bis die Strukturen so klein sind, daß die relative Reynoldszahl klein ist und Energie praktisch nur noch durch Dissipation verloren geht. In Bild 10.9 sind einzelne Phasen einer sich entwickelnden dreidimensionalen Scherschicht zu sehen, aus denen die Zunahme an Komplexität deutlich wird. Die Anfangsbedingung für die entsprechende Rechnung ist die dreidimensionale Version der Anfangsbedingung aus Abschnitt 10.3. Die Viskosität ist jedoch größer als dort gewählt worden, $\nu = 5.04e-6$. Die restlichen Verfahrensparameter stimmen mit denen der 2D Rechnung überein.

Bei einer solchen turbulenten Scherschicht ist es nicht mehr sinnvoll, momentane Geschwindigkeitsfelder direkt z.B. mit Referenzlösungen zu vergleichen. Stattdessen betrachtet man üblicherweise statistische Größen, für deren Verhalten es theoretische Aussagen gibt. So ist z.B. in [HH84, Tow76] gezeigt worden, daß nach einer Anlaufphase, in der die Strömung die Anfangsbedingungen "vergift", die integrale Impuls-Grenzschichtdicke

$$\delta_m(t) := \int_{-1/2}^{1/2} \left(\frac{1}{4} - \frac{\bar{u}(y,t)}{2U} \right) dy \quad \text{wobei} \quad \bar{u}(y,t) := \int_{x,z} u(x,y,z,t) dx dz$$

linear anwächst. In unserem Experiment war dies nach etwa 60 Zeiteinheiten der Fall, siehe Bild 10.6. Diese Phase geht bis etwa 100 Zeiteinheiten. Danach kommt es zu einem Abflachen des Anstiegs, da in dem endlichen Berechnungsgebiet die Grenzschichtdicke nicht beliebig groß werden kann. In der Phase des linearen Anwachsens der Grenzschichtdicke sollen sich weiterhin z.B. die Profile der mittleren Geschwindigkeit selbstähnlich verhalten, d.h. $\bar{u}(y/\delta_m(t), t)$ soll nur von $y/\delta_m(t)$ abhängig sein, aber nicht mehr von t selbst. In Bild 10.6 sind für $t = 60, 70, \dots, 120$ die Profile $\bar{u}(\cdot, t)$ im entsprechend skalierten Koordinatensystem aufgetragen. Offensichtlich stimmen die Rechnungen gut mit der Vorhersage überein. Ein analoges

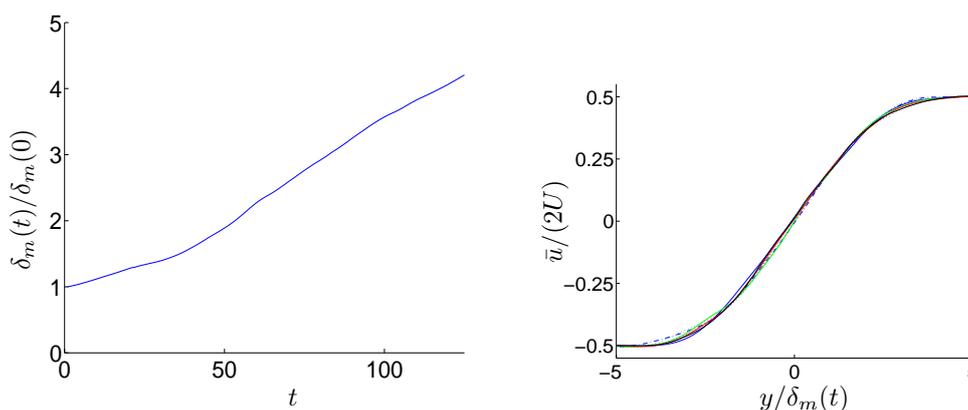


Abbildung 10.6: Links: Zeitliche Entwicklung der Impuls-Grenzschichtdicke. Rechts: Selbstähnliche Profile der mittleren Geschwindigkeitskomponente $\bar{u}(\cdot, t)$ für $t = 60, \dots, 120$.

selbstähnliches Verhalten wird auch für statistische Momente höherer Ordnung bzw. r.m.s.-Werte vorhergesagt. Jedoch sind hier große Anzahlen von Stichproben notwendig, um eine auskonvergierte Statistik zu erhalten. Die momentan verfügbaren Daten reichen dafür noch nicht aus.

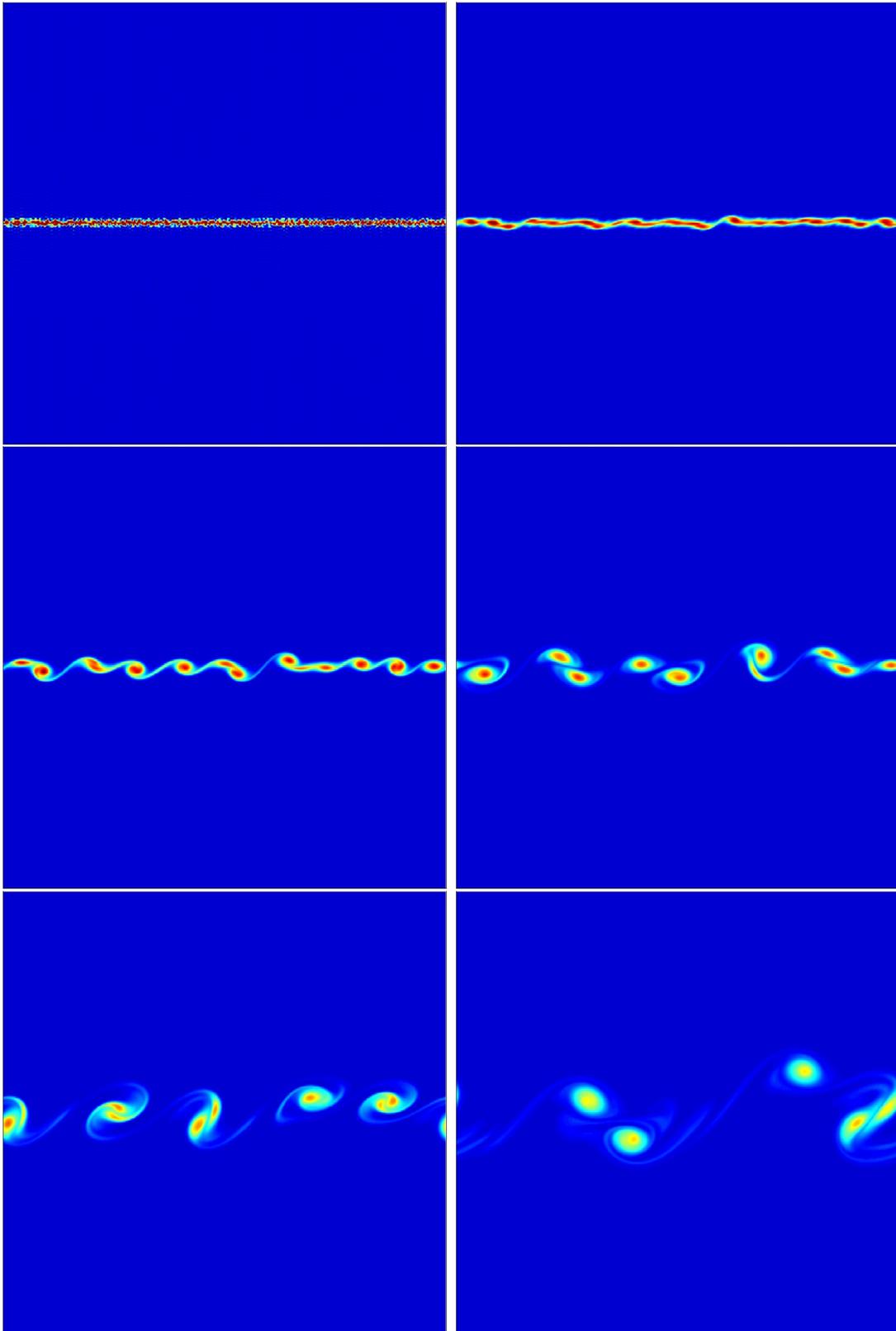


Abbildung 10.7: Zeitliche Entwicklung einer Scherschicht: Rotation zu den Zeitpunkten $t = 0, 4, 8, 16, 24, 40$. Die Daten wurden mit einem Fourier-Spektral-Code berechnet.

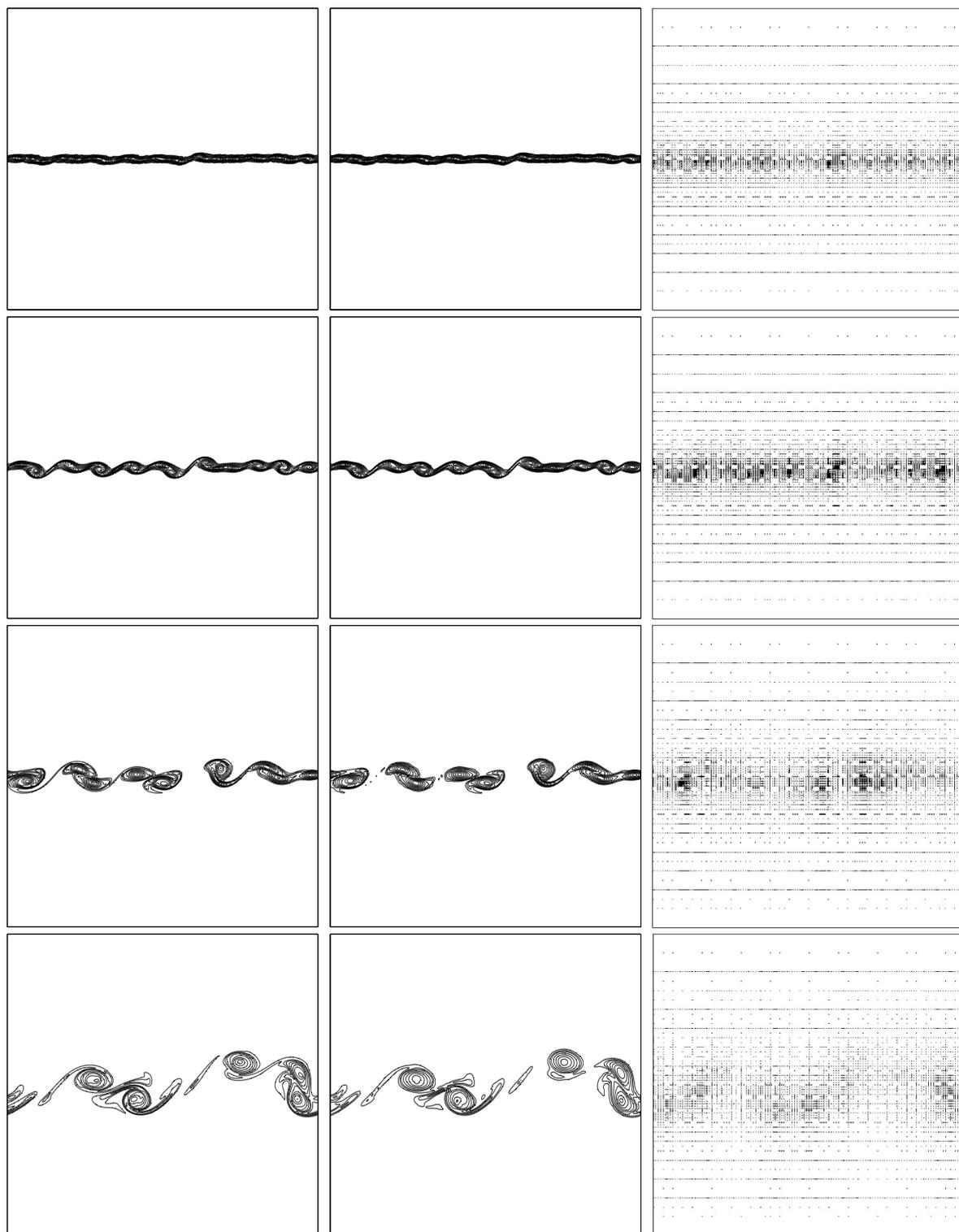


Abbildung 10.8: *Niveau-Linien der Rotation für die Werte $\{0.1, 0.2, 0.3, 0.5, 0.75, 1.0, \dots, 2.25, 2.5\}$ zu den Zeitpunkten $t = 4, 8, 16, 36$. Links: Spektral-Code. Mitte: adaptives Wavelet-Verfahren. Rechts: adaptive Gitter.*

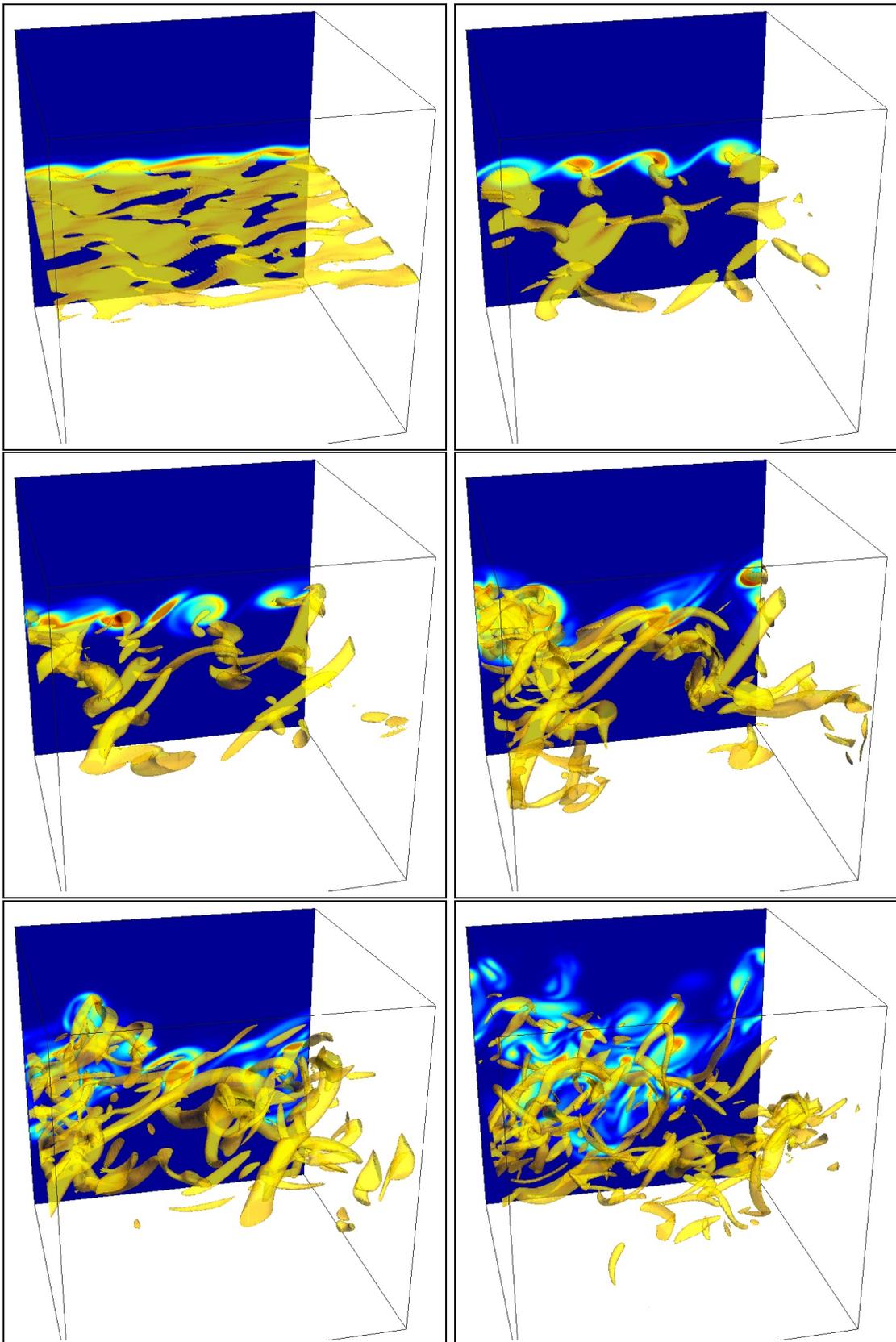


Abbildung 10.9: Niveauflächen/Schnittdarstellung der Rotation für die turbulente Scherschicht zu den Zeitpunkten 0, 15, 30, 60, 90, 150.

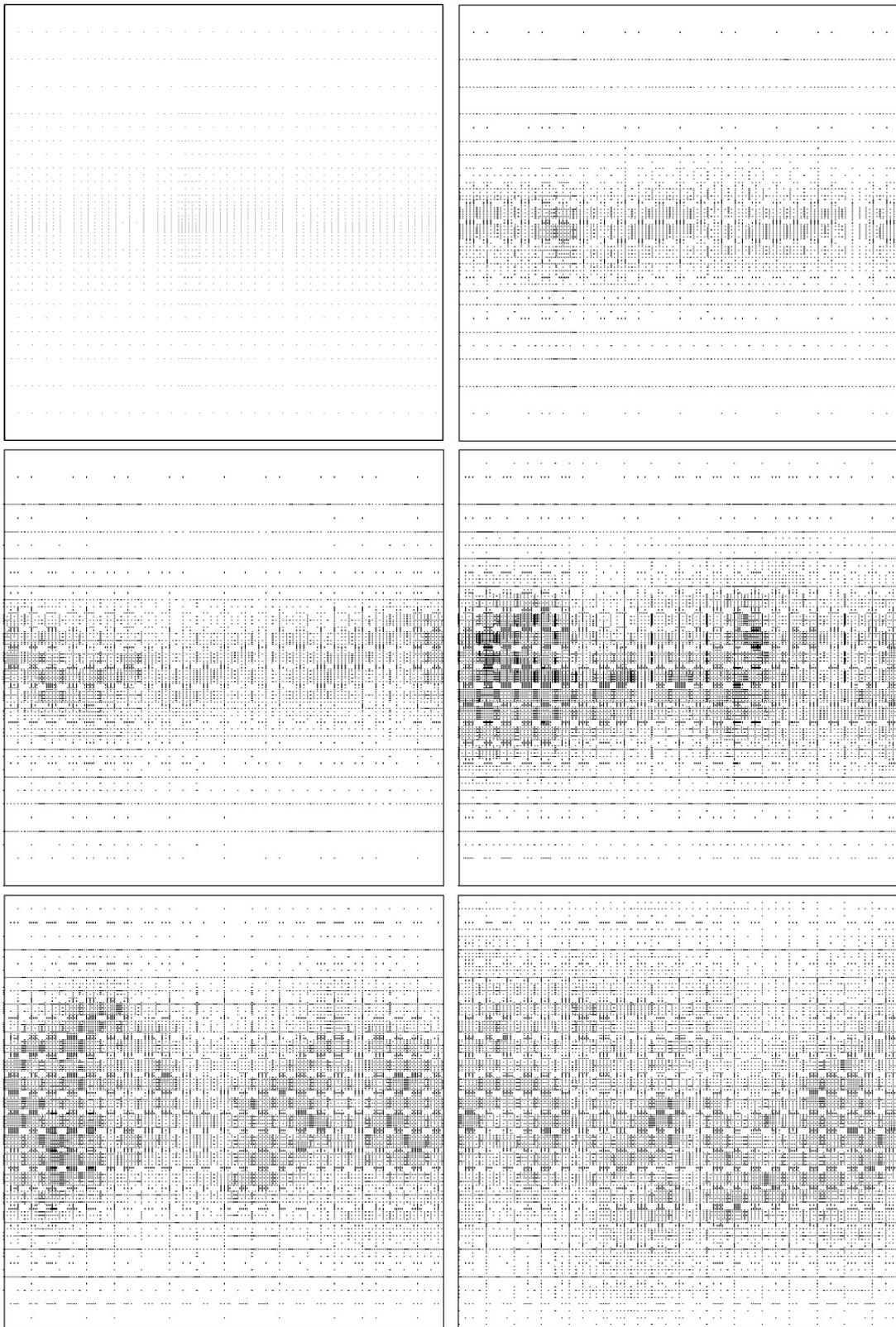


Abbildung 10.10: Schnitte ($z = 0.5$) durch die adaptiven Gitter zu den Zeitpunkten 0, 15, 30, 60, 90, 150.

Kapitel 11

Zusammenfassung und Ausblick

In dieser Arbeit wurden Lösungsverfahren für partielle Differentialgleichungen vorgestellt und analysiert, die auf adaptiven Diskretisierungen mit anisotropen Tensorprodukten von Interpolets basieren. Diese Verfahren vereinen die guten Approximationseigenschaften von anisotropen Tensorprodukt-Multiskalenfunktionen mit der algorithmischen Einfachheit der Interpolets, welche unter anderem eine schnelle approximative Auswertung nichtlinearer Funktionale erlaubt.

Für die eigentliche Diskretisierung wurden Finite-Differenzen-Schemen bzw. das biorthogonale Petrov-Galerkin-Schema betrachtet, bei dem die Testfunktionen die dualen Multiskalenfunktionen sind. In beiden Fällen kann für einfache Differentialoperatoren, wie D^{re_i} , die Matrix-Vektor-Multiplikation mit der Steifigkeitsmatrix durch zeilenweises Anwenden von univariaten Algorithmen realisiert werden. Die Operatorauswertung ist damit deutlich einfacher und schneller als beispielsweise beim Galerkin-Verfahren.

Durch die Finite-Differenzen-Technik ist es sehr leicht möglich, spezielle an die Differentialgleichung angepaßte Diskretisierungen zu konstruieren. Beispiele sind die Verwendung von WENO-Schemen für Konvektionsprobleme oder die in Kapitel 10 entwickelten stabilen Finite-Differenzen-Sterne für die Druck-Poisson-Gleichung, die aus einem Sattelpunktproblem resultiert.

Der Mangel der Interpolets, daß nur für einen kleinen Bereich von Sobolev-Räumen Normäquivalenzen gelten, kann gegebenenfalls durch einen Basiswechsel zu den Lifting-Wavelets behoben werden. Eine Anwendung davon ist der in Kapitel 7 vorgestellte Lifting-Vorkonditionierer. Für den einfachen nichtadaptiven Fall konnte für diesen Vorkonditionierer gezeigt werden, daß die l_2 -Konditionszahlen der resultierenden linearen Gleichungssysteme unabhängig von der Dimension d des Gebietes für Interpolets höherer Ordnung, $N \geq 4$, von der Größenordnung 1 sind und für die hierarchische Basis, $N = 2$, von der Größenordnung $\sqrt{L}2^L$. L bezeichnet das maximale Diskretisierungslevel. Mit dem Lifting-Vorkonditionierer ist es damit erstmalig gelungen, nahezu optimal vorkonditionierte elliptische Gleichungssysteme für Finite-Differenzen-Diskretisierungen auf dünnen Gittern zu erreichen. Für den einfachen Diagonal-Vorkonditionierer ohne Lifting-Wavelettransformation konnte die obere Abschätzung

$$\frac{\|\mathbf{A}_h \mathbf{D}^{-1} \cdot \mathbf{u}\|_{l_2}}{\|\mathbf{u}\|_{l_2}} \lesssim \begin{cases} 2^L & \text{für } N = 2 \\ \sqrt{2^L} & \text{für } N \geq 4 \end{cases} \quad \text{für } \mathbf{u} \neq 0$$

gezeigt werden. Die entsprechenden Abschätzungen nach unten sind leider noch nicht befrie-

digend. Hier besteht aber die Vermutung, daß

$$\frac{\|\mathbf{A}_h \mathbf{D}^{-1} \cdot \mathbf{u}\|_{l_2}}{\|\mathbf{u}\|_{l_2}} \gtrsim \begin{cases} L^{-1} & \text{für } N = 2 \\ 1 & \text{für } N \geq 4 \end{cases} .$$

Die resultierenden Konditionszahlen, $L2^L$ bzw. $\sqrt{2^L}$, würden damit unabhängig von d nur noch moderat vom maximalen Level abhängen. In den numerischen Experimenten in Abschnitt 7.2 haben wir dieses Verhalten tatsächlich beim iterativen Lösen von elliptischen Gleichungssystemen beobachtet – sowohl im nichtadaptiven Fall als auch im adaptiven Fall. Für Interpolets höherer Ordnung stellt das einfache diagonale Skalieren damit eine echte Alternative zum aufwendigeren Lifting-Vorkonditionierer dar.

Den zweiten Theorie-Schwerpunkt dieser Arbeit bildet die Konvergenzanalyse des Finite-Differenzen- und des biorthogonalen Petrov-Galerkin-Schemas am Beispiel des Poisson-Problems. Die Idee war hier, mittels der vorausgesetzten $(s, s - 2)$ -Stabilität der diskreten Operatoren den Konvergenzfehler über den Interpolationsfehler und den Konsistenzfehler abzuschätzen:

$$\|u - u_{\mathcal{T}}\|_{\mathcal{H}^s} \leq \|u - I_{\mathcal{T}}u\|_{\mathcal{H}^s} + \frac{1}{\alpha_s} \|A_{\mathcal{T}} I_{\mathcal{T}}u - I_{\mathcal{T}}Au\|_{\mathcal{H}^{s-2}} .$$

Wir haben uns dann auf den schwierigeren Fall von Abschätzungen für den Konsistenzfehler beschränkt. Die Vorgehensweise bestand aus den Schritten:

1. Wähle eine Indexmenge, die an die Glattheit von u , den Operator und die Fehler-Norm angepaßt ist.
2. Abschätzen des Fehlers mittels des Abfallverhaltens der Multiskalenkoeffizienten und der einseitigen Normabschätzungen.

Ein Problem dabei ist, daß die Interpolets leider keine Normabschätzungen für negative Sobolev-Normen, $s - 2 < 0$, erlauben, so daß es nicht möglich war, den Einfluß des Operators auf die Indexmengen \mathcal{T} durch die Fehler-Norm zu kompensieren. Dadurch waren die Indexmengen \mathcal{T} deutlich größer als Indexmengen, die aus der Best- n -Term-Approximation von u bzgl. $\|\cdot\|_{\mathcal{H}^s}$ resultieren. Für ein biorthogonales Petrov-Galerkin-Schema mit hinreichend glatten dualen Multiskalenfunktionen würde das Konzept aber durchgehen.

Ein weiterer Schwerpunkt dieser Arbeit war die Anwendung des Lösungsverfahrens auf diverse partielle Differentialgleichungen: Poisson-Gleichungen, Konvektionsprobleme und die Navier-Stokes-Gleichungen. Dazu haben wir aus der Literatur bekannte, einfache Strategien für die Steuerung der Adaptivität übernommen bzw. leicht modifiziert. Speziell für das Poisson-Problem wurde gezeigt, wie Teile der in [CDD01] vorgestellten Konvergenztheorie für adaptive Wavelet-Galerkin-Verfahren auf das Schätzen des Fehlers bei unseren Finite-Differenzen-Diskretisierungen übertragen werden können. Das darauf basierende Verfeinerungskriterium besteht darin, jeweils Nachbarn von Indizes (\mathbf{l}, \mathbf{t}) , für die der Multiskalenkoeffizient der aktuellen Lösung multipliziert mit $2^{-|\mathbf{l}|/2}$ größer als ein Schwellwert ist, in die verfeinerte Indexmenge aufzunehmen. Jedoch hat sich dieses Vorgehen als weniger robust erwiesen, als die modifizierte Strategie, bei der auf den Gewichtungsfaktor $2^{-|\mathbf{l}|/2}$ verzichtet wird. Mit dieser vereinfachten Strategie konnte beim Poisson-Problem mit einer glatten Lösung für das Petrov-Galerkin- und das Finite-Differenzen-Schema genau die Konvergenzrate erzielt werden, die für die benutzten Ansatzfunktionen bzw. die resultierende Konsistenzordnung der Diskretisierung zu erwarten war. Beim Poisson-Problem mit einer nahezu singulären Lösung entsprach die Konvergenzrate nur für die Finite-Differenzen-Technik mit dem AFD-Schema

der vermutlich optimalen Approximationsordnung. Bei der Kollokationsmethode hat hingegen offenbar die reduzierte Konsistenzordnung zu einer nicht optimalen Konvergenzrate geführt. Für die zeitabhängigen Konvektionsprobleme wurde als Verfeinerungsstrategie die Hartensche Heuristik benutzt, bei der ebenfalls die Nachbarn von Indizes eingefügt werden, für die die Multiskalenkoeffizienten größer als ein Schwellwert sind. Als Beispiel wurde die zweidimensionale Kreiskonvektion eines Skalars mit nichtglatter Anfangsbedingung gerechnet. Erwartungsgemäß wurde für das adaptive Verfahren eine etwa doppelt so große Konvergenzrate beobachtet, wie für das nichtadaptive Verfahren. Jedoch waren beim adaptiven Verfahren die unphysikalischen Oszillationen um einen Faktor von etwa vier größer. Möglicherweise hängt das auch mit dem Verlust an Konsistenzordnung bei der adaptiven Finite-Differenzen-Diskretisierung mit dem AFD2-Schema und dem damit verbundenen springenden Konsistenzfehler zusammen. Der konvektive Term wurde explizit behandelt, so daß die Zeitschrittweite durch eine CFL-Bedingung beschränkt war. Dies ist gerade für adaptive Ortsdiskretisierungen sehr unangenehm. Hier sind in Zukunft modernere Methoden, wie etwa lokale Zeitschrittverfahren oder eine implizite Behandlung des konvektiven Terms, auszutesten.

Anhand der Simulation von zwei- bzw. dreidimensionalen turbulenten Scherschichten wurde gezeigt, daß die in dieser Arbeit vorgestellten Lösungsverfahren auch auf komplizierte Probleme angewendet werden können. Für die Validierung wurden hier statistische Größen, wie die Energie, die Enstrophie oder Profile der mittleren Geschwindigkeit, betrachtet. Es wurde eine gute Übereinstimmung mit Referenzlösungen bzw. statistischen Theorien erreicht.

Alles in allem haben sich die Multiskalen-basierte Finite-Differenzen-Diskretisierung sowie das biorthogonale Petrov-Galerkin-Schema als sehr vielversprechende Ansätze erwiesen.

Entsprechend unserer bisherigen Erfahrungen mit den in dieser Arbeit beschriebenen Verfahren bieten sich die folgenden Schwerpunkte für künftige Weiterentwicklungen an:

- Damit auch größere Probleme angegangen werden können, muß das Verfahren für Rechner mit verteiltem Speicher parallelisiert werden. Eine Möglichkeit wäre hier, die Technik der Raum-füllenden Kurven für die dynamische Lastbalancierung [Zum01a].
- Um verstärkt Probleme mit praktischer Relevanz lösen zu können, müssen Gebietszerlegungsverfahren und Diskretisierungen für partielle Differentialgleichungen auf transformierten, gekrümmten Gebieten entwickelt werden. Dabei kann auf die Erfahrungen aus [DP96, Sch98a] aufgebaut werden.
- Bisher hatten wir für die Diskretisierung vor allem die Kombination aus Interpolets und Finite-Differenzen-Schemen betrachtet. Überlegenswert ist, ob nicht auf die gleiche Weise wie bisher, die Lifting-Wavelets mit den Finite-Differenzen-Schemen kombiniert werden sollten. Entsprechend unserer Untersuchungen für den Lifting-Vorkonditionierer wären die elliptischen Gleichungssysteme nach Diagonal-Skalierung unmittelbar gut konditioniert und es sollte auch möglich sein, verbesserte Konvergenzaussagen zu erhalten.

Neben diesen algorithmischen Schwerpunkten gibt es natürlich auch auf theoretischer Seite noch viel zu tun. Wichtig erscheinen hier vor allem Aussagen zur Stabilität der diskreten Operatoren, verbesserte Konsistenzabschätzungen sowie eine Fundierung der Verfeinerungsstrategien für elliptische und hyperbolische Differentialgleichungen.

Anhang A

Beweise und Nebenrechnungen

A.1 Beweis für Theorem 2.1

- (i) Ein Beweis für diese Aussage findet sich z.B. in [Chu94].
- (ii) Wie bereits erwähnt sind die Wavelets $\psi_{(l,t)}$ Dilate/Translate einer Mutterfunktion ψ . Es reicht demnach aus, $\int_{\mathbb{R}} \psi(x)x^i dx$ zu betrachten. Die Funktion ψ erfüllt die Skalierungsgleichung

$$\psi(x) = \sum_{s \in \mathbb{Z}} g_s \phi(2x - s) .$$

Daraus folgt

$$\int_{\mathbb{R}} \psi(x)x^i dx = \frac{1}{2^{i+1}} \sum_{l=0}^i \binom{i}{l} \int_{\mathbb{R}} \phi(y)y^{i-l} dy \sum_{s \in \mathbb{Z}} g_s s^l . \quad (1.1)$$

Das Symbol $g(z)$ hat genau dann eine M -fache Nullstelle in $z = -1$, falls

$$0 = \frac{d^{(i)}g}{dz} \Big|_{z=-1} = \sum_{s \in \mathbb{Z}} g_s s(s-1)\dots(s-i+1) .$$

Mit der Darstellung (1.1) folgt nun ganz einfach die die Behauptung.

- (iii) Wir beweisen nur $\tilde{M} = N$. Wegen (2.4) gilt

$$\tilde{g}(z) = \frac{1}{2} \sum_{s \in \mathbb{Z}} (-1)^{1-s} h_{1-s} z^s = zh\left(\frac{-1}{z}\right)$$

Nach der Voraussetzung an h hat $\tilde{g}(z)$ die Faktorisierung

$$\tilde{g}(z) = z^{1-N} \left(\frac{z-1}{2}\right)^N L\left(\frac{-1}{z}\right) ,$$

woraus folgt, daß $\tilde{g}(z)$ in $z = 1$ eine Nullstelle N -ten Grades hat.

- (iv) Unter den angegebenen Voraussetzungen kann gezeigt werden ([Dau92] Lemma 7.1.2), daß die Fouriertransformierte $\hat{\phi}$ von ϕ für beliebiges $\epsilon > 0$ der Abschätzung

$$|\hat{\phi}(\zeta)| \leq C(1 + |\zeta|)^{-N+K+\epsilon}$$

genügt. Daraus folgt dann die Abschätzung für θ^* , siehe [Dau92] Lemma 7.1.1 oder [DD89] Lemma 7.1. Weiterhin folgt

$$|\hat{\phi}(\zeta)|^2(1 + |\zeta|^2)^{N-K-\frac{1}{2}-\frac{3}{2}\epsilon} \leq |\hat{\phi}(\zeta)|^2(1 + |\zeta|)^{2N-2K-1-3\epsilon} \leq C(1 + |\zeta|)^{-1-\epsilon} .$$

So daß $\phi \in \mathcal{H}^{N-K-\frac{1}{2}-\frac{3}{2}\epsilon}$. Da $\epsilon > 0$ beliebig war, folgt die Behauptung.

A.2 Beweis für Korollar 2.1

(i) Wir benutzen die Darstellung $\tilde{G}^l = \left(I_{S(l)} - \tilde{H}^l H^{lT} \right) G^l$ aus Theorem 2.5. Die ungeraden Zeilen von \tilde{H}^l sind alle Null. Weiterhin hat $G_{\cdot,t}^l$ nur in der Zeile $2t + 1$ einen nichtverschwindenden Eintrag. Daraus folgt die Behauptung.

(ii) Wegen Theorem 2.5 gilt $\tilde{H}^{lT} H^l = I_{S(l-1)}$, so daß aus der speziellen Form von \tilde{H}^l folgt

$$H_{2s,\cdot}^l = e_{2s}^T \cdot H^l = \left(e_s^T \cdot \tilde{H}^{lT} \right) H^l = e_s^T .$$

(iii) Es gilt

$$e_t^T \cdot \tilde{G}^{lT} = e_t^T \cdot G^{lT} - e_t^T \cdot G^{lT} H^l \tilde{H}^{lT} .$$

Da aber $e_t^T \cdot G^{lT} = e_{2t+1}^T$ und wegen der speziellen Form von \tilde{H}^l ist $e_t^T \cdot \tilde{G}^{lT} =$

$$e_{2t+1}^T - \left(H_{2t+1,0}^l, 0, H_{2t+1,1}^l, 0, H_{2t+1,2}^l, 0, \dots, 0, H_{2t+1,2^{l-1}}^l \right)$$

Aus dieser Darstellung der t -ten Zeile von \tilde{G}^{lT} folgt die Behauptung. \square

A.3 Beweis für Theorem 3.1

Die Richtung (3.5) \Rightarrow (3.6) ist klar, da $x^{-1}(\Omega_{\mathcal{T}})$ per Definition alle hierarchischen Väter von $(l, t) \in \mathcal{T}$ enthält. Für die andere Richtung nehmen wir an, daß \mathcal{T} die Bedingung (3.6) erfüllt und daß die Indexmengen $\mathcal{S}(l)$ mittels Algorithmus (\mathcal{S}) berechnet wurden. Wir zeigen zunächst $\mathcal{T} \subseteq x^{-1}(\Omega_{\mathcal{T}})$. Sei $(l, t) \in \mathcal{T}$, $l > l_0$. Da $\tilde{G}_{2t+1,t}^l \neq 0$ ist, ist $2t + 1 \in R(\tilde{G}^{lT}, t) \subseteq \mathcal{S}(l)$. Also ist $x(l, t) \in \Omega_{\mathcal{T}}$.

Der Beweis für $x^{-1}(\Omega_{\mathcal{T}}) \subseteq \mathcal{T}$ erfolgt durch Induktion über das Level l . Sei $(L, t) \in x^{-1}(\Omega_{\mathcal{T}})$. D.h. $x(L, t) = (2t+1)2^{-L}$. Da die Spalten von \tilde{G}^L jeweils nur einen Eintrag in einer ungeraden Zeile haben und L das maximale Level ist, muß $(2t+1) \in \mathcal{S}(L)$ sein. Deshalb ist $t \in \mathcal{T}(L)$. Angenommen, wir haben nun für alle $(k, r) \in x^{-1}(\Omega_{\mathcal{T}})$ mit $k > l > l_0$ gezeigt, daß $(k, r) \in \mathcal{T}$. Wir wollen nun für $(l, t) \in x^{-1}(\Omega_{\mathcal{T}})$ zeigen daß $(l, t) \in \mathcal{T}$. Definiere

$$l' := \max\{k : x(l, t) \in 2^{-k}\mathcal{S}(k)\} \quad \text{und} \quad s' := 2^{l'} x(l, t) \in \mathcal{S}(l') . \quad (1.2)$$

Es sind zwei Fälle möglich

- (1) s' ist wegen Bedingung (3.2) in $\mathcal{S}(l')$, so daß es ein $t' \in \mathcal{T}(l')$ gibt mit $s' \in R(\tilde{G}^{l'T}, t')$. Dann ist aber wegen der Kegel-Bedingung $(l, t) \in \mathcal{T}$.
- (2) s' ist wegen Bedingung (3.4) in $\mathcal{S}(l')$, so daß es ein $s'' \in \mathcal{S}^{l'+1}$ gibt mit $s' \in R(H^{l'}, s'')$. Angenommen, s'' wäre gerade: Da die Matrix $H^{l'+1}$ in der Zeile $H_{s'',\cdot}^{l'+1}$ nur einen nicht verschwindenden Eintrag hat ($H_{s'',s''/2}^{l'+1}$) muß $s' = s''/2$ sein. Dann gilt aber

$$x(l, t) = s'2^{-l'} = (2s')2^{-(l'+1)} \in 2^{-(l'+1)}\mathcal{S}(l'+1) ,$$

was ein Widerspruch zur Definition (1.2) von l' und s' ist. s'' kann also nur ungerade sein, so daß $s'' = 2t'' + 1$. Nach Induktionsvoraussetzung muß aber $(l' + 1, t'') \in \mathcal{T}$ sein, da $s'' \in \mathcal{S}(l' + 1)$, $x(l' + 1, t'') = s''2^{-(l'+1)}$ und weiterhin $l' + 1 > l$. Es gilt dann aber

$$2s' \in 2R(H^{l'+1}, s'' = 2t'' + 1) \subset R(\tilde{G}^{l'+1T}, t'') \subseteq \mathcal{S}(l' + 1) .$$

Auch dies ist ein Widerspruch zu (1.2), da wiederum $x(l, t) \in 2^{-(l'+1)}\mathcal{S}(l' + 1)$. Also ist tatsächlich nur Fall (1) möglich, für den die Behauptung aber gezeigt wurde.

□

A.4 Beweis für Lemma 3.2

Sei \mathcal{T} eine Indexmenge, die die Kegel-Bedingung (3.6) erfüllt, dann erfüllt sie offensichtlich auch (3.8). Sei nun umgekehrt \mathcal{T} eine Indexmenge, die (3.8) erfüllt. Für $(l, t) \in \mathcal{T}$ definieren wir

$$V^n(l, t) := V(l, t) \bigcap \{l - n\} \times T(l - n)$$

als die Menge der hierarchischen Väter auf Level $l - n$. Wir werden nun mittels Induktion über n zeigen, daß für $(l, t) \in \mathcal{T}$ auch

$$V^n(l, t) \subset \mathcal{T} \tag{1.3}$$

gilt. Da (l, t) und n als beliebig angenommen wurden, folgt daraus die Behauptung.

Die reduzierte Kegel-Bedingung besagt, daß (1.3) für $n = 1$ und beliebiges $(l, t) \in \mathcal{T}$ gilt. Betrachte nun den Induktionsschritt $1, \dots, n - 1 \mapsto n$. Seien $(l, t) \in \mathcal{T}$ und $(l - n, s) \in V^n(l, t)$ beliebige Indizes. Wir können annehmen, daß $l - n \geq l_0$, da ansonsten $V^n(l, t) = \emptyset$. Für $l - n = l_0$ ist wegen (3.1) $(l - n, s)$ auf jeden Fall in \mathcal{T} . Wir brauchen also nur $l - n > l_0$ zu betrachten.

Die Idee ist nun, einen Index $(l - 1, t') \in V^1(l, t) \subset \mathcal{T}$ zu finden, für den weiterhin $(l - n, s) \in V^{n-1}(l - 1, t')$ gilt. Nach der Induktionsvoraussetzung für $n - 1$ und $(l - 1, t')$ muß dann auch $(l - n, s) \in \mathcal{T}$ sein.

Definiere

$$s'' := 2^n(2s + 1) ,$$

so daß $x(l - n, s) = s''2^{-l}$. Da $(l - n, s) \in V(l, t)$ ist $s'' \in R(\tilde{G}^{lT}, t)$. Da $n \geq 2$ ist s'' sogar in der Teilmenge $2R(H^l, 2t + 1)$, vergleiche Korollar 2.1 (iii). Wegen Voraussetzung (3.7) und $4 \leq s'' \leq 2^l - 4$ können wir einen weiteren Index s' als $s'' + 2$ oder $s'' - 2$ wählen, so daß

$$s' \in 2R(H^l, 2t + 1) .$$

Da $n \geq 2$ ist s'' durch 4 teilbar und $s'/2$ ist ungerade: $s'/2 = (2t' + 1)$. Weiterhin ist $s'/2 \in \mathcal{S}(l - 1)$ und $t' \in T(l - 1)$, so daß $(l - 1, t') \in V^1(l, t)$. Wir haben also einen geeigneten Kandidaten gefunden.

Wir werden nun $H_{2t'+1, s''/4}^{l-1} \neq 0$ zeigen, woraus $s''/2 \in R(\tilde{G}^{l-1T}, t')$ folgt. Somit wäre dann

$$\left(\frac{s''}{2}\right)2^{-(l-1)} = x(l - n, s) \in 2^{-(l-1)}R\tilde{G}^{l-1T}, \{t'\} ,$$

also $(l - n, s) \in V^{n-1}(l - 1, t')$. Damit wäre der Beweis beendet.

Je nach Wahl von s' ist

$$2t' + 1 = s'/2 = 2\left(\frac{s''}{4}\right) + 1 \quad \text{oder} \quad 2t' + 1 = s'/2 = 2\left(\frac{s''}{4} - 1\right) + 1 ,$$

also $t' = s''/4$ oder $t' + 1 = s''/4$. Falls Skalierungsfunktionen auf dem Intervall ohne Randbedingungen benutzt werden, oder falls $1 \leq t' \leq 2^{l-2} - 2$, dann gilt wegen (3.7)

$$H_{2t'+1, t'}^{l-1} \neq 0 \quad \text{bzw.} \quad H_{2t'+1, t'+1}^{l-1} \neq 0 ,$$

so daß $H_{2t'+1, s''/4}^{l-1} \neq 0$. Falls Skalierungsfunktionen mit homogene Randbedingungen benutzt werden und falls $t' = 0$ oder $t' = 2^{l-2} - 1$ muß nachgeprüft werden, ob $H_{1,1}^{l-1} \neq 0$ bzw. $H_{2^{l-1}-1, 2^{l-2}-1}^{l-1} \neq 0$ ist. In diesen Fällen ist $s''/4 = 1$ oder $2^{l-2} - 1$. Für die in dieser Arbeit konstruierten Interpolets auf dem Intervall ist dies der Fall. Somit ist den Induktionsbeweis abgeschlossen. \square

A.5 Beweis für Lemma 3.3

Trivialerweise ist $R(\tilde{G}^{lT}, \mathcal{T}(l)) \subseteq \mathcal{S}(l)$. Hier gilt es auszuschließen, daß $\mathcal{S}(l)$ darüber hinaus noch Einträge hat. Wir beweisen (3.9) mittels Induktion über l . Für $l = L$ ist die Aussage klar. Für $l < L$ betrachte

$$s \in \mathcal{S}(l) = R(\tilde{G}^{lT}, \mathcal{T}(l)) \cup R(H^{l+1}, \mathcal{S}(l+1)) .$$

Angenommen $s \in \mathcal{S}(l) \setminus R(\tilde{G}^{lT}, \mathcal{T}(l))$, dann gibt es $s' \in \mathcal{S}(l+1)$ mit $s \in R(H^{l+1}, s')$. Nach Induktionsvoraussetzung gibt es $t' \in \mathcal{T}(l+1)$ mit

$$s' \in R(\tilde{G}^{l+1T}, t') \stackrel{\text{Korollar 2.1, (iii)}}{=} 2R(H^{l+1}, 2t' + 1) \cup \{2t' + 1\} .$$

Falls s' ungerade ist, dann ist $s' = 2t' + 1$. Also bedeutet $s \in R(H^{l+1}, 2t' + 1)$, daß $H_{2t'+1, s}^{l+1} \neq 0$. Wegen Bedingung (3.7) ist dann aber auch mindestens einer der benachbarten Werte $H_{2t'+1, s''}^{l+1} \neq 0$, $s'' = s + 1$ oder $s'' = s - 1$. Also $s'' \in \mathcal{S}(l)$. Wegen (3.6) und Korollar 2.1 ist $s \notin R(\tilde{G}^{lT}, \mathcal{T}(l))$ gerade. Somit ist s'' ungerade: $s'' = 2t'' + 1$. Nun ist aber t'' ein hierarchischer Vater von $(l+1, t')$. Demnach ist $t'' \in \mathcal{T}(l)$. Wegen Bedingung ist $s \in R(\tilde{G}^{lT}, t'')$. Dies ist ein Widerspruch zur Annahme $s \in \mathcal{S}(l) \setminus R(\tilde{G}^{lT}, \mathcal{T}(l))$. Analog kann der Fall "s' ist gerade" abgehandelt werden. \square

A.6 Beweis für Lemma 3.4

Zur Vereinfachung der Notation definieren wir

$$\mathcal{V}^l := \text{span}\{\Psi_{\mathcal{T}(k)}\}_{k=l_0}^l \quad \text{bzw.} \quad \bar{\mathcal{V}}^l := \text{span}\{\bar{\Psi}_{\mathcal{T}(k)}\}_{k=l_0}^l .$$

Zunächst wird gezeigt, daß aus (3.14) $\text{span}\Psi_{\mathcal{T}} = \text{span}\bar{\Psi}_{\mathcal{T}}$ folgt. Wir können $u = \Psi_{\mathcal{T}} \cdot \mathbf{u}_{\mathcal{T}}$ auf die folgende Weise zerlegen.

$$u = \underbrace{\left(\Psi_{\mathcal{T}(L)} + \Phi^{L-1} Q_{S(L-1), \mathcal{T}(L)}^L\right)}_{= \bar{\Psi}_{\mathcal{T}(L)}} \cdot \mathbf{u}_{\mathcal{T}(L)} + \underbrace{\left(-\Phi^{L-1} Q_{S(L-1), \mathcal{T}(L)}^L \cdot \mathbf{u}_{\mathcal{T}(L)} + \sum_{l=l_0}^{L-1} \Psi_{\mathcal{T}(l)} \cdot \mathbf{u}_{\mathcal{T}(l)}\right)}_{=: u'} \quad (1.4)$$

Nach Voraussetzung (3.14) ist $u' \in \mathcal{V}^{L-1}$. Man kann somit dieses Zerlegungsschema auf u' anwenden und erhält auf diese Weise rekursiv die Koeffizienten $\bar{\mathbf{u}}_{\mathcal{T}}$ mit $\bar{\Psi}_{\mathcal{T}} \cdot \mathbf{u}_{\mathcal{T}} = u$. Also ist $\text{span} \Psi_{\mathcal{T}} \subseteq \text{span} \bar{\Psi}_{\mathcal{T}}$. Die andere Inklusion folgt mit

$$\begin{aligned} \text{span} \bar{\Psi}_{\mathcal{T}} &= \text{span} \Phi^{l_0} \oplus \bigoplus_{l=l_0+1}^L \text{span} \left(\Psi_{\mathcal{T}(l)} + \Phi^{l-1} \cdot Q_{S(l-1), \mathcal{T}(l)}^l \right) \\ &\subseteq \text{span} \Phi^{l_0} \oplus \left(\bigoplus_{l=l_0+1}^L \text{span} \Psi_{\mathcal{T}(l)} \right) \oplus \underbrace{\bigoplus_{l=l_0+1}^L \text{span} \Phi^{l-1} \cdot Q_{S(l-1), \mathcal{T}(l)}^l}_{\subseteq \mathcal{V}^l} \\ &\subseteq \text{span} \Psi_{\mathcal{T}} . \end{aligned}$$

Damit ist die eine Richtung der Behauptung bewiesen. Die andere Richtung ($\text{span} \Psi_{\mathcal{T}} = \text{span} \bar{\Psi}_{\mathcal{T}} \Rightarrow (3.14)$) wird indirekt bewiesen. Wie wir in (3.11) gesehen hatten, muß für eine beliebige Funktion u mit den Darstellungen $u = \Psi_{\mathcal{T}} \cdot \mathbf{u}_{\mathcal{T}} = \bar{\Psi}_{\mathcal{T}} \cdot \bar{\mathbf{u}}_{\mathcal{T}}$ gelten: $\mathbf{u}_{\mathcal{T}(L)} = \bar{\mathbf{u}}_{\mathcal{T}(L)}$. Ist demnach $\mathbf{u}_{\mathcal{T}(L)} = 0$, dann ist $u \in \mathcal{V}^{L-1}$ aber auch in $\bar{\mathcal{V}}^{L-1}$, da auch $\bar{\mathbf{u}}_{\mathcal{T}(L)}$ gleich 0 ist. Somit folgt aus $\text{span} \Psi_{\mathcal{T}} = \text{span} \bar{\Psi}_{\mathcal{T}}$ auch $\mathcal{V}^{L-1} = \bar{\mathcal{V}}^{L-1}$. Angenommen (3.14) gilt nicht, dann gibt es ein u mit den alternativen Darstellung $u = \Psi_{\mathcal{T}} \cdot \mathbf{u}_{\mathcal{T}} = \bar{\Psi}_{\mathcal{T}} \cdot \bar{\mathbf{u}}_{\mathcal{T}}$, so daß u' aus (1.4) nicht in \mathcal{V}^{L-1} ist. Andererseits ist aber $u' = \sum_{l=l_0}^{L-1} \bar{\Psi}_{\mathcal{T}} \cdot \bar{\mathbf{u}}_{\mathcal{T}} \in \bar{\mathcal{V}}^{L-1}$. Dies ist wegen $\mathcal{V}^{L-1} = \bar{\mathcal{V}}^{L-1}$ ein Widerspruch. Daraus folgt die Behauptung.

A.7 Beweis für Lemma 3.5

Entsprechend unserer Überlegungen, die zu $W(l)$ bzw. Bedingung 1. und 2. geführt hatten, ist für $l > l_0$

$$\text{span} \Phi^l \cdot Q_{S(l), \mathcal{T}(l+1)}^{l+1} \subseteq \text{span} \Phi^{W(l)} \subseteq \text{span} \Psi_{R(\tilde{G}^l, W(l))} \oplus \text{span} \Phi^{R(\tilde{H}^l, W(l))} . \quad (1.5)$$

Da nach Voraussetzung (3.15) $R(\tilde{G}^l, W(l)) \subseteq \mathcal{T}(l)$ ist $\text{span} \Psi_{R(\tilde{G}^l, W(l))} \subseteq \text{span} \Psi_{\mathcal{T}(l)}$. Wir werden gleich zeigen, daß für $l_0 < k \leq L$ auch gilt $R(\tilde{H}^k, W(k)) \subseteq W(k-1)$. Daraus folgt dann $\text{span} \Phi^{W(l)} \subseteq \text{span} \Psi_{\mathcal{T}(l)} \oplus \text{span} \Phi^{W(l-1)}$, so daß mittels Rekursion

$$\text{span} \Phi^{W(l)} \subseteq \text{span} \Psi_{\mathcal{T}(l)} \oplus \dots \oplus \text{span} \Psi_{\mathcal{T}(l_0)}$$

folgt. Zusammen mit (1.5) ergibt dies die Behauptung. Nun zu $R(\tilde{H}^k, W(k)) \subseteq W(k-1)$. Aus dem Besetzungsmuster der Matrizen Q^l ergibt sich

$$W(k) = R(Q^{(k+1)T}, \mathcal{T}(k+1)) = S(k) \cap \{\mathcal{T}(k+1) \cup (\mathcal{T}(k+1) +^* 1)\} .$$

Wegen des einfachen Besetzungsmusters von \tilde{H}^k ist deshalb

$$R(\tilde{H}^k, W(k)) = \{s : 2s \in S(k) \cap \{\mathcal{T}(k+1) \cup (\mathcal{T}(k+1) +^* 1)\}\} .$$

Diese Menge soll in $W(k-1) = S(k-1) \cap \{\mathcal{T}(k) \cup (\mathcal{T}(k) +^* 1)\}$ enthalten sein. 1. Fall: s sei so, daß $2s \in \mathcal{T}(k+1) \cap S(k)$. Da unter dieser Voraussetzung $2s \leq 2^k - 2$, folgt mit (3.7) $4s + 2 \in R(\tilde{G}^{s+1T}, \mathcal{T}(s+1))$. Also ist $2s + 1 \in S(k)$. Somit ist (k, s) ein hierarchischer Vater von $(k+1, 2s)$. Dieser Index ist aber aus \mathcal{T} , woraus $s \in \mathcal{T}(k) \subset W(k)$ folgt.

Der 2. Fall, $2s \in \mathcal{T}(k+1) + 1$, führt analog zu $s \in \mathcal{T}(k) + 1$. Daraus folgt die Behauptung. \square

A.8 Beweis für Lemma 3.8

Wir betrachten nur den Fall für den linken Rand. Sei $0 \leq t < N/2$. Für hierarchische Wavelets ist $\psi_{(l,t)} = \phi^{(l,2t+1)}$, wobei $2t+1 < N$. Entsprechend des Konstruktionsschemas für die Randskalierungsfunktionen wissen wir, daß $\phi^{(l,2t+1)} \in \text{span } {}^+\Phi^l$ ist. Dabei sind ${}^+\Phi^l$ die Skalierungsfunktionen ohne Randbedingungen. Es gibt also einen Vektor \mathbf{v}^l , so daß $\phi^{(l,2t+1)} = {}^+\Phi^l \cdot \mathbf{v}^l$. Nun sind die Funktionen ${}^+\Phi^l$ aber interpolierend in $2^{-l}\{0, 1, 2, \dots\}$. Andererseits ist nach Konstruktion $\phi^{(l,2t+1)}$ gleich 1 in $2t+1$ und 0 in den Punkten $2^{-l}\{0, 1, 2, \dots\} \setminus \{2t+1\}$. Also ist $v^{(l,s)}$ gleich 0 für $s \neq 0$ und $s \neq 2t+1$. Es reicht also zu zeigen, daß

$${}^+\phi^{(l,0)}, {}^+\phi^{(l,2t+1)} \subset \text{span } {}^+\Psi_{\mathcal{T}^+}$$

ist. Dann wäre $\psi_{(l,t)} \in \text{span } {}^+\Psi_{\mathcal{T}^+}$. Nun ist ${}^+\phi^{(l,2t+1)} = {}^+\psi_{(l,t)}$ mit $t < N/2$, so daß dieser Fall nach Voraussetzung erledigt ist. Analog zur Argumentation für die Lifting-Bedingung folgert man aus ${}^+\Phi^l = {}^+\Phi^{l-1} \cdot {}^+\tilde{H}^{lT} + {}^+\Psi_l \cdot {}^+\tilde{G}^{lT}$, daß ${}^+\phi^{(l,0)} \in \text{span } {}^+\Psi_{\mathcal{T}^+}$, falls

(i) für $s \in R({}^+\tilde{H}^l, 0) = \{0\}$ ist ${}^+\phi^{(l-1,s)} \in \text{span } {}^+\Psi_{\mathcal{T}^+}$

(ii) $R({}^+\tilde{G}^l, 0) \subset \mathcal{T}(l)$.

Es gilt nun aber $R({}^+\tilde{G}^l, 0) \subset \{0, \dots, N/2 - 1\}$, denn für $N/2 \leq r$ ist ${}^+\tilde{G}_{q,r}^l$ nur für $q \in 2r+1 + \{1-N, 3-N, \dots, -1, 0, 1, \dots, N-3, N-1\}$ ungleich 0, vergleiche (2.4) und Theorem 2.4. Jedoch ist $2r+1 + 1 - N \geq 2 > 0$, so daß r nicht in $R({}^+\tilde{G}^l, 0)$ sein kann. D.h. (ii) ist erfüllt. Wenn aber (ii) erfüllt ist, dann folgt mit Induktion auch, daß für $l_0 \leq k < l$ auch ${}^+\phi^{(k,0)} \in \text{span } {}^+\Psi_{\mathcal{T}^+}$ ist. Also Bedingung (ii).

Induktionsanfang: Auch im adaptiven Fall gilt ${}^+\Phi^0 \in \text{span } {}^+\Psi_{\mathcal{T}^+}$.

Induktionsschritt $k-1 \mapsto k$: Nach Induktionsvoraussetzung ist ${}^+\phi^{(k-1,0)} \in \text{span } {}^+\Psi_{\mathcal{T}^+}$. Wegen der Kegelbedingung und der Voraussetzung von Lemma 3.8 ist ${}^+\psi_{(k,s)}$ für $0 \leq s < N/2$ in $\text{span } {}^+\Psi_{\mathcal{T}^+}$. Daraus folgt wie oben (siehe Bedingungen (i) und (ii)), daß auch ${}^+\phi^{(k,0)} \in \text{span } {}^+\Psi_{\mathcal{T}^+}$ ist. Damit ist die Behauptung von Lemma 3.8 bewiesen.

A.9 Beweis für Lemma 3.9

Um die Beweisidee nicht zu stark hinter Formalismen zu verbergen, skizzieren wir nur den allgemeinen Fall, wo x nicht unmittelbar am Rand liegt.

Sei l das maximale Level, so daß $x \in 2^{-l}\mathcal{S}(l)$. D.h. es gibt $s \in \mathcal{S}(l)$ mit $x = 2^{-l}s$ und $2s \notin \mathcal{S}(l+1)$. Dann folgt zunächst einmal, daß weder die links- noch die rechtsseitige Maschenweite kleiner als 2^{-l} ist: Angenommen, z.B. die rechtsseitige Maschenweite wäre $2^{-(l+1)}$, d.h. $x + 2^{-(l+1)} \in \Omega_{\mathcal{T}}$. Dann ist $2s+1 \in \mathcal{S}(l+1)$. Da aber $2s+1$ ungerade ist, muß demnach wegen Korollar 2.1 und der Kegelbedingung $s \in \mathcal{T}(l+1)$ sein. Daraus folgt mit der Annahme 3.7, die für Interpolets erfüllt ist, daß $2s \in \mathcal{S}(l+1)$ sein muß, was ein Widerspruch zur Definition von l ist. Auf die gleiche Weise zeigt man, daß die Maschenweite nicht noch kleiner ist.

Wir wissen jetzt also, daß die Gitterpunkte in der unmittelbaren Nähe von x durch Wavelets auf Level l induziert werden, so daß $s \in R(\tilde{G}^{lT}, t)$ für ein $t \in \mathcal{T}(l)$. In Bild 1.1 ist $R(\tilde{G}^{lT}, t)$ einmal für nicht randangepaßte Interpolets der Ordnung $N = 4$ skizziert (rote und blaue Punkte). Gehört s zu einem der blauen Punkte, dann sind die Nachbarn $s+1$ oder $s+2$ bzw. $s-$ oder $s-2$ auch in $\mathcal{S}(l)$. Die links- bzw. rechtsseitige Maschenweite ist damit 2^{-l} oder 2^{-l+1} . Kritisch ist nur der Fall, daß s zu einem roten Punkt gehört. Wir betrachten als

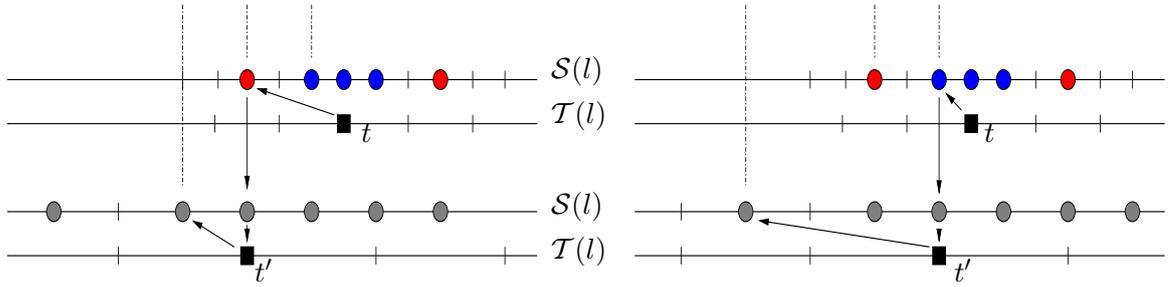


Abbildung 1.1: Index $t \in \mathcal{T}(l)$ induziert zunächst, daß die roten und blauen Indizes in $\mathcal{S}(l)$ liegen müssen. Die Kegelbedingung induziert, daß der Index t' in $\mathcal{T}(l)$ sein muß. Links: der Fall daß t gerade ist. Rechts: t ist ungerade.

Beispiel den linken roten Punkt. In diesem Fall ist $s = 2t - 2$, also gerade. Die rechtsseitige Maschenweite ist 2^{-l+1} . Es gibt nun zwei Möglichkeiten: entweder $s/2 = t - 1$ ist ungerade, dann ist $t' = (t - 2)/2 \in \mathcal{T}(l - 1)$ und wegen Annahme 3.7 ist dann $2t = t - 2$ auch in $\mathcal{S}(l - 1)$, so daß die linksseitige Maschenweite maximal 2^{-l+1} ist, siehe Bild 1.1 (links).

Falls t gerade ist, dann ist wegen der Kegelbedingung $t' = (t - 1)/2 \in \mathcal{T}(l - 1)$. Für Interpolets der Ordnung $N \geq 4$ ist dann auch $2t' - 2 = t - 3 \in \mathcal{S}(l - 1)$, so daß die linksseitige Maschenweite maximal 22^{-l+1} ist, siehe Bild 1.1 (rechts).

Die obige Argumentation geht mit den entsprechenden Modifikation auch für die Fälle durch, daß x in der Nähe des Randes liegt, oder daß das maximale Level l gleich l_0 ist.

A.10 Beweis für Lemma 4.3

Wir betrachten nur den zweidimensionalen Fall, da die Beweise direkt auf höhere Dimensionen übertragen werden können.

Zur Kegel-Bedingung: Für einen Index (\mathbf{l}, \mathbf{t}) und $i = 1, 2$ bezeichnen wir mit $V^i(\mathbf{l}, \mathbf{t})$ die Menge der hierarchischen Väter entlang der i -ten Koordinatenrichtung, z.B. ist $V^1((l_1, l_2), (t_1, t_2)) = \{((k, l_2), (s, t_2)) : (k, s) \in V(l_1, t_1)\}$. Auf Grund dieser Konstruktion ist dann $(\mathbf{k}, \mathbf{s}) \in V^2(V^1(\mathbf{l}, \mathbf{t}))$ gleichbedeutend mit $(k_1, s_1) \in V(l_1, t_1)$ und $(k_2, s_2) \in V(l_2, t_2)$. Das ist dann aber auch gleichbedeutend mit $(\mathbf{k}, \mathbf{s}) \in V^1(V^2(\mathbf{l}, \mathbf{t}))$, so daß im allgemeinen gilt

$$V^1(V^2(\mathbf{l}, \mathbf{t})) = V^2(V^1(\mathbf{l}, \mathbf{t})) .$$

Sei nun \mathcal{T}^1 das Zwischenergebnis der Anwendung von Fill auf \mathcal{T}^0 bzgl. der ersten Koordinatenrichtung. Dann gilt offensichtlich

$$V^1(\mathcal{T}^1) \subseteq \mathcal{T}^1 .$$

\mathcal{T} sei das Resultat von Fill angewendet auf \mathcal{T}^1 bzgl. der zweiten Koordinatenrichtung. Dann gilt wie oben $V^2(\mathcal{T}) \subseteq \mathcal{T}$. Wir müssen nur noch $V^1(\mathcal{T}) \subseteq \mathcal{T}$ nachrechnen, um die Behauptung zu zeigen. Algorithmus Fill sorgt dafür, daß neben \mathcal{T}^1 auch $V^2(\mathcal{T}^2)$, $V^2(V^2(\mathcal{T}^2))$, $V^2(V^2(V^2(\mathcal{T}^2)))$ u.s.w. in \mathcal{T} enthalten sind. Falls L das maximale Level bezeichnet, dann gilt sogar

$$\mathcal{T} = \bigcup_{i=0}^L \underbrace{V^2 \circ \dots \circ V^2}_{i\text{-mal}}(\mathcal{T}^1) .$$

Also

$$V^1(\mathcal{T}) \subseteq \bigcup_{i=0}^L V^1 \circ V^2 \circ \dots \circ V^2(\mathcal{T}^1) = \bigcup_{i=0}^L V^2 \circ \dots \circ V^2 \circ V^1(\mathcal{T}^1) \subseteq \bigcup_{i=0}^L V^2 \circ \dots \circ V^2(\mathcal{T}^1) = \mathcal{T}.$$

Daraus folgt die Behauptung.

Zur Lifting-Bedingung: Auch hier gibt es so etwas wie hierarchische Väter. Entsprechend Algorithmus LiftingFill sind diese im univariaten Fall als

$$\bar{V}(l, t) := \{(l-1, s) : 2s+1 \in R(\tilde{G}^{lT}, t)\} \cup \{(l-1, s) : s \in R(\tilde{G}^{l-1}, R(Q^l, t))\}$$

definiert. Wie bei der Kegel-Bedingung können wir darauf aufbauend die Väter $\bar{V}^i(\mathbf{l}, \mathbf{t})$ entlang der i -ten Koordinatenrichtung definieren. Auch hier ist $(\mathbf{k}, \mathbf{s}) \in \bar{V}^2(\bar{V}^1(\mathbf{l}, \mathbf{t}))$ gleichbedeutend mit $(k_1, s_1) \in \bar{V}(l_1, t_1)$ und $(k_2, s_2) \in \bar{V}(l_2, t_2)$, so daß $\bar{V}^2(\bar{V}^1(\mathbf{l}, \mathbf{t})) = \bar{V}^1(\bar{V}^2(\mathbf{l}, \mathbf{t}))$. Die restlichen Schritte des Beweises sind analog zum Beweis für die Kegel-Bedingung.

A.11 Sobolev-/Besov-Regularität von $|\mathbf{x}|^\beta$

Wir benutzen für das Kriterium (5.6) orthogonale Daubechies–Wavelets $\{\psi_{(l, \mathbf{t}, \mathbf{e})}\}$ mit N verschwindenden Momenten. Für die einzelnen Waveletkoeffizienten erhält man mit Standardtechniken [DD97] die folgenden Abschätzungen.

1. Fall $l = l_0$ oder $t < 2N$:

$$|\langle \psi_{(l, t)}, u \rangle| \lesssim \|u\|_{\mathcal{L}^2(\text{supp}\psi_{(l, \mathbf{t}, \mathbf{e})})} \sim \left(\int_{r=0}^{2N2^{-l}} r^{2\beta} \cdot r^{d-1} dr \right)^{1/2} \lesssim 2^{-l(\beta+d/2)}.$$

2. Fall $l > l_0$ oder $t > 2N$:

$$\begin{aligned} |\langle \psi_{(l, t)}, u \rangle| &\lesssim \|u - P_N u\|_{\mathcal{L}^2(\text{supp}\psi_{(l, \mathbf{t}, \mathbf{e})})} \\ &\lesssim \left(\int_{\text{supp}\psi_{(l, \mathbf{t}, \mathbf{e})}} 2^{-2lN} d\mathbf{x} \right)^{1/2} \cdot \sup_{\mathbf{k}: |\mathbf{k}|_1=N} \sup_{\mathbf{x} \in \text{supp}\psi_{(l, \mathbf{t}, \mathbf{e})}} |D^{\mathbf{k}} u(\mathbf{x})| \\ &\lesssim 2^{-l(N+d/2)} \cdot (2^{-l} |\mathbf{t}|)^{\beta-N} = 2^{-l(d/2+\beta)} |\mathbf{t}|^{\beta-N}. \end{aligned}$$

Für $u \in \mathcal{H}^s$ ist $\sum_l \sum_{\mathbf{t}} |\langle \psi_{(l, \mathbf{t}, \mathbf{e})}, u \rangle|^2 2^{2sl} < \infty$ nachzuweisen. Wir wählen dazu $N \geq \beta + d/2 - 1$ fest und erhalten

$$\begin{aligned} \sum_l \sum_{\mathbf{t}} |\langle \psi_{(l, \mathbf{t}, \mathbf{e})}, u \rangle|^2 2^{2sl} &\lesssim \sum_{l=l_0}^{\infty} 2^{2ls} \cdot 2^{-l(2\beta+d)} + \sum_{l=l_0+1}^{\infty} 2^{2ls} \cdot 2^{-l(2\beta+d)} \sum_{2N \leq |\mathbf{t}| \leq 2^l} |\mathbf{t}|^{2(\beta-N)} \\ &\sim \sum_{l=l_0}^{\infty} 2^{l(2s-2\beta-d)} + \sum_{l=l_0+1}^{\infty} 2^{l(2s-2\beta-d)} \sum_{r=2N}^{2^l} r^{2(\beta-N)} \cdot r^{d-1} \end{aligned}$$

Nach Voraussetzung an N ist $2\beta - 2N + d < 0$, so daß $\sum_{r=2N}^{2^l} r^{2(\beta-N)} \cdot r^{d-1} \lesssim 1$. Damit $u \in \mathcal{H}^s$ ist, muß also $2s - 2\beta - d < 0$ sein. Also ist u in \mathcal{H}^s für $s < \beta + d/2$.

Für die $\mathcal{B}_{\mathcal{H}^0}^s$ -Regularität muß $\sum_l \sum_{\mathbf{t}} |\langle \psi_{(l, \mathbf{t}, \mathbf{e})}, u \rangle|^\tau < \infty$, $\tau = (s/d + 1/2)^{-1} > 0$, nachgewiesen werden. Wir wählen dazu $N > \beta + s + d/2$ und erhalten analog zu oben

$$\sum_l \sum_{\mathbf{t}} |\langle \psi_{(l, \mathbf{t}, \mathbf{e})}, u \rangle|^\tau \lesssim \sum_{l=l_0}^{\infty} 2^{-l\tau(\beta+d/2)} + \sum_{l=l_0+1}^{\infty} 2^{-l\tau(\beta+d/2)} \sum_{r=2N}^{2^l} r^{\tau(\beta-N)} \cdot r^{d-1}$$

Unter der Voraussetzung an N ist $\tau(\beta - N) + d < 0$, so daß $\sum_{r=2N}^{2^l} r^{\tau(\beta-N)} \cdot r^{d-1} \lesssim 1$. Daraus folgt, daß obige Summe für jedes $\tau > 0$ und $\beta > -d/2$ beschränkt ist. Eine etwas kompliziertere Analyse, bei der auch die in den Abschätzungen auftretenden Konstanten und deren Abhängigkeit von τ berücksichtigt werden, zeigt, daß $\|u\|_{\mathcal{B}_{\mathcal{H}^0}^s}$ durch einen Ausdruck der Form s^s beschränkt werden kann (sehr grobe Abschätzung!).

A.12 Abschätzungen für die Punktsingularität

Zunächst soll der \mathcal{L}^2 -Fehler für den Bestapproximanten bzgl. der durch die Indexmengen Λ_δ definierten Ansatzräume abgeschätzt werden. Für das Quadrat dieses Fehlers sind einfach die Quadrate der Koeffizienten der nicht in Λ_δ enthaltenen Wavelets aufzusummieren.

Für $l < l_\delta$ ist der Fehleranteil gleich 0. Für $l_\delta < l \leq L_\delta$ hat man

$$\sum_{\mathbf{e}, \mathbf{t}: |\mathbf{t}| > R(l)} 2^{-2l(1+\beta)} |\mathbf{t}|^{2(\beta-N)} \sim 2^{-2l(1+\beta)} \sum_{r=R(l)}^{2^l} r^{2(\beta-N)} r$$

O.B.d.A. können wir annehmen, daß $N > \beta + 1$, so daß dann $\sum_{r=R(l)}^{2^l} r^{2(\beta-N)} r \sim R(l)^{2(1+\beta-N)}$. Der Anteil des quadratischen Fehlers für den Bereich $l_\delta \leq l \leq L_\delta$ ist somit

$$\begin{aligned} \sum_{l=l_\delta}^{L_\delta} 2^{-2l(1+\beta)} R(l)^{2(1+\beta-N)} &= \delta^{2\frac{1+\beta-N}{\beta-N}} \sum_{l=l_\delta}^{L_\delta} 2^{2l\frac{1+\beta}{\beta-N}} \quad , \quad \beta - N < 0 \\ &\sim \delta^{2\frac{1+\beta-N}{\beta-N}} \cdot 2^{-2\frac{1}{N+1} \log_2(\delta) \frac{1+\beta}{\beta-N}} \\ &= \delta^{-\frac{2}{\beta-N} [1+\beta-N-\frac{1+\beta}{N+1}]} = \delta^{2\frac{N}{N+1}} \quad . \end{aligned}$$

Für $l > L_\delta$ kann der quadratische Fehleranteil von Level l durch $2^{-2l(1+\beta)}$ abgeschätzt werden, so daß von allen Level $> L_\delta$ zusammen nur ein Anteil von

$$\sum_{l > \frac{1}{1+\beta} \log_2(\delta)} 2^{-2l(1+\beta)} \sim \delta^2$$

stammt. Insgesamt ist das Quadrat des Fehlers durch $\delta^{2\frac{N}{N+1}}$ abschätzbar.

Zur Zahl der Freiheitsgrade: Für alle Level kleiner gleich l_δ zusammen hat man ungefähr $2^{2l_\delta} \sim \delta^{-2/(N+1)}$ Freiheitsgrade. Für $l_\delta < l \leq L_\delta$ ergeben sich je Level etwa $R(l)^2$ Freiheitsgrade, insgesamt also

$$\sum_{l=l_\delta}^{L_\delta} R^2(l) \sim \delta^{-2/(N+1)} \quad .$$

Daraus folgt $|\Lambda_\delta| \sim \delta^{-2/(N+1)}$.

A.13 Beweis für Lemma 6.2

Analog wie im Beweis für Lemma 6.1 kann der Beweis wegen der Biorthogonalität von Ansatz- und Testfunktionen auf den zweidimensionalen Fall reduziert werden, was die Notation erheblich vereinfacht.

Wir definieren

$$\mathcal{T}_1 := \{(k_1, s_1) : \exists(k_2, s_2) \text{ mit } (\mathbf{k}, \mathbf{s}) \in \mathcal{T}\} \text{ sowie } \mathcal{T}_2 := \{(k_2, s_2) : \exists(k_1, s_1) \text{ mit } (\mathbf{k}, \mathbf{s}) \in \mathcal{T}\}$$

Dazu werden die Indexmengen

$$\mathcal{T}^1(k_1, s_1) := \{(k_2, s_2) : (\mathbf{k}, \mathbf{s}) \in \mathcal{T}\} \text{ sowie } \mathcal{T}^2(k_2, s_2) := \{(k_1, s_1) : (\mathbf{k}, \mathbf{s}) \in \mathcal{T}\}$$

definiert, mit denen disjunkte Zerlegungen von \mathcal{T} realisiert werden können:

$$\begin{aligned} \mathcal{T} &= \bigcup_{(k_2, s_2) \in \mathcal{T}_2} \{(\mathbf{k}, \mathbf{s}) : (k_1, s_1) \in \mathcal{T}^2(k_2, s_2)\} \\ &= \bigcup_{(k_1, s_1) \in \mathcal{T}_1} \{(\mathbf{k}, \mathbf{s}) : (k_2, s_2) \in \mathcal{T}^1(k_1, s_1)\} \end{aligned}$$

Entsprechend kann man für beliebiges l_2 die disjunkten Zerlegungen

$$\begin{aligned} \mathcal{T} &= \bigcup_{(k_2, s_2) \in \mathcal{T}_2, k_2 < l_2} \{(\mathbf{k}, \mathbf{s}) : (k_1, s_1) \in \mathcal{T}^2(k_2, s_2)\} \\ &\cup \bigcup_{(k_1, s_1) \in \mathcal{T}_1} \{(\mathbf{k}, \mathbf{s}) : (k_2, s_2) \in \mathcal{T}^1(k_1, s_1), k_2 > l_2\} \end{aligned}$$

definieren. Mit dieser Zerlegung wird

$$\begin{aligned} v_{(\mathbf{1}, \mathbf{t})} &= \sum_{(\mathbf{k}, \mathbf{s}) \in \mathcal{T}} \langle \tilde{\psi}_{(\mathbf{k}, \mathbf{s})}, \partial_x \partial_y u_{(\mathbf{k}, \mathbf{s})} \psi_{(\mathbf{k}, \mathbf{s})} \rangle \\ &= \sum_{(k_2, s_2) \in \mathcal{T}_2, k_2 < l_2} \langle \tilde{\psi}_{(l_2, t_2)}, \partial_y \psi_{(k_2, s_2)} \rangle \underbrace{\sum_{(k_1, s_1) \in \mathcal{T}^2(k_2, s_2)} \langle \tilde{\psi}_{(l_1, t_1)}, \partial_x u_{((k_1, k_2), (s_1, s_2))} \psi_{(k_1, s_1)} \rangle}_{=: w_{(l_1, k_2, t_1, s_2)}^1} \\ &+ \sum_{(k_1, s_1) \in \mathcal{T}_1} \langle \tilde{\psi}_{(l_1, t_1)}, \partial_x \psi_{(k_1, s_1)} \rangle \underbrace{\sum_{(k_2, s_2) \in \mathcal{T}^1(k_1, s_1), k_2 \geq l_2} \langle \tilde{\psi}_{(l_2, t_2)}, \partial_y u_{((k_1, k_2), (s_1, s_2))} \psi_{(k_2, s_2)} \rangle}_{=: w_{(k_1, l_2, s_1, t_2)}^2} \quad (1.6) \end{aligned}$$

Die Terme $w_{(l_1, k_2, t_1, s_2)}^1$ können auf effiziente Weise simultan mit $\text{MV}(\text{APG}, 1, \mathbf{u}_{\mathcal{T}})$ berechnet werden. Allerdings nur für $(l_1, t_1) \in \mathcal{T}^2(k_2, s_2)$. Benötigt werden sie aber für $(l_1, t_1) \in \mathcal{T}^2(l_2, t_2)$ falls $\langle \tilde{\psi}_{(l_2, t_2)}, \partial_y \psi_{(k_2, s_2)} \rangle \neq 0$. Diese Bedingung und $k_2 < l_2$ implizieren aber für allgemeine Indexmengen \mathcal{T} , die die multivariate Kegel-Bedingung erfüllen, daß $\mathcal{T}^2(l_2, t_2) \subset \mathcal{T}^2(k_2, s_2)$, da (k_2, s_2) dann ein hierarchischer Vater von (l_2, t_2) ist. Aus diesem Grund ist die Doppelsumme auf der rechten Seite von (1.6) mit $\text{MV}(\text{APG}, j, \text{MV}(\text{APG}, i, \mathbf{u}_{\mathcal{T}}))$ berechenbar.

Ähnlich wird für die zweite Doppelsumme argumentiert: $w_{(k_1, l_2, s_1, t_2)}^2$ ist für $(l_2, t_2) \in \mathcal{T}^1(k_1, s_1)$ effizient mit $\text{MV}(\text{APG}, 2, \mathbf{u}_{\mathcal{T}})$ berechenbar. Benötigt werden dieser Terme aber für $(l_2, t_2) \in \mathcal{T}^1(l_1, t_1)$.

Sei nun (l_2, t_2) ein beliebiger Index aus $\mathcal{T}^1(l_1, t_1) \setminus \mathcal{T}^1(k_1, s_1)$, dann ist l_2 größer als das maximale Level von $\mathcal{T}^1(k_1, s_1)$ und somit ist für alle $(k_2, s_2) \in \mathcal{T}^1(k_1, s_1)$ die Bedingung $k_2 \geq l_2$ nicht erfüllt. Also ist in diesem Fall $w_{(k_1, l_2, s_1, t_2)}^2 = 0$. Die Werte für die zweite Doppelsumme sind also mit $\text{MV}(\text{APG}, i, \text{MV}(\text{APG}, j, \mathbf{u}_{\mathcal{T}}))$ berechenbar.

A.14 Beweis für Lemma 6.5

Es reicht, den eindimensionalen Fall zu betrachten. Der multivariate Fall folgt dann durch rekursives Anwenden der 1D Argumentation. Da die dualen Skalierungsfunktionen Dirac-Funktionale sind, ist

$$\langle \tilde{\psi}_{(l,t)}, u \rangle = \sum_{s \in S(l)} \tilde{G}_{s,t}^l u(2^l s) .$$

Die Funktionswerte $u(2^{-l}s)$ werden durch Taylor-Entwicklungen um $x_t = 2^{-l}t$ bis zur Ordnung $k := \min([\alpha]_-, N) - 1$ ersetzt:

$$\langle \tilde{\psi}_{(l,t)}, u \rangle = \sum_{s \in S(l)} \tilde{G}_{s,t}^l \left[\sum_{r=0}^k \frac{D^r u(x_t)}{r!} (x_s - x_t)^r + \frac{1}{k!} \int_{x_t}^{x_s} (x_s - z)^k D^{k+1} u(z) dz \right] \quad (1.7)$$

Da die dualen Wavelets N verschwindende Momente haben, ist $\sum_{s \in S(l)} \tilde{G}_{s,t}^l s^r = 0$ für $0 \leq r \leq k < N$. Somit ist für beliebiges y auch $\sum_{s \in S(l)} \tilde{G}_{s,t}^l (x_s - y)^r = 0$ für $0 \leq r \leq k$. Daraus folgt, daß der Anteil für den ersten Term in der eckigen Klammer in (1.7) verschwindet. Außerdem ist auch

$$\sum_{s \in S(l)} \tilde{G}_{s,t}^l \int_{x_t}^{x_s} (x_s - z)^r dz = 0 ,$$

so daß

$$\begin{aligned} |\langle \tilde{\psi}_{(l,t)}, u \rangle| &= \left| \frac{1}{k!} \sum_{s \in S(l)} \tilde{G}_{s,t}^l \int_{x_t}^{x_s} (x_s - z)^k (D^{k+1} u(z) - D^{k+1} u(x_t)) dz \right| \\ &\lesssim \sum_{s: \tilde{G}_{s,t}^l \neq 0} \int_{x_t}^{x_s} |x_s - z|^k |z - x_t|^{\beta - (k+1)} \sup_{y, y+h \in [x_s, x_t]} \frac{|\Delta_h D^{k+1} u(y)|}{h^{\beta - (k+1)}} dz \end{aligned}$$

Da aber $\tilde{G}_{s,t}^l$ nur für $x_s \in \text{supp} \psi_{(l,t)}$ ungleich 0 sein kann, d.h. für maximal $2N$ Indizes s , folgt daraus die Behauptung:

$$|\langle \tilde{\psi}_{(l,t)}, u \rangle| \lesssim 2^{-l} 2^{-lk} 2^{-l(\beta - k - 1)} \left| u \Big|_{\text{supp} \psi_{(l,t)}} \right|_{\beta} .$$

□

A.15 Beweis für Lemma 6.6

Offensichtlich ist für $\text{supp} \tilde{\eta}_{(l,t)} \cap \text{supp} \eta_{(k,s)} = \emptyset$ der Operatorkoeffizient gleich 0. Wir brauchen also nur die anderen Fälle zu betrachten. Speziell für die Interpolets und ihre dualen Funktionen ist die Aussage eine einfache Folgerung aus Lemma 6.5. Für beliebiges $r \leq \theta < \theta^*$ ist

$$\begin{aligned} |\langle \tilde{\psi}_{(l,t)}, D^r \psi_{(k,s)} \rangle| &\lesssim 2^{-l(\theta-r)} |D^r \psi_{(k,s)}|_{\theta-r} = 2^{-l(\theta-r)} |\psi_{(k,s)}|_{\theta} = 2^{-l(\theta-r)} |\phi(2^k \cdot - (2s+1))|_{\theta} \\ &\lesssim 2^{-l(\theta-r)} 2^{k\theta} |\phi|_{\theta} \\ &\lesssim 2^{lr - \theta(l-k)} \end{aligned}$$

Für $l \geq k$ folgt sofort die Behauptung, für $l < k$ wird der letzte Ausdruck minimal, falls $\theta = r$. In diesem Fall ist $2^{lr - r(l-k)} = 2^{rk}$.

Im allgemeineren Fall benutzen wir die Argumentation in [Dah97], welche im wesentlichen auf verallgemeinerten Jackson/Bernstein-Ungleichungen beruht. Wir bezeichnen mit $a_{lt,ks} = |\langle \tilde{\eta}_{(l,t)}, D^r \eta_{(k,s)} \rangle|$ den Operatorkoeffizienten und mit M bzw. \tilde{M} die Anzahl der verschwindenden Momente von $\eta_{(k,s)}$ bzw. $\tilde{\eta}_{(k,s)}$. Dann haben wir die (Bernstein-)Ungleichungen $\|\eta_{(k,s)}\|_{\mathcal{H}^q} \lesssim 2^{kq} \|\eta_{(k,s)}\|_{\mathcal{H}^0}$ für $-M < q < \gamma^*$ bzw. $\|\tilde{\eta}_{(k,r)}\|_{\mathcal{H}^q} \lesssim 2^{kq} \|\tilde{\eta}_{(k,s)}\|_{\mathcal{H}^0}$ für $-\tilde{M} < q < \tilde{\gamma}^*$, siehe Theoreme 6.1.3 und 6.1.4 in [Sch98b]. Damit erhalten wir für beliebiges q mit $-\tilde{M} < q \leq \tilde{\gamma}$ und $-M < r - q \leq \gamma$ die Abschätzung

$$a_{lt,ks} \lesssim \|\tilde{\eta}_{(l,t)}\|_{\mathcal{H}^q} \|\eta_{(k,s)}\|_{\mathcal{H}^{r-q}} \lesssim 2^{lq} 2^{k(r-q)} 2^{(l-k)/2} .$$

Im Fall $l \geq k$ können wir $q = r - \gamma^* - \epsilon$ wählen und im Fall $l < k$ $q = \tilde{\gamma} - \epsilon$. Daraus folgt dann die Behauptung.

A.16 Beweis für Theorem 7.1

(ii) Aus der Stabilitätsbedingung (3.7), der Poincare-Friedrichs-Ungleichung [Bra92] und den Normäquivalenzen aus Theorem 2.3 folgt

$$\begin{aligned} \mathbf{u}^T \cdot T_h^T B T_h \cdot \mathbf{u} &\sim \mathbf{u}^T \cdot T_h^T \langle \partial_x 2^{L/2} \Phi^{S(L)}, \partial_x 2^{L/2} \Phi^{S(L)} \rangle T_h \cdot \mathbf{u} \\ &= \left| \sum_{l=l_0}^L (2^{l/2} \Psi_l) \cdot \mathbf{u}_l \right|_{\mathcal{H}^1}^2 \sim \left\| \sum_{l=l_0}^L (2^{l/2} \Psi_l) \cdot \mathbf{u}_l \right\|_{\mathcal{H}^1}^2 \sim \mathbf{u}^T \cdot D \cdot \mathbf{u} . \end{aligned}$$

Der Beweis ist für T_l analog.

(iii) Falls $\gamma^* > 2$ ist, folgt (7.8) wie oben, da dann sowohl die hierarchischen als auch die Lifting-Wavelets \mathcal{H}^2 -stabil sind. Die in den Abschätzungen enthaltenen Konstanten $\underline{b}_j(L)$ und $\bar{b}_j(L)$ sind unabhängig von L nach unten bzw. oben beschränkt. Für den Fall $\gamma^* \leq 2$ benötigt man den erweiterten Bereich der einseitigen Normabschätzungen aus Theorem 2.2. Wir beginnen mit der rechten Ungleichung in (7.8) und T_h . Die Beweise sind für T_l identisch. Sei \mathbf{u}_l der Teilvektor aus \mathbf{u} für Level l . R_l ist der entsprechende Restriktionsoperator, so daß

$$\mathbf{u}_l = R_l \mathbf{u} \quad \text{und umgekehrt} \quad \mathbf{u} = \sum_{l=l_0}^L R_l^T \mathbf{u}_l .$$

Es reicht zu zeigen, daß für beliebiges $\epsilon > 0$ gilt:

$$\begin{aligned} \mathbf{u}^T \cdot D^{-1} T_h^T B^T B T_h D^{-1} \cdot \mathbf{u} &= \sum_{l,k=l_0}^L (4^{-k} R_k^T \mathbf{u}_k)^T \cdot T_h^T B^T B T_h \cdot (4^{-l} R_l^T \mathbf{u}_l) \\ &\lesssim 4^{(2-\gamma^*+\epsilon)L} \sum_{l=l_0}^L \|\mathbf{u}_l\|_{l_2}^2 . \end{aligned} \tag{1.8}$$

Aus Gründen einer einfacheren Notation setzen wir $\mathbf{v}^l := 4^{-l} T_h R_l^T \cdot \mathbf{u}_l$. Interpolets der Ordnung 4, hier mit $\Phi_4^{S(L)}$ bezeichnet, sind in \mathcal{H}^2 , so daß mit der Stabilitätsbedingung (3.7) $B^T B \sim \langle \partial_{xx} 2^{L/2} \Phi_4^{S(L)}, \partial_{xx} 2^{L/2} \Phi_4^{S(L)} \rangle$ gilt. Daraus folgt

$$\mathbf{v}^{lT} \cdot B^T B \cdot \mathbf{v}^l \lesssim \|2^{L/2} \Phi_4^{S(L)} \cdot \mathbf{v}^l\|_{\mathcal{H}^2}^2$$

$$\begin{aligned}
&\lesssim 4^{(2-\gamma^*+\epsilon)L} \|2^{L/2} \Phi_4^{S(L)} \cdot \mathbf{v}^l\|_{\mathcal{H}^{\gamma^*-\epsilon}}^2 \\
&\lesssim 4^{(2-\gamma^*+\epsilon)L} \|2^{L/2} \Phi^{S(L)} \cdot \mathbf{v}^l\|_{\mathcal{H}^{\gamma^*-\epsilon}}^2 \\
&= 4^{(2-\gamma^*+\epsilon)L} 4^{-2l} \|2^{l/2} \Psi_l \cdot \mathbf{u}_l\|_{\mathcal{H}^{\gamma^*-\epsilon}}^2 \\
&\lesssim 4^{(2-\gamma^*+\epsilon)(L-l)} \|\mathbf{u}_l\|_{l_2}^2 .
\end{aligned} \tag{1.9}$$

Für die Abschätzung (1.9) haben wir die $\mathcal{H}^{\gamma^*-\epsilon}$ -Stabilität der biorthogonalen Projektoren $\Phi^{S(L)} \cdot \langle \tilde{\Phi}^{S(L)}, \cdot \rangle$ sowie $\Phi^{S(L)} \cdot \langle \tilde{\Phi}^{S(L)}, \Phi_4^{S(L)} \rangle = \Phi^{S(L)}$ genutzt. Diese Stabilität ist eine Konsequenz aus den Jackson- und Bernstein-Ungleichungen. Somit folgt

$$\begin{aligned}
\sum_{l,k} \mathbf{v}_l^T \cdot B^T B \cdot \mathbf{v}_k &\leq \sum_{l,k} \|B \cdot \mathbf{v}_l\|_{l_2} \|B \cdot \mathbf{v}_k\|_{l_2} \\
&\lesssim \sum_{l,k} 2^{(2-\gamma^*+\epsilon)(2L-|l-k|)} \|\mathbf{u}_l\|_{l_2} \|\mathbf{u}_k\|_{l_2} \\
&\lesssim 4^{(2-\gamma+\epsilon)L} \sum_{l=l_0}^L \|\mathbf{u}_l\|_{l_2}^2 ,
\end{aligned}$$

also die Behauptung (1.8).

Sei $c := L$ falls $N = 2$ und $c := 1$ falls $N > 2$. Für die Abschätzung nach unten ist nach der Transformation $\mathbf{v} := T_h \cdot \mathbf{u}$,

$$\mathbf{v}^T \cdot T_h^{-T} D^2 T_h^{-1} \cdot \mathbf{v} \lesssim c \mathbf{v}^T \cdot B^T B \cdot \mathbf{v} , \quad \text{für alle } \mathbf{v} \text{ mit } \mathbf{1}^T \cdot \mathbf{v} = 0 \tag{1.10}$$

zu zeigen. Definiert man $\mathbf{w} := B \cdot \mathbf{v}$, dann ist wegen $\mathbf{1}^T \cdot \mathbf{v} = 0$ \mathbf{v} der kleinste (in der l_2 -Norm) aller Vektoren \mathbf{q} mit $\mathbf{w} = B \cdot \mathbf{q}$. Somit ist $\mathbf{v} = B^+ \cdot \mathbf{w}$, wobei B^+ die Pseudoinverse der singulären Matrix B ist. In unserem Fall ist

$$B^+ = U^T \begin{pmatrix} 0 & & & \\ & \lambda_1^{-1} & & \\ & & \ddots & \\ & & & \lambda_{2^{L-1}}^{-1} \end{pmatrix} U \quad \text{wobei} \quad B = U^T \begin{pmatrix} 0 & & & \\ & \lambda_1 & & \\ & & \ddots & \\ & & & \lambda_{2^{L-1}} \end{pmatrix} U . \tag{1.11}$$

Also ist (1.10) äquivalent zu

$$\mathbf{w}^T \cdot (B^+ T_h^{-T} D) (D T_h^{-1} B^+) \cdot \mathbf{w} \lesssim c \mathbf{w}^T \cdot \mathbf{w} , \quad \forall \mathbf{w} .$$

D.h. es ist der maximale Eigenwert von $Q^T Q$, mit $Q = D T_h^{-1} B^+$, abzuschätzen und dieser soll $\lesssim c$ sein. Der maximale Eigenwert von $Q^T Q$ ist gleich dem maximalen Eigenwert von $Q Q^T$, [Hac91] Satz 2.4.6, so daß

$$\begin{aligned}
&\mathbf{w}^T \cdot D T_h^{-1} B^+ B^+ T_h^{-T} D \cdot \mathbf{w} \lesssim c \mathbf{w}^T \cdot \mathbf{w} \\
\Leftrightarrow &\mathbf{q}^T \cdot T_h^{-1} (B^T B)^+ T_h^{-T} \cdot \mathbf{q} \lesssim c \mathbf{q}^T \cdot D^{-2} \cdot \mathbf{q}
\end{aligned} \tag{1.12}$$

zu verifizieren ist.

Wegen (2.9) ist $T_h^{-T} = \tilde{T}_h$, wobei \tilde{T}_h die Transformationsmatrix \tilde{T}_h für die dualen Wavelets ist, die durch $2^{-L/2} \tilde{\Phi}^{S(l)} \cdot \tilde{T}_h = (2^{-l_0/2} \tilde{\Psi}_{l_0}, \dots, 2^{-L/2} \tilde{\Psi}_L)$ definiert ist, denn

$$\begin{aligned}
I &= \langle (2^{-l_0/2} \tilde{\Psi}_{l_0}, \dots, 2^{-L/2} \tilde{\Psi}_L), (2^{l_0/2} \Psi_{l_0}, \dots, 2^{L/2} \Psi_L) \rangle \\
&= \tilde{T}_h^T \cdot \langle 2^{-L/2} \tilde{\Phi}^{S(L)}, 2^{L/2} \Phi^{S(L)} \rangle T_h = \tilde{T}_h^T T_h .
\end{aligned}$$

Die Beweisidee für (1.12) ist deshalb, die einseitigen Normabschätzung analog Theorem 2.2 für die dualen Wavelets und \mathcal{H}^{-2} zu nutzen. Als Vorbereitung betrachten wir für einen beliebigen Vektor \mathbf{p} die Dual-Norm $\|\tilde{\Phi}^{S(L)} \cdot \mathbf{p}\|_{\mathcal{H}^{-2}}^2 =$

$$\begin{aligned} \sup_{r \in \mathcal{H}^2} \frac{\langle \tilde{\Phi}^{S(L)} \cdot \mathbf{p}, r \rangle^2}{\|r\|_{\mathcal{H}^2}^2} &\geq \sup_{\mathbf{r}: \mathbf{1}^T \mathbf{r} = 0} \frac{\langle \tilde{\Phi}^{S(L)} \cdot \mathbf{p}, \Phi_4^{S(L)} \cdot \mathbf{r} \rangle^2}{\|\Phi_4^{S(L)} \cdot \mathbf{r}\|_{\mathcal{H}^2}^2} \\ &\gtrsim 2^L \sup_{\mathbf{r}: \mathbf{1}^T \mathbf{r} = 0} \frac{(\mathbf{p}^T \cdot K \cdot \mathbf{r})^2}{\mathbf{r}^T \cdot B^T B \cdot \mathbf{r}}, \quad K = \langle \tilde{\Phi}^{S(L)}, \Phi_4^{S(L)} \rangle \\ &= 2^L \sup_{\mathbf{s}} \frac{(\mathbf{p}^T \cdot K B^+ \cdot \mathbf{s})^2}{\mathbf{s}^T \cdot \mathbf{s}} = 2^L \frac{(\mathbf{p}^T \cdot K B^+ B^+ K^T \cdot \mathbf{p})^2}{\mathbf{p}^T \cdot K B^+ B^+ K^T \cdot \mathbf{p}} \\ &= 2^L \mathbf{p}^T \cdot K (B^T B)^+ K^T \cdot \mathbf{p} \end{aligned}$$

Für die Interpolets ist K natürlich gleich I , für die Lifting-Wavelets ist $K = \langle \tilde{\Phi}^{S(L)}, \Phi_4^{S(L)} \rangle$ zu setzen. Diese Matrix ist zirkulant und hat somit die gleichen Eigenvektoren wie B oder B^+ . Die Eigenwerte sind jeweils ~ 1 . Außerdem ist $\mathbf{1}$ ein Eigenvektor von K , so daß für Vektoren \mathbf{p} mit $\mathbf{1}^T \cdot \mathbf{p} = 0$

$$2^{-L} \|\tilde{\Phi}^{S(L)} \cdot \mathbf{p}\|_{\mathcal{H}^{-2}} \gtrsim \mathbf{p}^T \cdot K (B^T B)^+ K^T \cdot \mathbf{p} \sim \mathbf{p}^T \cdot (B^T B)^+ \cdot \mathbf{p}$$

gilt. Daraus folgt für \mathbf{q} mit $\mathbf{1}^T T_h^{-T} \cdot \mathbf{q} = 0$

$$\begin{aligned} c \mathbf{q}^T \cdot D^{-2} \cdot \mathbf{q} &\gtrsim 2^{-L} \|2^{L/2} \sum_{l=l_0}^L (2^{-l/2} \tilde{\Psi}_l) \cdot \mathbf{q}_l\|_{\mathcal{H}^{-2}}^2 = 2^{-L} \|\tilde{\Phi}^{S(L)} \cdot \tilde{T} \cdot \mathbf{q}\|_{\mathcal{H}^{-2}}^2 \\ &\gtrsim \mathbf{q} \tilde{T}_h^T \cdot (B^T B)^+ \tilde{T}_h \cdot \mathbf{q}, \end{aligned}$$

also die Behauptung (1.12).

(i) Für die Lifting-Wavelets folgt (i) ganz einfach analog zu (ii) aus den Normäquivalenzen. Für die hierarchischen Wavelets ergibt Theorem 2.2 mit $M = 0$

$$\mathbf{u}^T \cdot T_h^T T_h \cdot \mathbf{u} \sim \mathbf{u}^T \cdot T_h^T \langle \Phi^{S(l)}, \Phi^{S(l)} \rangle T_h \cdot \mathbf{u} = \left\| \sum_{l=l_0}^L (2^{l/2} \Psi_l) \cdot \mathbf{u}_l \right\|_{\mathcal{L}^2} \lesssim L \mathbf{u}^T \cdot \mathbf{u}.$$

Für die andere Seite der Ungleichung ist zu zeigen, daß

$$\mathbf{u}^T \cdot \tilde{T}_h^T \tilde{T}_h \cdot \mathbf{u} \lesssim 2^{(1+\epsilon)L} \mathbf{u}^T \cdot \mathbf{u}.$$

Dazu kann wie im Beweis für die erste Ungleichung in (iii) für den Fall $\gamma^* \leq 2$ vorgegangen werden.

(iv) Ist eine Folgerung aus (i) und (iii):

$$\begin{aligned} \mathbf{u}^T \cdot (T_j^{-1} B T_j)^T T_j^{-1} B T_h \cdot \mathbf{u} &= \mathbf{u}^T \cdot (B T_j)^T (T_j T_j^T)^{-1} (B T_j) \cdot \mathbf{u} \stackrel{(i)}{\lesssim} \frac{1}{\underline{a}_j(L)} \mathbf{u}^T \cdot (B T_j)^T (B T_j) \cdot \mathbf{u} \\ &\lesssim \frac{\bar{b}_j(L)}{\underline{a}_j(L)} \mathbf{u}^T \cdot D^2 \cdot \mathbf{u}. \end{aligned}$$

Analog zeigt man die Abschätzung nach unten. \square

A.17 Beweis für Lemma 8.1

Wir zeigen $\|u^k - u^{k+1}\|_A^2 \geq \frac{c}{2\bar{c}} \|u - u^k\|_A^2$. Mit der Galerkin-Orthogonalität $\|u - u^k\|_A^2 = \|u - u^{k+1}\|_A^2 + \|u^{k+1} - u^k\|_A^2$ folgt dann daraus die Behauptung.

$$\begin{aligned}
\|u^k - u^{k+1}\|_A^2 &= \sup_{v \in \mathcal{V}^{k+1}} \frac{\langle v, A(u^k - u^{k+1}) \rangle^2}{\langle v, Av \rangle} \quad (\text{Symmetrie von } A \text{ erlaubt diese Darstellung}) \\
&= \sup_{v \in \mathcal{V}^{k+1}} \frac{\langle v, Au^k - Au + Au - Au^{k+1} \rangle^2}{\langle v, Av \rangle} \\
&= \sup_{v \in \mathcal{V}^{k+1}} \frac{\langle v, Au^k - Au \rangle^2}{\langle v, Av \rangle} \quad (\text{Galerkin-Orthogonalität}) \\
&\geq \frac{1}{\bar{c}} \sup_{v \in \mathcal{V}^{k+1}} \frac{\langle v, -r^k \rangle^2}{\|v\|_{\mathcal{H}^{r/2}}^2} \\
&= \frac{1}{\bar{c}} \sup_{v \in \mathcal{V}^{k+1}} \frac{\langle v, -I_{\Lambda^{k+1}}^* r^k \rangle^2}{\|v\|_{\mathcal{H}^{r/2}}^2} \quad (\text{da } v \in \mathcal{V}^{k+1}) \\
&= \frac{1}{\bar{c}} \|I_{\Lambda^{k+1}}^* r^k\|_{\mathcal{H}^{-r/2}}^2 \\
&\geq \frac{1}{2\bar{c}} \|r^k\|_{\mathcal{H}^{-r/2}}^2 \quad (\text{Voraussetzung}) \\
&= \frac{1}{2\bar{c}} \sup_{w \in \mathcal{H}^{r/2}} \frac{\langle w, Au - Au^k \rangle^2}{\|w\|_{\mathcal{H}^{r/2}}^2} \\
&\geq \frac{c}{2\bar{c}} \sup_{w \in \mathcal{H}^{r/2}} \frac{\langle w, Au - Au^k \rangle^2}{\langle w, Aw \rangle} \\
&= \frac{c}{2\bar{c}} \langle u - u^k, Au - Au^k \rangle \quad (\text{siehe oben})
\end{aligned}$$

Mit der Argumentation aus den letzten 4 Zeilen folgt sofort, daß

$$\|r^k\|_{H^{-r/2}}^2 \geq \frac{c}{2\bar{c}} \|u - u^k\|_A^2 .$$

Analog kann man zeigen, daß

$$\|r^k\|_{H^{-r/2}}^2 \leq \frac{1}{c} \|u - u^k\|_A^2 .$$

Anhang B

Notation

(Räume) Für ein Gebiet Ω bezeichnet $\mathcal{L}^2(\Omega)$ den Hilbert-Raum der Quadrat-integrierbaren Funktionen mit Skalarprodukt $\langle u, v \rangle = \int_{\Omega} uv dx$. $\mathcal{H}^s(\Omega)$, $s \in \mathbb{R}$, ist der entsprechende Sobolev-Raum. Weiterhin werden wir die Hölder-Räume $C^s(\Omega)$, $s > 0$, sowie die Lipschitz-Räume $\mathcal{C}^s(\Omega)$, $s > 0$, [Tri78] benötigen. Bei der Notation für Funktionenräume wird im allgemeinen auf die Angabe des Gebietes, auf dem die Funktionen definiert sein sollen, verzichtet. In dieser Arbeit kommen dafür ohnehin nur \mathbb{R} , $[0, 1]$ oder $[0, 1]^d$ in Betracht. Welches der Gebiete gemeint ist, sollte aus dem jeweiligen Kontext eindeutig hervorgehen. Der Raum l_2 ist der Raum der Folgen $\mathbf{u} = \{u_\lambda\}$ mit $\|\mathbf{u}\|_{l_2} = (\sum_{\lambda} |u_\lambda|^2)^{1/2} < \infty$. Für endliche Vektoren \mathbf{u} ist $\|\mathbf{u}\|_{l_2} = (\mathbf{u}^T \cdot \mathbf{u})^{1/2}$.

(Dirac) Das Symbol $\delta(\cdot)$ bezeichnet das Dirac-Funktional und $\delta_{..}$ ist das Kronecker-Symbol.

(Relationen) $X \lesssim Y$ bedeutet, daß es eine Konstante $c > 0$ mit $X \leq cY$ gibt, die unabhängig ist von den Termen X und Y bzw. den verschiedenen Parametern von denen X und Y abhängen können. $X \sim Y$ bedeutet $X \lesssim Y$ und $Y \lesssim X$.

Speziell für quadratische $n \times n$ -Matrizen A, B bedeutet $A \lesssim B$, daß es $c > 0$ gibt, so daß für alle $\mathbf{u} \in \mathbb{R}^n$ gilt $\mathbf{u}^T \cdot A \cdot \mathbf{u} \leq c \mathbf{u}^T \cdot B \cdot \mathbf{u}$. Entsprechend ist $A \sim B$ definiert.

(Multiindizes und Vektoren) Multiindizes und Vektoren werden fettgedruckt geschrieben, z.B. $\mathbf{l} = (l_1, \dots, l_d)$. Funktionen wie $|\cdot|$ oder $[\cdot]$, die zunächst nur für Elemente aus \mathbb{R} definiert sind, werden komponentenweise auf Vektoren angewandt. Ist z.B. $\mathbf{x} = (x_1, \dots, x_d)$, dann ist $|\mathbf{x}| = (|x_1|, \dots, |x_d|)$. Die Relation $\mathbf{x} > \mathbf{y}$ bedeutet, daß für alle Komponenten $x_i > y_i$ gelten soll. Analog ist $\mathbf{x} \geq \mathbf{y}$ definiert. Ist $c \in \mathbb{R}$, dann bedeutet $\mathbf{x} > c$, daß für alle Komponenten $x_i > c$ gilt. Wir schreiben $\mathbf{x} \neq 0$, falls mindestens eine Komponente von \mathbf{x} ungleich null ist. Mit $\mathbf{x}^T \cdot \mathbf{y} = \sum_i x_i y_i$ wird das übliche Skalarprodukt bezeichnet. Spezielle Vektoren sind $\mathbf{1} := (1, \dots, 1)$ und $\mathbf{0} := (0, \dots, 0)$.

(Matrizen) Etwas ausführlicher soll an dieser Stelle auf die verwendete Matrix-Notation eingegangen werden. Entgegen der üblichen Konvention, bei der die Indexmengen von 0 oder 1 bis $n \in \mathbb{N}$ laufen, wollen wir beliebige abzählbare Mengen $\Sigma, \Gamma \subset \mathbb{Z}$ als Indexmengen zulassen. Eine $\Sigma \times \Gamma$ -Matrix A wäre dann beispielsweise durch

$$A = (a_{s,t})_{s \in \Sigma, t \in \Gamma}$$

definiert. Der Vorteil dieser Notation ist, daß wir oft mit Teilmatrizen anderer Matrizen um-

gehen müssen und wir somit die natürlichen Indexmengen für diese Teilmatrizen beibehalten können. Seien zum Beispiel $\Sigma' \subset \Sigma$ und $\Gamma' \subset \Gamma$, dann ist $A_{\Sigma', \Gamma'}$ die Teilmatrix

$$A_{\Sigma', \Gamma'} = (a_{s,t})_{s \in \Sigma', t \in \Gamma'} .$$

Die unschöne und fehlerträchtige Umrechnung der Indexmengen zu $\{1, \dots, m\} \times \{1, \dots, n\}$ entfällt. Die Menge aller $\Sigma \times \Gamma$ -Matrizen wird mit $M(\Sigma, \Gamma)$ bezeichnet. Eine spezielle Matrix, die oft gebraucht wird, ist die $\Sigma \times \Sigma$ Einheitsmatrix

$$I_{\Sigma} := (\delta_{s,t})_{s,t \in \Sigma} .$$

Ein weiterer Typ von Matrizen entsteht durch Tensorprodukte (auch Kronecker-Produkt) von Vektoren. Seien z.B. $\Phi = \{\phi^s\}_{s \in \Sigma}$ und $\Psi = \{\psi^t\}_{t \in \Gamma}$ Zeilenvektoren von Funktionen, dann ist $\Phi^T \cdot \Psi$ eine Matrix deren Komponenten Funktionen sind. Für ein beliebiges Argument x ist dann $(\Phi^T \cdot \Psi)(x) \in M(\Sigma, \Gamma)$. Entsprechend definieren wir die $\Sigma \times \Gamma$ Matrix

$$\langle \Phi, \Psi \rangle := (\langle \phi^s, \psi^t \rangle)_{s \in \Sigma, t \in \Gamma} .$$

Die kanonischen Einheitsvektoren werden mit \mathbf{e}_s (Spaltenvektor) bezeichnet. Analog zu unserer obigen Matrix-Notation ist dabei $s \in \mathbb{Z}$ zulässig.

(zirkulante Toeplitz-Matrizen) Zu einem gegebenen Vektor (a_{-R}, \dots, a_R) definieren wir die zirkulante $m \times m$ -Matrix G durch

$$G_{s,t} = \begin{cases} a_{s-t-m} & \text{falls } s-t > m/2 \\ a_{s-t+m} & \text{falls } s-t < -m/2 \\ a_{s-t} & \text{sonst} \end{cases} .$$

Diese Matrix wird mit $G = [a_{-R}, \dots, a_R]$ bezeichnet. Die Größe m wird nicht extra mit angegeben, sie sollte aus dem jeweiligen Kontext eindeutig hervorgehen. Der Vektor (a_{-R}, \dots, a_R) wird als *Generator* der Toeplitz-Matrix G bezeichnet.

Beispiel: $G = [-1, 0, 1]$ ist die Matrix

$$\begin{pmatrix} 0 & 1 & & & -1 \\ -1 & 0 & 1 & & \\ & -1 & 0 & \ddots & \\ & & \ddots & \ddots & 1 \\ 1 & & & -1 & 0 \end{pmatrix} .$$

(Indexmengen) Für Indexmengen, z.B. Σ , benötigen wir an einigen Stellen abgeleitete Indexmengen, wie z.B. $\{2t + 1 : t \in \Sigma\}$ solche Mengen bezeichnen wir einfach als z.B. $2\Sigma + 1$.

Symbole

$\lfloor s \rfloor$	größtes Ganzes einer Zahl s
$\lfloor s \rfloor_-$	$\lfloor s \rfloor_- = \lfloor s \rfloor$ falls $s \notin \mathbb{Z}$, ansonsten $\lfloor s \rfloor_- = s - 1$
$\lfloor \mathbf{s} \rfloor_-$	\mathbf{s} ist ein Multiindex, dann ist $\lfloor (s_1, \dots, s_d) \rfloor_- = (\lfloor s_1 \rfloor_-, \dots, \lfloor s_d \rfloor_-)$
$\alpha^{iso}(\mathcal{F})$	Konvergenzrate der Best- n -Term-Approximation von Funktionen aus \mathcal{F} mit isotropen Wavelets, Kapitel 5
$\alpha^{aniso}(\mathcal{F})$	Konvergenzrate der Best- n -Term-Approximation von Funktionen aus \mathcal{F} mit anisotropen Wavelets, Kapitel 5
β	Exponent für Beispiel-Funktion $ \mathbf{x} ^\beta$ in Kapitel 5
γ^*	maximale Sobolev-Regularität der Skalierungsfunktionen $\phi^{(l,s)}$, Definition 2.1
$\tilde{\gamma}^*$	” der dualen Skalierungsfunktionen $\tilde{\phi}^{(l,s)}$, Definition 2.1
$\tilde{\tilde{\gamma}}^*$	” der dualen Skalierungsfunktionen für die Lifting-Wavelets
δ	Schwellwert für Abschneide-Verfahren
θ^*	maximale Hölder-Regularität der Skalierungsfunktionen $\phi^{(l,s)}$, Definition 2.1
ϕ	Mutterfunktion der primalen Skalierungsfunktionen, (2.1)
$\tilde{\phi}$	” dualen Skalierungsfunktionen, (2.4)
$\phi^{(l,t)}$	Skalierungsfunktion auf Level l , Abschnitte 2.1 und 2.4
Φ^l	Zeilenvektor aller Skalierungsfunktionen $\phi^{(l,s)}$ auf Level l
$\Phi^{\mathcal{S}(l)}$	Zeilenvektor von $\phi^{(l,s)}$, $s \in \mathcal{S}(l)$, analog sind $\Phi^{\mathcal{S}^a(l)}$ oder $\tilde{\Phi}^{\mathcal{S}^b(l)}$ definiert
$\psi_{(l,t)}$	Wavelet auf Level l , (2.2) bzw. (2.25)
$\tilde{\psi}_{(l,t)}$	duales Wavelet auf Level l , (2.4)
$\tilde{\tilde{\psi}}_{(l,t)}$	Lifting-Wavelet auf Level l , (2.27)
Ψ_l	Zeilenvektor aller Wavelets auf Level l
$\Psi_{\mathcal{T}(l)}$	Zeilenvektor von $\psi_{(l,t)}$, $t \in \mathcal{T}(l)$, analog ist $\tilde{\Psi}_{\mathcal{T}(l)}$ oder $\tilde{\tilde{\Psi}}_{\mathcal{T}(l)}$ definiert
c	generische Konstante
\mathbf{C}	Matrix-Darstellung für den Basiswechsel 'hierarchisch \rightarrow lifting', Abschnitt 7.1
$C(.,.)$	$C(\text{Matrix}, \text{Indermenge})$, exakte Komponenten bei adaptivem Matrix-Vektor-Produkt, Kapitel 3
d	Dimension des Berechnungsgebietes $[0, 1]^d$
D^r	r -te Ableitung $\partial^r / \partial x^r$
$D^{\mathbf{r}}$	gemischte Ableitung $\frac{\partial^{r_1}}{\partial x^{r_1}} \dots \frac{\partial^{r_d}}{\partial x^{r_d}}$
\mathbf{e}_i	$= (0, \dots, 0, 1, 0, \dots, 0)$, kanonischer Einheitsvektor
g_s	Skalierungskoeffizient für ψ , (2.2)
\tilde{g}_s	” für $\tilde{\psi}$, (2.4)
G^l	Skalierungsmatrix für Level l , (2.7) bzw. Theorem 2.5
\tilde{G}^l	wie G^l
h_s	Skalierungskoeffizient für ϕ , (2.1)
\tilde{h}_s	” für $\tilde{\phi}$, (2.4)
H^l	wie G^l
\tilde{H}^l	wie G^l

$I_{\mathcal{T}}$	biorthogonaler Projektor $I_{\mathcal{T}}u = \Psi_{\mathcal{T}} \cdot \langle \tilde{\Psi}_{\mathcal{T}}, u \rangle$
k	Verfeinerungslevel oder Iterationszähler
\mathbf{k}	$= (k_1, \dots, k_d)$, Multiindex, wie \mathbf{l}
K	Konsistenzordnung
l	Verfeinerungslevel von $\phi^{(l,s)}$ oder $\psi_{(l,t)}$
\mathbf{l}	$= (l_1, \dots, l_d)$, Multiindex für multivariates Verfeinerungslevel, (4.4)
L	maximales Level der adaptiven Indexmengen \mathcal{T} oder \mathcal{T}
N	Grad der polynomialen Exaktheit der primalen Skalierungsfunktionen, Definition 2.1
M	Zahl der verschwindenden Momente der primalen Wavelets, Definition 2.1 bzw. Theorem 2.1
\overline{M}	Zahl der verschwindenden Momente der Lifting-Wavelets, (2.28)
$R(.,.)$	$R(\text{Matrix}, \text{Indexmenge})$, siehe Kapitel 3
s	oft Ortsindex für Skalierungsfunktionen, aber auch Glattheitsparameter in z.B. \mathcal{H}^s
\mathbf{s}	$= (s_1, \dots, s_d)$
$S(l)$	Indexmenge aller Skalierungsfunktionen auf Level l , siehe Tabelle 2.2
$\mathcal{S}(l)$	$\subseteq S(l)$, für adaptive Algorithmen in Kapitel 3
$\mathcal{S}^a(l)$	$\subseteq S(l)$, für Algorithmus APG in Abschnitt 3.4
$\mathcal{S}^b(l)$	$\subseteq S(l)$, wie $\mathcal{S}^a(l)$
$\mathcal{S}^e(l)$	$\subseteq S(l)$, für Algorithmus AFD2 in Abschnitt 3.6
$\mathcal{S}^f(l)$	$\subseteq S(l)$, für Algorithmus AFD2 in Abschnitt 3.6
t	oft Ortsindex für Wavelets, aber auch Glattheitsparameter in z.B. \mathcal{H}^t
\mathbf{t}	$= (t_1, \dots, t_d)$
$T(l)$	Indexmenge aller Wavelets auf Level l , (2.25)
T	End-Zeit bei Zeit-abhängigen Problemen
T_h	Transformationsmatrix für hierarchische Wavelets im nichtadaptiven Fall, (7.3)
T_l	Transformationsmatrix für Lifting-Wavelets im nichtadaptiven Fall, (7.3)
$\mathcal{T}(l)$	$\subseteq T(l)$, Indexmenge für Level l der adaptiven Indexmenge \mathcal{T}
\mathcal{T}	adaptive Indexmenge im univariaten Fall
$\mathcal{T}(\mathbf{l})$	$\subseteq T(l_1) \otimes \dots \otimes T(l_d)$, Indexmenge für Level l der adaptiven Indexmenge \mathcal{T}
\mathcal{T}	adaptive Indexmenge im multivariaten Fall
$u^{(l,s)}$	Koeffizient für Skalierungsfunktion $\phi^{(l,s)}$, generische Schreibweise, analog z.B. $v^{(l,s)}$
$u_{(l,t)}$	Koeffizient für Wavelet $\psi_{(l,t)}$, generische Schreibweise, analog z.B. $v_{(l,s)}$
$u_{(\mathbf{l},\mathbf{t})}$	wie oben, aber multivariater Fall
\mathbf{u}^l	Spaltenvektor von Koeffizienten für Skalierungsfunktionen auf Level l
$\mathbf{u}^{\mathcal{S}(l)}$	Spaltenvektor von $u^{(l,s)}$, $s \in \mathcal{S}(l)$
\mathbf{u}_l	Spaltenvektor von Koeffizienten für Multiskalenfunktionen auf Level l
$\mathbf{u}_{\mathcal{T}(l)}$	Spaltenvektor von $u_{(l,t)}$, $t \in \mathcal{T}(l)$
$\mathbf{u}_{\mathcal{T}(\mathbf{l})}$	wie oben, aber multivariater Fall
$\mathbf{u}_{\mathcal{T}}$	Spaltenvektor aller Multiskalenkoeffizienten für \mathcal{T}
$\mathbf{u}_{\mathcal{T}}$	wie oben, aber multivariater Fall

Literaturverzeichnis

- [AW85] ARNOLD, D.N. und W. WENDLAND: *On the asymptotic convergence of spline collocation of strongly elliptic equations on curves*. Num. Math., 47:317–341, 1985.
- [BA72] BABUSKA, I. und A.K. AZIZ: *Survey lectures on the mathematical foundations of the finite element method*. In: AZIZ, A.K. (Herausgeber): *The mathematical Foundations of the Finite Element Method*. Academic Press, 1972.
- [BBC⁺94] BARRET, R., M. BERRY, T. CHAN, J. DEMMEL, J. DONATO, J. DONGARRA, V. EIJKHOUT, R. POZO, C. ROMINE und H. VAN DER VORST: *Templates for the solution of linear systems: Building Blocks for Iterative Methods*. SIAM, Philadelphia, 1994.
- [BC89] BERGER, M.J. und P. COLELLA: *Local adaptive mesh refinement for shock hydrodynamics*. J. Comp. Phys., 82:360–367, 1989.
- [BCG89] BELL, J.B., P. COLELLA und H.M. GLAZ: *A second order projection method for the incompressible Navier-Stokes equations*. J. Comp. Phys., 85:257–283, 1989.
- [BCM99] BROWN, D.L., R. CORTEZ und M.L. MINION: *Accurate projection methods for the incompressible Navier-Stokes equations*. Technischer Bericht UCRL-JC-135227, Lawrence Livermore National Laboratory, 1999.
- [BCU00] BERTOLUZZA, S., C. CANUTO und K. URBAN: *On the adaptive computation of integrals of wavelets*. Appl. Numer. Math., 34:13–38, 2000.
- [BD98] BUNGARTZ, H.-J. und T. DORNSEIFER: *Sparse grids: Recent developments for elliptic partial differential equations*. In: HACKBUSCH, W. und G. WITTUM (Herausgeber): *Multigrid Methods V*, LNCSE, Seiten 45–70. Springer, 1998.
- [BEK96] BORNEMANN, F., B. ERDMANN und R. KORNHUBER: *A posteriori error estimates for elliptic problems in two or three space dimensions*. SIAM J. Num. Anal., 33:1188–1204, 1996.
- [Ber96] BERTOLUZZA, S.: *A Wavelet Collocation Method for the Numerical Solution of Partial Differential Equations*. Applied and Computational Harmonic Analysis, 3:1–9, 1996.
- [BJR95] BECKER, R., C. JOHNSON und R. RANNACHER: *Adaptive error control for multigrid finite element methods*. Computing, 55:271–288, 1995.

- [BK97] BEYLKIN, G. und J. M. KEISER: *An Adaptive Pseudo-Wavelet Approach for Solving Nonlinear Partial Differential Equations*. In: DAHMEN, W., A.J. KURDILA und P. OSWALD (Herausgeber): *Multiscale Wavelet Methods for Partial Differential Equations*. Academic Press, 1997.
- [BPX90] BRAMBLE, J., J. PASCIAK und J. XU: *Parallel multilevel preconditioners*. Math. Comp., 31:1–22, 1990.
- [BR78] BABUŠKA, I. und W.C. RHEINBOLDT: *Estimates for Adaptive Finite Element Computations*. SIAM J. Num. Anal., 15:736–754, 1978.
- [BR96] BECKER, R. und R. RANNACHER: *Weighted A Posteriori Error Control in FE Methods*. Technischer Bericht, IWR Universität Heidelberg, 1996.
- [Bra72] BRANDT, A.: *Multi-level adaptive technique for fast numerical solution to boundary value problems*. In: CABANNES, H. und R. TEMAM (Herausgeber): *Proc. of the Third Int. Conf. on Num. Methods in Fluid Mechanics*, Lecture Notes in Physics. Springer-Verlag, 1972.
- [Bra92] BRAESS, D.: *Finite Elemente: Theorie, schnelle Löser und Anwendungen in der Elastizitätstheorie*. Springer-Verlag, 1992.
- [BS93] BANK, R.E. und R.K. SMITH: *A posteriori error estimates based on hierarchical bases*. SIAM J. Num. Anal., 30:921–935, 1993.
- [BS94] BABUŠKA, I. und M. SURI: *The p and $h-p$ versions of the finite element method, basic principles and properties*. SIAM Rev., 36:579–632, 1994.
- [Bun92] BUNGARTZ, H.-J.: *Dünne Gitter und deren Anwendung bei der adaptiven Lösung der dreidimensionalen Poisson-Gleichung*. Doktorarbeit, Technische Universität München, 1992.
- [Bun98] BUNGARTZ, H.-J.: *Finite Elements of Higher Order on Sparse Grids*. Habilitationsschrift, Technische Universität München, 1998.
- [CDD01] COHEN, A., W. DAHMEN und R.A. DEVORE: *Adaptive Wavelet Methods for Elliptic Operator Equations*. Math. Comp., 70:27–75, 2001.
- [CDDD99] COHEN, A., W. DAHMEN, I. DAUBECHIES und R. DEVORE: *Tree Approximation and Optimal Encoding*. Technischer Bericht, RWTH Aachen, 1999.
- [CDF92] COHEN, A., I. DAUBECHIES und J.-C. FEAUVEAU: *Biorthogonal bases of compactly supported wavelets*. Comm. Pure Appl. Math., 45:485–560, 1992.
- [CDH97] COHEN, A., R. DEVORE und R. HOCHMUTH: *Restricted Nonlinear Approximation*. Technischer Bericht, University of South Carolina, 1997.
- [CDP96] CARNICER, J.M., W. DAHMEN und J.M. PENA: *Local Decomposition of Refinable Spaces*. Appl. Comp. Harm. Anal., 3:127–153, 1996.
- [Cho68] CHORIN, A.J.: *Numerical solution of the Navier-Stokes equations*. Math. Comp., 22:745–762, 1968.

- [Cho98] CHORIN, A.J.: *Vorticity and Turbulence*. Applied Mathematical Sciences. Springer-Verlag, 1998.
- [CHQZ88] CANUTO, C., M.Y. HUSSAINI, A. QUARTERONI und T.A. ZHANG: *Spectral Methods in Fluid Dynamics*. SSCP. Springer-Verlag, 1988.
- [Chu94] CHUI, C.K.: *Wavelets and Spline Interpolation*. In: CHUI, C.K. (Herausgeber): *Wavelets : theory, algorithms, and applications*. Academic Press, 1994.
- [CKMP00] COHEN, A., S.M. KABER, S. MÜLLER und M. POSTEL: *Fully Adaptive Multiresolution Schemes for Conservation Laws*. Technischer Bericht, Laboratoire d'Analyse Numerique, Paris VI, 2000.
- [Cor95] CORTEZ, R.: *Impulse-based Particle Methods for Fluid Flow*. PhD thesis, University of California, Berkeley, 1995.
- [Cor96] CORTEZ, R.: *An impulse-based approximation of fluid motion due to boundary forces*. J. Comp. Phys., 123:341–353, 1996.
- [CTU99] CANUTO, C., A. TABACCO und K. URBAN: *The Wavelet Element Method, Part I: Construction and Analysis*. Appl. Comp. Harm. Anal., 6:1–52, 1999.
- [Dah97] DAHMEN, W.: *Wavelet and Multiscale Methods for Operator Equations*. Acta Numerica, 6:55–228, 1997.
- [Dah99a] DAHLKE, S.: *Besov Regularity for the Stokes Problem*. In: HAUSSMANN, W., K. JETTER und M. REIMER (Herausgeber): *Advances in Multivariate Approximation*. Wiley, 1999.
- [Dah99b] DAHMEN, W.: *Wavelet Methods for PDEs – Some Recent Developments*. Technischer Bericht, IGPM, RWTH Aachen, 1999.
- [Dau88] DAUBECHIES, I.: *Orthonormal bases of compactly supported wavelets*. Comm. Pure Appl. Math., 41:909–996, 1988.
- [Dau92] DAUBECHIES, I.: *Ten Lectures on Wavelets*. CBMS-NSF Regional Conf. Series in Appl. Math. 61. SIAM, 1992.
- [Dau93] DAUBECHIES, I.: *Orthonormal bases of compactly supported wavelets II*. SIAM J. Math. Anal., 24:499–519, 1993.
- [DD89] DESLAURIERS, G. und S. DUBUC: *Symmetric Iterative Interpolation Processes*. Constr. Approx., 5:49–68, 1989.
- [DD97] DAHLKE, S. und R.A. DEVORE: *Besov Regularity for Elliptic Boundary Value Problems*. Comm. Partial Differential Equations, 22:1–16, 1997.
- [DDD97] DAHLKE, S., W. DAHMEN und R.A. DEVORE: *Nonlinear Approximation and Adaptive Techniques for Elliptic Operator Equations*. In: DAHMEN, W., A. KURDILA und P. OSWALD (Herausgeber): *Multiscale Wavelet Methods for PDEs*. Academic Press, 1997.

- [DDU01] DAHLKE, S., W. DAHMEN und K. URBAN: *Adaptive Wavelet Methods for Saddle Point Problems – Optimal Convergence Rates*. Technischer Bericht, IGPM, RWTH Aachen, 2001.
- [DeV98] DEVORE, R.A.: *Nonlinear Approximation*. Acta Numerica, 7:51–150, 1998.
- [DF97] DYMNIKOV, V.P. und A.N. FILATOV: *Mathematics of Climate Modeling*. Birkhäuser, 1997.
- [DK92] DAHMEN, W. und A. KUNOTH: *Multilevel preconditioning*. Numer. Math., 63:315–344, 1992.
- [DM93] DAHMEN, W. und C.A. MICHELLI: *Using the refinement equation for evaluating integrals of wavelets*. SIAM J. Num. Analysis, 30:507–537, 1993.
- [Doe96] DOERFLER, W.: *A Convergent Adaptive Algorithm for Poisson’s Equation*. SIAM J. Num. Anal., 33:1106–1124, 1996.
- [Don92] DONOHO, D.: *Interpolating wavelet transforms*. Technischer Bericht, Stanford University, 1992.
- [Don99] DONOHO, D.: *Orthonormal Ridgelets and Linear Singularities*. SIAM J. Math. Anal., 31:1062 – 1099, 1999.
- [DP96] DORNSEIFER, T. und C. PFLAUM: *Discretization of elliptic differential equations on curvilinear bounded domains with sparse grids*. Computing, 56:197–213, 1996.
- [DS98] DAUBECHIES, I. und W. SWELDENS: *Factoring Wavelet Transforms into Lifting Steps*. J. Fourier Anal. Appl., 4(3):245–267, 1998.
- [EL95] E, W. und J.G. LIU: *Projection method I: Convergence and numerical boundary layers*. SIAM J. Num. Anal., 32:1017–1057, 1995.
- [Fab09] FABER, G.: *Über stetige Funktionen*. Math. Annalen, 66:81–94, 1909.
- [FJ90] FRAZIER, M. und B. JAWERTH: *A discrete transform an decomposition of distribution spaces*. J. Func. Anal., 93:34–170, 1990.
- [FKBS99] FARGE, M., N. KEVLAHAN, C. BARDOS und K. SCHNEIDER: *Combining deterministic and statistical approaches to compute two-dimensional turbulent flows with walls*. In: TSINOBER, A. (Herausgeber): *Trends in Mathematics*. Birkhäuser, 1999.
- [FN91] FREUND, R. und N.M. NACHTIGAL: *QMR: a quasi-minimal residual method for non-Hermitian linear systems*. Numer. Math., 60:315–339, 1991.
- [FQ99] FLOATER, M.S. und E. QUAK: *Piecewise Linear Prewavelets on Arbitrary Triangulations*. Numer. Math., 82:221–252, 1999.
- [GB86] GUI, W. und I. BABUSKA: *The h , p and $h - p$ versions of the finite element method in one dimension, parts I-III*. Numer. Math., 40:577–683, 1986.

- [Ger01] GERSTNER, T.: *Multiskalenmethoden zur Kompression und interaktiven Verarbeitung großer Datenmengen*. Doktorarbeit, Universität Bonn, 2001.
- [GGM84] GOUPILLAUD, P., A. GROSSMANN und J. MORLET: *Cycle-Octave and Related Transforms in Seismic Signal Analysis*. *Geoexploration*, 23:85–102, 1984.
- [GK99] GRIEBEL, M. und S. KNAPEK: *Optimized approximation spaces for operator equations*. *Constr. Approximation*, 1999. submitted, also as SFB 256 report 568, Universität Bonn.
- [GO95a] GRIEBEL, M. und P. OSWALD: *On the abstract theory of additive and multiplicative Schwarz algorithms*. *Numer. Math.*, 70:163–180, 1995. also as SFB-Report 342/6/93 A, Institut für Informatik, Technische Universität München, 1993.
- [GO95b] GRIEBEL, M. und P. OSWALD: *Tensor Product Type Subspace Splitting and Multilevel Iterative Methods for Anisotropic Problems*. *Adv. Comput. Math.*, 4:171–206, 1995.
- [GO95c] GRIEBEL, M. und P. OSWALD: *Tensor Product Type Subspace Splitting and Multilevel Iterative Methods for Anisotropic Problems*. *Adv. Comput. Math.*, 4:171–206, 1995. also as SFB-Bericht 342/15/94A, Institut für Informatik, TU München, 1994.
- [GOS99] GRIEBEL, M., P. OSWALD und T. SCHIEKOFER: *Sparse Grids for Boundary Integral Equations*. *Numer. Mathematik*, 83(2):279–312, 1999.
- [Gri72] GRIGORIEFF, R.D.: *Numerik gewöhnlicher Differentialgleichungen 1/2*. Teubner Studienbücher. Teubner Verlag, 1972.
- [Gri98] GRIEBEL, M.: *Adaptive Sparse Grid Multilevel Methods for elliptic PDEs based on finite differences*. *Computing*, 61(2):151–179, 1998.
- [GS87] GRESHO, P.M. und R.L. SANI: *On pressure boundary conditions for the incompressible Navier-Stokes equations*. *Int. J. Numer. Meth. Fluids*, 7:371–394, 1987.
- [GSB96] GROPP, W.D., B. SMITH und P. BJØRSTAD: *Domain decomposition*. Cambridge University Press, 1996.
- [GZ99] GRIEBEL, M. und G. ZUMBUSCH: *Adaptive sparse grids for hyperbolic conservation laws*. In: FEY, M. und R. JELTSCH (Herausgeber): *Hyperbolic Problems, Theory, Numerics, Applications, Volume 1*, ISNM 129, Seiten 411–422. Birkhäuser, 1999.
- [Haa10] HAAR, A.: *Zur Theorie der orthogonalen Funktionen-Systeme*. *Math. Annalen*, 69:331–371, 1910.
- [Hac78] HACKBUSCH, W.: *On the multi-grid method applied to difference equations*. *Computing*, 20:291–306, 1978.
- [Hac91] HACKBUSCH, W.: *Iterative Lösung großer schwachbesetzter Gleichungssysteme*. Studienbücher Mathematik. Teubner Verlag, 1991.

- [Hac96] HACKBUSCH, W.: *Theorie und Numerik elliptischer Differentialgleichungen*. Studienbücher Mathematik. Teubner Verlag, 1996.
- [Har94] HARTEN, A.: *Adaptive multiresolution schemes for shock computations*. J. Comp. Phys., 115:319–338, 1994.
- [Har95] HARTEN, A.: *Multiresolution algorithms for the numerical solution of hyperbolic conservation laws*. Comm. Pure Appl. Math., 70:1305–1342, 1995.
- [HB97] HOWELL, L.H. und J.B. BELL: *An adaptive mesh projection method for viscous incompressible flow*. SIAM J. Sci. Comput., 18:996–1013, 1997.
- [Hem95] HEMKER, P. W.: *Sparse-grid finite-volume multigrid for 3D-problems*. Adv. Comput. Math., 4:83–110, 1995.
- [HH84] HO, C.M. und P. HUERRE: *Perturbed free shear layers*. SIAM J. Sci. Comput. WAnnu. Rev. Fluid Mech., 16:365–, 1984.
- [Hoc99] HOCHMUTH, R.: *Wavelet Bases in Numerical Analysis and Restricted Nonlinear Approximation*. Habilitationsschrift, Freie Universität Berlin, 1999.
- [Hol99] HOLMSTROM, M.: *Solving hyperbolic PDEs using interpolating wavelets*. SIAM J. Sci. Comput., 21:405–420, 1999.
- [HP97] HEMKER, P. und C. PFLAUM: *Approximation of partially ordered sets of regular grids*. Appl. Numer. Math., 25:55–87, 1997.
- [HS99] HEMKER, P. W. und F. SPRENGEL: *On the representation of functions and finite difference operators on adaptive sparse grids*. Technischer Bericht, 1999. CWI-Report MAS-R9933.
- [HW65] HARLOW, F. und J. WELCH: *Numerical calculation of time-dependent viscous incompressible flow of fluid with free surface*. Phys. Fluids, 8:2182–2189, 1965.
- [HW91] HAIRER, E. und G. WANNER: *Solving Ordinary Differential Equations I/II*. SSCM. Springer-Verlag, 1991.
- [JRB95] JOHNSON, C., R. RANNACHER und M. BOMAN: *On transition to turbulence and error control in CFD*. Technischer Bericht, IWR, Universität Heidelberg, 1995.
- [KIO91] KARNIADAKIS, G.E., M. ISRAELI und S.A. ORSZAG: *High-order splitting methods for incompressible Navier-Stokes equations*. J. Comp. Phys., 97:414–, 1991.
- [Kna00] KNAPEK, S.: *Approximation und Kompression mit Tensorprodukt-Multiskalen-Approximationsräumen*. Doktorarbeit, Universität Bonn, Institut für Angewandte Mathematik, 2000.
- [Kos00] KOSTER, F.: *A Proof of the Consistency of the Finite Difference Technique on Sparse Grids*. Computing, 65:247–261, 2000.
- [KS99] KARNIADAKIS, G.E. und S.J. SHERWIN: *Spectral/hp Element Methods for CFD*. Oxford University Press, 1999.

- [Kun01] KUNOTH, A.: *Wavelet Methods - Elliptic Boundary Value Problems and Control Problems*. Advances in Numerical Mathematics. Teubner, 2001.
- [KV92] KOVACEVIC, J. und M. VETTERLI: *Nonseparable multidimensional perfect reconstruction filter banks and wavelet bases for \mathbb{R}^n* . IEEE Trans. Inform. Theory, special issue on wavelets, 38:533–555, 1992.
- [Lax71] LAX, P.: *Hyperbolic Systems of Conservation Laws and the Mathematical Theory of Shock Waves*, Band 11 der Reihe *CBMS-NSF Regional Conference Series*. SIAM, 1971.
- [LCUC96] LIU, W.K., Y. CHEN, R.A. URAS und C.T. CHANG: *Generalized multiple scale reproducing kernel particle methods*. Comp. Meth. Appl. Mech. Eng., 139:91–157, 1996.
- [LeV94] LEVEQUE, R.J.: *The immersed interface method for elliptic equations with discontinuous coefficients and singular sources*. SIAM J. Numer. Anal., 31:1019–1044, 1994.
- [LeV97] LEVEQUE, R.J.: *An immersed interface method for Stokes flow*. SIAM J. Sci. Comput., 18:709–735, 1997.
- [LL99] LI, S. und W.K. LIU: *Reproducing kernel hierarchical partition of unity, part I*. Int. J. numer. Meth. Eng., 45:251–288, 1999.
- [LM86] LEMARIÉ, P.G. und Y. MEYER: *Ondelettes et bases Hilbertiennes*. Rev. Mat. Iberoamer., 2:1–18, 1986.
- [LPR94] LOUIS, A.K., P. MAASS und A. RIEDER: *Wavelets*. Teubner Studienbücher. Teubner Verlag, 1994.
- [Mal89a] MALLAT, S.: *Multiresolution Approximations and Wavelet Orthonormal Bases*. Trans. Amer. Mat. Soc., 315(1):69–87, 1989.
- [Mal89b] MALLAT, S.: *A theory for multiresolution signal decomposition: the wavelet representation*. IEEE Trans. Patt. Anal. and Mach. Intell., 7:674–693, 1989.
- [Mey92] MEYER, Y.: *Wavelets and Operators*. Cambridge University Press, 1992.
- [Mic64] MICHALKE, A.: *On the inviscid instability of the hyperbolic tangent velocity profile*. J. Fluid Mech., 19:543–556, 1964.
- [MP98] MONASSE, P. und V. PERRIER: *Orthonormal wavelet bases adapted for partial differential equations with boundary conditions*. SIAM J. Math. Anal., 29:1040–1065, 1998.
- [Nov76] NOVIKOV, E.A.: *Dynamics and statistics of a system of vortices*. Sov. Phys. JETP, 41,5:937–943, 1976.
- [OID86] ORSZAG, S.A., M. ISRAELI und M.O. DEVILLE: *Boundary conditions for incompressible flows*. J. Sci. Comp., 1:75–, 1986.

- [Ose89] OSELEDETS, V.I.: *On a new way of writing the Navier-Stokes equation: The Hamiltonian formalism*. Commun. Moscow Math. Soc., 44:210–211, 1989.
- [Osw94] OSWALD, P.: *Multilevel Finite Element Approximation. Theory and Applications*. Teubner Skripten zur Numerik. Teubner, Stuttgart, 1994. [Bibl. We6: InvNr. 5161].
- [Osw00] OSWALD, P.: *Best-N-Term capacitance approximation on sparse grids*. In: AL., T. CHAN ET (Herausgeber): *Proc. 11th Intern. Conf. on Domain Decomposition Methods in Science and Engineering*. Domain Decomposition Press, 2000.
- [Pes76] PESKIN, C.S.: *Partial differential equations in biology*. Courant Institute of Mathematical Sciences, 1976.
- [Pfl95] PFLAUM, C.: *Discretization of second order elliptic differential equations on sparse grids*. In: BANDLE, C. (Herausgeber): *Calculus of Variations, Applications and Computations*, Pitman Res. Notes Math. Ser. 326, Seiten 176–187, Harlow, 1995. Longman Scientific & Technical.
- [Pro97] PROHL, A.: *Projection and quasi-compressibility methods for solving the incompressible Navier-Stokes equations*. Advances in Numerical Mathematics. B.G. Teubner, 1997.
- [SB86] STOER, J. und R. BULIRSCH: *Einführung in die Numerische Mathematik II*. Springer Verlag, 1986.
- [SB91] SZABO, B.A. und I. BABUSKA: *Finite Element Analysis*. Wiley, 1991.
- [Sch97] SCHNEIDER, K.: *Wavelets, turbulence and chemical reactions*. Doktorarbeit, Universität Kaiserslautern, 1997.
- [Sch98a] SCHIEKOFER, T.: *Die Methode der Finiten Differenzen auf Dünnen Gittern*. Doktorarbeit, Universität Bonn, 1998.
- [Sch98b] SCHNEIDER, R.: *Multiskalen- und Wavelet-Matrixkompression: Analysisbasierte Methoden zur effizienten Lösung großer vollbesetzter Gleichungssysteme*. Advances in Numerical Mathematics. Teubner Verlag, 1998.
- [Shu99] SHU, C.W.: *High order ENO and WENO schemes for computational fluid dynamics*. In: BARTH, T.J. und H. DECONINCK (Herausgeber): *High-Order Methods for Computational Physics*, Band 9 der Reihe LNCSE. Springer-Verlag, 1999.
- [Spr01] SPRENGEL, F.: *Multilevel Algorithms for Finite-Difference Discretizations on Sparse Grids*. Numer. Algorithms, 26:111–121, 2001.
- [SS95] SCHRÖDER, P. und W. SWELDENS: *Spherical Wavelets: Efficiently Representing Functions on the Sphere*. Computer Graphics Proceedings (SIGGRAPH 95), Seiten 161–172, 1995.
- [Ste96] STEVENSON, R.: *Piecewise linear (pre-)wavelets on non-uniform meshes*. In: HACKBUSCH, W. und G. WITTUM (Herausgeber): *Multigrid Methods V*, LNCSE, Seiten 306–319. Springer-Verlag, 1996.

- [Str97] STRICKWERDA, J.C.: *High-order accurate Schemes for incompressible Flows*. Int. J. Num. Meth. Fluids, 24:715–734, 1997.
- [Swe97] SWELDENS, W.: *The Lifting Scheme: a Construction of Second Generation Wavelets*. SIAM J. Math. Anal., 29,2:511–546, 1997.
- [Tch87] TCHAMITCHIAN, P.H.: *Biorthogonalité et théorie des opérateurs*. Rev. Math. Iberoamer., 3:163–189, 1987.
- [Tem69] TEMAM, R.: *Sur l'approximation de la solution des equation de Navier-Stokes par la méthode des fractionnaires (ii)*. Arch. Rat. Mech. Anal., 33:377–385, 1969.
- [Tem93] TEMLYAKOV, V. N.: *Approximation of Periodic Functions*. Nova Science, New York, 1993.
- [Tow76] TOWNSEND, A.A.: *Structure of Turbulent Shear Flow*. Cambridge University Press, 1976.
- [Tri78] TRIEBEL, H.: *Interpolation Theory, Function Spaces, Differential Operators*. Dt. Verlag der Wissenschaften, 1978.
- [TW85] THOMPSON, J.F. und Z.U. WARSI: *Numerical Grid Generation, Foundations and Applications*. North-Holland, 1985.
- [Urb95] URBAN, K.: *Multiskalenverfahren fr das Stokes-Problem und angepate Wavelet-Basen*. Doktorarbeit, RWTH Aachen, 1995.
- [Van93] VANDEWALLE, S.: *Parallel multigrid waveform relaxation for parabolic problems*. B.G. Teubner, Stuttgart, 1993.
- [Ver96] VERFÜRTH, R.: *A Review of a Posteriori Error Estimation and Adaptive Mesh-Refinement Techniques*. Wiley-Teubner, 1996.
- [VK93] VREUGDENHIL, C.B. und B. KOREN: *Numerical methods for Advection-Diffusion Problems*. In: *Notes on Numerical Fluid Mechanics*. Teubner Verlag, 1993.
- [VW97] VASSILEVSKI, P. und J. WANG: *Stabilizing the Hierarchical Basis by approximate Wavelets I, Theory*. Num. Lin. Algebra Appl., 4:103–126, 1997.
- [WDH97] WILMOTT, P., J. DEWYNNE und S. HOWISON: *Option pricing: Mathematical models and computation*. Oxford Financial Press, 1997.
- [Wet] WETTON, B.: *Error analysis for pressure increment schemes*. SIAM J. Num. Anal. to appear.
- [Yse86] YSERENTANT, H.: *On the multilevel splitting of finite element spaces*. Num. Math., Seiten 379–412, 1986.
- [Zen91] ZENGER, C.: *Sparse Grids*. In: HACKBUSCH, W. (Herausgeber): *Parallel Algorithms for Partial Differential Equations*, Notes on Numerical Fluid Mechanics 31, Seiten 241–251. Vieweg, Braunschweig, 1991.
- [Zha99] ZHANG, F.: *Matrix Theory*. UTX. Springer-Verlag, 1999.

- [ZISC70] ZIENKIEWICZ, O.C., B.M. IRONS, F.E. SCOTT und J.S. CAMPBELL: *High speed computing of elastic structures*. In: *Proc. of the Symposium of Int. Union of Theor. Appl. Mechanics, Liege*, 1970.
- [ZKGB82] ZIENKIEWICZ, O.C., D.W. KELLY, J. GAGO und I. BABUSKA: *Hierarchical finite element approaches, error estimates and adaptive refinement*. In: WHITEMAN, J. (Herausgeber): *Mathematics of Finite Elements and Applications IV*, Seiten 313–346. Academic Press, 1982.
- [Zum01a] ZUMBUSCH, G.: *Adaptive parallel multilevel methods for PDEs*. Habilitationsschrift, Universität Bonn, 2001.
- [Zum01b] ZUMBUSCH, G.: *private communication*, April 2001.