Multiskalen-Verfahren für Konvektions-Diffusions Probleme

Dissertation

zur

Erlangung des Doktorgrades (Dr. rer. nat.)

der

Mathematisch–Naturwissenschaftlichen Fakultät

der

Rheinischen Friedrich–Wilhelms–Universität Bonn

vorgelegt von

Frank Kiefer

aus

Mettlach

Bonn 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: 03.07.2001

Zusammenfassung

In dieser Arbeit werden erstmalig über einen zur Nichtstandardform gehörenden Erzeugendensystem-Ansatz robuste Wavelet-basierte Multiskalen-Löser für allgemeine zweidimensionale stationäre Konvektions-Diffusions Probleme entworfen und praktisch umgesetzt.

Für Multiskalen-Verfahren, die lediglich direkte Unterraumzerlegungen verwenden, ist es im allgemeinen nicht mehr möglich, zugehörige Multiskalen-Glätter zu konstruieren, die im Grenzfall sehr starker Konvektion auf jeder Skala zu einem direkten Löser entarten. Als eine Möglichkeit zur Konstruktion robuster Multiskalen-Methoden bleibt die Wahl der Multiskalen-Zerlegungen selbst. Es ist sicherzustellen, daß man sowohl hinsichtlich der singulären Störung stabile Grobgitterprobleme als auch bezüglich der Maschenweite stabile Unterraumzerlegungen erhält. Gleichzeitig muß der Aspekt der approximativen Gauß-Elimination beachtet werden, der durch das Zusammenspiel matrixabhängiger Prolongationen und Restriktionen mit einer hierarchischen Basis Zerlegung gegeben ist.

Um alle diese Forderungen zu erfüllen, wird zunächst ausgehend von geometrischen Vergröberungen ein allgemeines Petrov–Galerkin Multiskalen-Konzept entwickelt, bei dem die Zerlegungen auf der Ansatz- und Testseite unterschiedlich sind. Es werden matrixabhängige Prolongationen, die von robusten Mehrgitter-Techniken her bekannt sind, verwendet, zusammen mit Wavelet-artigen und hierarchischen Multiskalen-Zerlegungen der Ansatz- und Testräume bezüglich des feinsten Gitters. Die Kernidee bei den vorgeschlagenen Verfahren ist, jeweils einen der Komplementräume auf der Ansatz- oder Testseite hierarchisch zu wählen, um zusammen mit einer problemabhängigen Vergröberung auf der anderen Seite physikalisch sinnvolle Grobgitterdiskretisierungen und gleichzeitig einen approximativen Eliminationseffekt zu erreichen. Die Komplementräume auf der entsprechend anderen Seite werden hingegen Wavelet-artig aufgespannt, was insbesondere zu einer Stabilisierung des Verfahrens bezüglich der Abhängigkeit von der Maschenweite der Diskretisierung führt. Mit den weiterhin entwickelten AMGlet-Zerlegungen, die auf rein algebraischen Prinzipien beruhen, gelingt es, geometrisch orientierte Tensorprodukt-Konstruktionen, die für separable Probleme erfolgreich sind, zu verlassen, um schwierige nichtseparable Aufgaben in unter Umständen kompliziert berandeten Gebieten behandeln zu können. Dies eröffnet darüberhinaus auch den Übergang von Modellproblemen hin zu praxisnahen Fragestellungen.

Unterschiedliche numerische Beispiele zeigen, daß man durch die vorgeschlagenen Konstruktionen zu verallgemeinerten Hierarchische Basis Mehrgitter-Verfahren mit robusten Konvergenzeigenschaften gelangt.

Dank

An dieser Stelle möchte ich zunächst meinen Dank an all diejenigen Personen aussprechen, die mich beim Erstellen dieser Arbeit unterstützt haben. An erster Stelle ist hier Prof. Dr. Michael Griebel zu nennen, der mir zu jeder Zeit mit wertvollen Ratschlägen und Ideen zur Seite stand. Bei Frau Prof. Dr. Angela Kunoth bedanke ich mich für die Übernahme des Zweitgutachtens. Besonderen Dank schulde ich meinem Kollegen Dipl.-Math. Marc Alexander Schweitzer, der mir bei vielen Fragen der Programmierung stets bereitwillig weitergeholfen hat. Ich danke ihm ebenfalls für die Überlassung seines AMG-Codes. Herzlichen Dank ihm sowie auch Dr. Stephan Knapek und Daniel Schittko für die vielen fachlichen Diskussionen und das mühevolle Korrekturlesen. Edgar Kraft danke ich für das Erstellen einiger Graphiken mit

dem Visualisierungsprogramm AVS. Nicht zuletzt möchte ich allen MitarbeiterInnen, Kollegen und Studenten der Abteilung für Wissenschaftliches Rechnen und Numerische Simulation an der Universität Bonn für das gute Arbeitsklima Dank sagen.

Diese Arbeit entstand im Rahmen des Paketprojekts "Mehrgitteralgorithmen für unsymmetrische und heterogene Probleme", das von der Deutschen Forschungsgemeinschaft finanziell gefördert wurde. Darüberhinaus kam mir finanzielle Unterstützung von Seiten des Sonderforschungsbereichs 256 "Nichtlineare Partielle Differentialgleichungen" an der Universität Bonn zu. Dank auch dafür.

Bonn, im April 2001 Frank Kiefer

Inhaltsverzeichnis

Kapitel 1

Einleitung

Die Vorgänge in unserer natürlichen Umwelt sind äußerst komplex und vielfältig. Eine zentrale Aufgabe des *Wissenschaftlichen Rechnens* und der *Numerischen Simulation* ist, diese mit Hilfe des Computers besser zu beschreiben und zu verstehen, als es alleine mit Experimenten und Beobachtungen oder mathematischen Formulierungen möglich ist. Mit diesen beiden noch relativ jungen und zusammengehörenden Disziplinen hat sich neben dem traditionellen experimentellen Zugang und dem theoretischen Ansatz mittlerweile ein vielversprechender dritter Weg zur Beschreibung der Wirklichkeit herausgebildet.

Häufig treten in den zur Numerischen Simulation verwendeten mathematischen Modellen für die Beschreibung von Vorgängen, wie sie in den Natur-, Ingenieur-, Wirtschafts- und Sozialwissenschaften untersucht werden, *konvektive* und *diffusive* Prozesse *gleichzeitig* auf. Ein wichtiges Beispiel ist etwa die Schadstoffausbreitung in einem fließenden Gewässer. Eine Verunreinigung breitet sich durch molekulare Diffusion langsam um ihre Quelle herum aus, die selbst jedoch einem Transport durch das Fließen des Gewässers unterliegt. Durch Konvektion alleine würde sich eine konzentrierte punktförmige Verschmutzung lediglich entlang einer eindimensionalen Bahnkurve an der Oberfläche des Gewässers bewegen. Wegen der zusätzlich vorhandenen Diffusion verbreitert sich diese Kurve nach allen Richtungen hin und nimmt eine keilförmige Gestalt an. Ein dimensionsloser Parameter, der die relative Stärke des konvektiven Anteils gegenüber dem diffusiven Anteil mißt, kann hierbei sehr unterschiedliche und große Werte annehmen. Man spricht auch von einem *singulär gestörten Problem*. Nimmt die Transportgeschwindigkeit des Gewässers im obigen Beispiel betragsmäßig Werte zwischen 0.5 und 3 $\frac{m}{s}$ an und variiert die Diffusion zwischen 0.005 und 5 $\frac{m^2}{s}$, so ergeben sich bezogen auf eine Längenskala von einigen 100 Metern etwa *Pécletzahlen* zwischen 10 und 6000.

Für andere Anwendungen können diese Unterschiede noch dramatischer sein. Konvektions-Diffusions Prozesse und die sie beschreibenden partiellen Differentialgleichungen können hierbei selbstständig oder auch als Teilprobleme noch komplexerer Fragestellungen auftreten. Weitere Beispiele sind die Temperaturverteilung in einer Strömung, atmosphärische Verunreinigungen, Grundwassertransport, Halbleitervorgänge, aber ebenso die Optionspreisbewertung in finanzmathematischen Modellen [86]. Auch die inkompressiblen Navier–Stokes Gleichungen im Fall großer *Reynoldszahlen*, die dabei die Rolle des dimensionslosen Parameters übernehmen, werden häufig als Anlaß zur Untersuchung von Konvektions-Diffusions Gleichungen genannt. Durch ihren nichtlinearen Charakter und die zusätzliche Bedingung der Divergenzfreiheit ist aber nicht klar, ob es gerade die Kopplung von Konvektion und Diffusion ist, die die Hauptschwierigkeiten bei der Lösung der Navier–Stokes Gleichungen verursacht.

Modellhaft betrachten wir in dieser Arbeit das folgende stationäre Konvektions-Diffusions Problem mit homogenen Dirichletschen Randbedingungen auf dem Einheitsquadrat

$$Tu := -\Delta u + \vec{b} \cdot \nabla u = f \text{ in } \Omega :=]0,1[^2 \quad \text{und} \quad u = 0 \text{ auf } \partial\Omega. \tag{1.1}$$

Das Konvektionsfeld $\vec{b} = (b_1(x,y), b_2(x,y))$ ist hierbei im allgemeinen ortsabhängig und die zugehörige $\mathcal{L}^\infty(\Omega)$-Norm kann sehr große Werte annehmen. Je nach Richtung der Konvektion unterteilt man Konvektions-Diffusions Probleme in bestimmte Klassen. Für *achsenorientierte* Probleme gilt $\vec{b} = (b_1(x), 0)$ oder $\vec{b} = (0, b_2(y))$, *separable* Probleme erfüllen $\vec{b} = (b_1(x), b_2(y))$. Aufgaben, für die die letzte Bedingung nicht mehr zutrifft, sind etwa Fälle wirbelbehafteter Konvektion [76].

Zur Numerischen Simulation von Konvektions-Diffusions Vorgängen muß die kontinuierliche Gleichung (1.1) zuerst einmal diskretisiert werden. Dies geschieht zum Beispiel mittels eines geeigneten *Finite-Differenzen* oder *Finite-Elemente* Verfahrens, indem man die Gleichungen etwa für die Punkte (*Freiheitsgrade*) eines endlichen Gitters $\Omega_0 \subset \Omega$ betrachtet [44, 86, 102]. Man gelangt so zu großen schwachbesetzten linearen Gleichungssystemen der Form

$$T_0 u_0 = f_0, \tag{1.2}$$

die den Kern einer möglichen Simulationsschleife darstellen. Der untere Index kennzeichnet hierbei die Auflösungsstufe (*Level*, *Skala*), auf der die diskreten Probleme betrachtet werden. Entgegen der üblichen Notation [62] versehen wir gröbere Auflösungsstufen mit größeren Indizes. Die feinste Skala trägt den Index 0.

Es ist zumeist keine triviale Aufgabe, die entstehenden Gleichungssysteme effizient mit Hilfe eines Computers zu lösen. Die für realistische Simulationen benötigte hohe Auflösung des Raumes mit diskreten Punkten führt zu extrem hohen Anforderungen an Rechenzeit und Speicherplatz. Dies gilt insbesondere, wenn zeitabhängige Vorgänge in drei Raumdimensionen simuliert werden sollen. Die direkte Lösung mittels Gauß-Elimination benötigt für ein allgemeines Gleichungssystem mit n Unbekannten einen Aufwand $\mathcal{C}$, der kubisch mit n steigt, das heißt $\mathcal{C} = \mathcal{O}(n^3)$ [64]. Leider überträgt sich dieses Verhalten auch auf die Lösungszeit. Da die wachsende Rechnerleistung stets von dem Wunsch begleitet wird, noch größere Probleme zu rechnen, stoßen auch geschicktere und weniger aufwendige Implementierungen direkter Löser wie etwa die Band-Gauß-Elimination sofort an Grenzen [40]. Die Lösung der Gleichungssysteme (1.2) erfolgt daher heute zumeist mit Hilfe von Iterationsverfahren, die sich ausgehend von einer Startapproximation schrittweise der tatsächlichen Lösung nähern [64]. Sie haben gegenüber direkten Lösungsverfahren zunächst den Vorteil, daß sie nur einen geringen Rechenaufwand pro Iteration benötigen. Außerdem lassen sie sich ohne großen zusätzlichen Speicheraufwand implementieren. Neben den klassischen Iterationsverfahren, wie beispielsweise dem Jacobi-, dem Gauß–Seidel-Verfahren oder SOR-Varianten, werden heutzutage auch moderne Krylovraum-Methoden, etwa das GMRES- oder das BICGSTAB-Verfahren, zusammen mit einfachen Vorkonditionierern zur Lösung der Systeme (1.2) herangezogen [104]. Man trifft dabei in der Regel aber auf die folgenden beiden Schwierigkeiten. Die Konvergenz der Verfahren

- verschlechtert sich mit abnehmender Maschenweite h_0 des zur Diskretisierung verwendeten Gitters und
- hängt ab von den Koeffizientenfunktionen des Differentialoperators, hier also von der Gestalt und Stärke des Konvektionsfeldes $\vec{b}$.

Diese Problematik bezeichnet man auch als die Frage nach der *Robustheit* der Verfahren [126]. Dabei wird gefordert, daß der Aufwand für einen Schritt des betreffenden Iterationsverfahrens nur linear mit der Anzahl der vorhandenen Freiheitsgrade wächst. Der ersten Schwierigkeit begegnet man im Fall des ungestörten Poissonproblems (1.1) mit $\vec{b} = \mathbf{0}$ durch *Multilevel-Methoden* (*Mehrgitter-Verfahren*, *Multilevel-Vorkonditionierer*) [24, 56, 62]. Sie führen im Idealfall zu Konvergenzraten, die nicht mehr von der Maschenweite h_0 des Gitters abhängen, das zur Diskretisierung benutzt wurde [91]. Man kann sagen, daß die Entwicklung von Multilevel-Methoden für symmetrisch positiv definite elliptische Randwertaufgaben abgesehen von auch hier sich möglicherweise stellenden Robustheitsfragen mit den Arbeiten [57, 127] eine finale Reife erreicht hat. Dies gilt auch aus theoretischer Sicht. Ganz anders verhält es sich im Fall unsymmetrischer Probleme wie der Konvektions-Diffusions Gleichung (1.1). Robuste Multilevel-Verfahren, die also *zusätzlich* die zweite oben genannte Schwierigkeit beheben, sind im allgemeinen nur sehr schwer zu konstruieren, wie die Erfahrung lehrt. Es existieren zwar einige in der Praxis recht erfolgreiche Ansätze [38, 47, 52, 76, 79, 94, 133], wirkliche Robustheitsbeweise liegen bislang aber fast nur für (triviale) eindimensionale Probleme vor. Oft wird in theoretischen Arbeiten der konvektive Anteil als kleine Störung des symmetrischen Teils der Gleichung angesehen, so daß die Aussagen für den interessierenden konvektionsdominierten Fall nicht mehr gültig oder wertlos sind [23, 123]. Ein erstes theoretisches Robustheitsresultat für achsenorientierte Probleme in zwei Raumdimensionen und nichtperiodische Randbedingungen findet man in [99]. Es basiert auf einer *Semi-Vergröberung* und nutzt die Reduktion auf den eindimensionalen Fall.

Vor circa 15 Jahren, also fast ein viertel Jahrhundert nach Vorstellen des ersten Mehrgitter-Ansatzes [45], setzte mit dem systematischen Vorantreiben der Theorie der *Wavelets* und *Multiskalen-Analysen* eine stürmische Entwicklung in den unterschiedlichsten Bereichen der Natur- und Ingenieurwissenschaften sowie der Angewandten und Reinen Mathematik ein [36, 81, 82, 83]. Auf der Suche nach möglichen Anwendungen hatte diese sehr bald auch das Gebiet der Numerik partieller Differentialgleichungen erfaßt. Wavelet-Methoden zur Lösung elliptischer partieller Differentialgleichungen nutzen eine geschickte Multiskalen-artige Zerlegung des Lösungsraums in Unterräume, die von den Dilaten und Translaten einer einzigen Funktion, nämlich dem Wavelet, aufgespannt werden. Mit entsprechenden schnellen Algorithmen können die Lösungen dann in Bestandteile, die zu unterschiedlichen Skalen gehören, aufgespalten werden, was insbesondere nützlich zur Entwicklung von Fehlerschätzern und adaptiven Verfahren ist. Einen guten Überblick bietet der Übersichtsartikel [32]. Man kann Wavelet-Methoden aber ebenfalls zum Lösen der nach Diskretisierung entstehenden linearen Gleichungssysteme einsetzen. Für symmetrisch positiv definite elliptische Randwertprobleme erhält man mit Algorithmen, die den bekannten Multilevel-Methoden nachempfunden sind, *Multiskalen-Verfahren*, die ebenfalls zu gitterweitenunabhängigen Konvergenzraten führen [18, 33, 75, 121, 122]. Wavelet-basierte Multiskalen-Löser, die für Konvektions-Diffusions Probleme robuste Methoden liefern, sind hingegen kaum bekannt. Die theoretischen Resultate in [92, 93] sind aufgrund der dort verwendeten Normen fragwürdig, wie auch Reusken in [99] bemerkt. Das auf einer Semi-Vergröberung basierende Verfahren ist außerdem nur auf den Spezialfall achsenorientierter Konvektion zugeschnitten und wird unserer Meinung nach nicht ausreichend von numerischen Beispielen unterstützt.

Ziel dieser Arbeit ist die Konstruktion robuster Wavelet-basierter Multiskalen-Verfahren zur Lösung großer schwachbesetzter linearer Gleichungssysteme, die bei der numerischen Behandlung des Konvektions-Diffusions Problems (1.1) auftreten. Da wir in unserem Ansatz

entscheidend auf Techniken robuster Mehrgitter-Methoden zurückgreifen, wollen wir im folgenden zunächst kurz die Kernidee von Mehrgitter-Verfahren skizzieren und die wichtigsten Probleme und Lösungsstrategien in Bezug auf Konvektions-Diffusions Gleichungen diskutieren. Sogenannte *Hierarchische Basis Mehrgitter-Verfahren* und ihre problemabhängig gebildeten Verallgemeinerungen dienen uns dann als Ausgangspunkt, um das von uns vorgeschlagene Konzept zu motivieren und zu erklären.

Wir beschränken uns zuerst wieder auf die ungestörte Poissongleichung. Betrachtet man für ein mittels Finite-Differenzen gebildetes zugehöriges lineares Gleichungssystem klassische iterative Löser wie das Gauß–Seidel- oder das Jacobi-Verfahren, so dämpfen diese hochfrequente (rauhe) Anteile des Fehlers zwischen Lösung und Iterierter zwar stark, dafür aber niederfrequente (glatte) Anteile nur schwach. Das führt dazu, daß die Fehler nach einigen wenigen Iterationsschritten im wesentlichen nur noch glatte Anteile besitzen, für deren Dämpfung dann viele weitere Iterationsschritte notwendig sind. In diesem Zusammenhang spricht man von traditionellen Iterationsverfahren auch als *Glättern*. Niederfrequente Fehleranteile lassen sich gut auf einem gröberen Gitter approximieren und man braucht dazu weitaus weniger Freiheitsgrade als zuvor. Die zentrale Idee ist nun, die geglätteten Fehleranteile auf ein gröberes Gitter zu transportieren, um sie dort mittels eines kostengünstigeren Grobgitterproblems schneller reduzieren zu können. Da die zur Korrektur dienenden Grobgitterprobleme die gleiche Struktur aufweisen wie das ursprüngliche Feingitterproblem, läßt sich diese Prozedur rekursiv fortsetzen und man gelangt zu Mehrgitter-Verfahren [25, 45, 61]. Mit ihnen entstanden erstmals Iterationsverfahren linearen Aufwandes, für die die Zahl der notwendigen Iterationsschritte zum Erreichen der Lösung bis auf Diskretisierungsgenauigkeit unabhängig von der Gitterweite h_0 ist. Solche Löser heißen auch *optimal*.

Mehrgitter-Verfahren lassen sich ebenfalls zur Lösung von Gleichungssystemen einsetzen, die mittels einer Finite-Elemente Diskretisierung entstanden sind (*variationelles Mehrgitter*). Die für den Transport zwischen den unterschiedlichen Gittern verantwortlichen *Prolongations-* und *Restriktionsabbildungen* induzieren dabei Basisfunktionen von Unterräumen grobskaligerer Funktionen der endlichdimensionalen Ansatz- und Testräume $\mathcal{V}_0$ und $\mathcal{S}_0$, die in den Punkten gröberer Gitter verankert sind. Variationelle Mehrgitter-Verfahren können darüberhinaus als *multiplikative Unterraumkorrektur-Verfahren* angesehen werden [127] und die zugehörigen *additiven* Varianten stellen gerade die bekannten Multilevel-Vorkonditionierer BPX und MDS von Bramble, Pasciak und Xu respektive Zhang dar [24, 137]. Durch das gleichzeitige Betrachten verschiedener Diskretisierungslevel und das Zusammenfassen der zugehörigen Basisfunktionen erhält man *Erzeugendensysteme*. In [55, 56] wurde gezeigt, daß sich Mehrgitter-Verfahren und Multilevel-Vorkonditionierer auch als traditionelle iterative Methoden (Gauß–Seidel-Verfahren, Jacobi-Vorkonditionierer) über den mittels der Erzeugendensysteme entstehenden semidefiniten Gleichungssysteme interpretieren lassen. Durch Übertragen der Konvergenzbeweise für diese klassischen Verfahren, erhält man mit Hilfe des Erzeugendensystem-Zugangs im Fall symmetrisch positiv definiter Probleme sogar eine regularitätsfreie qualitative Mehrgitter-Konvergenztheorie [57].

Betrachten wir das Konvektions-Diffusions Problem (1.1), so führen mit zunehmender Konvektion konventionelle Mehrgitter-Verfahren zur Lösung der entstehenden Gleichungssysteme zu einer Reihe von Schwierigkeiten. Sie machen es notwendig, Modifikationen der Standardkomponenten klassischer Mehrgitter-Verfahren zu entwickeln. Diese betreffen die verwendeten Grobgitterdiskretisierungen und die Wahl der Prolongationen und Restriktionen. Darüberhinaus werden auch spezielle Glätter eingesetzt.

Berechnet man die Matrizen für die Grobgitterprobleme mit Hilfe des *Galerkin-Produkts* $T_{k+1} = (P_{k+1}^k)^t T_k P_{k+1}^k$, so wird bei Verwenden der (bi-)linearen Interpolation als Prolongation P_{k+1}^k der zugehörige Grobgitterstern durch jeden Galerkin-Vergröberungsschritt dem Stern der gewöhnlichen Finite-Elemente Diskretisierung ähnlicher, der bekanntlich zentrale Differenzen beinhaltet [54, 132]. Zentrale Differenzen Diskretisierungen führen aber für konvektionsdominierte Probleme bei nicht hinreichend feinen Gitterweiten zu oszillierenden und physikalisch unsinnigen Lösungen [60]. Grobgitterdiskretisierungen, die über eine einfache Galerkin-Vergröberung gebildet werden, neigen daher sogar ausgehend von einer stabilen Feingitterdiskretisierung zu Instabilitäten und können dadurch die Konvergenz des Mehrgitter-Verfahrens zerstören. Das eben geschilderte Problem in Bezug auf Grobgitterdiskretisierungen und den Galerkin-Ansatz kann durch eine geeignete matrixabhängige Wahl der Prolongationen im Galerkin-Vergröberungsprozeß vermieden werden. Hierzu gibt es Vorschläge von Dendy und de Zeeuw [38, 132]. Im Zusammenspiel mit ILU-artigen Glättungsverfahren konnte für einfache zweidimensionale Probleme im numerischen Experiment robuste Mehrgitter-Konvergenz erreicht werden. Daneben wurden Grobgitterkorrekturen vorgeschlagen, die nicht mehr auf dem Prinzip einer uniformen Vergröberung (*Standardvergröberung*) basieren. Verläuft die Konvektionsrichtung etwa entlang einer der Koordinaten, dann ist Semi-Vergröberung, die entgegen der Richtung der singulären Störung erfolgt, ein probates Mittel [62, 117]. *Algebraische Mehrgitter-Verfahren* (AMG) bestimmen darüberhinaus problemabhängig geeignete Grobgitterkorrekturen, indem zuerst aus den Einträgen der Feingittermatrix die zu verwendenden Grobgitterpunkte bestimmt und danach zugehörige Prolongationen und Restriktionen konstruiert werden. Sie erlangen so mit einfachen Gauß–Seidel- oder Jacobi-Glättern robuste Konvergenzraten [27, 52, 112]. Beweise liegen dafür aber bislang nicht vor.

Die Schwierigkeiten in Bezug auf den Glättungsprozeß gestalten sich wie folgt. Geschieht die Vergröberung auf klassische Weise, so beobachtet man, daß einfache Glätter wie das Jacobi- oder das Gauß–Seidel-Verfahren bei starker Konvektion im allgemeinen ihre Glättungseigenschaften verlieren [54]. Dann konvergiert das Mehrgitter-Verfahren trotz stabiler Grobgitterdiskretisierungen nicht mehr schnell genug und die Konvergenzrate ist von der Stärke der Konvektion abhängig. Für achsenorientierte Probleme eignet sich ein Linienglätter, das heißt ein Block-Gauß–Seidel-Verfahren, bei dem die Blöcke durch je eine Linie des Gitters in Konvektionsrichtung gebildet sind und in jedem Block eine exakte Invertierung stattfindet [62]. Der Linienglätter kann auch mit abwechselnder Richtung alternierend eingesetzt werden. Für Konvektion in Diagonalrichtung ergibt sich jedoch auch damit leider kein robustes Mehrgitter-Verfahren. Verallgemeinerungen hiervon basieren auf der Idee der *unvollständigen Faktorisierung* (ILU) [4]. ILU- und Block-ILU-Glätter funktionieren im 2D-Fall gut und bewirken weitgehende Robustheit. Ihre Wirkungsweise basiert nicht zuletzt auch auf der Eigenschaft, im Grenzfall $b \to \infty$ zu einem exakten Löser zu entarten. Das gesamte Mehrgitter-Verfahren konvergiert dann in nur einem Schritt. Man spricht auch von einem *robusten Glätter*. Um diesen Grenzfall zu realisieren, geht man zum Beispiel über zu der mittels $\varepsilon = 1/||\vec{b}||_\infty$ skalierten Differentialgleichung (1.1). Im 3D-Fall ist es im allgemeinen aber nicht mehr ohne weiteres möglich, kostengünstige ILU-Glätter mit dieser Eigenschaft anzugeben [69]. Setzt man als Glätter ein Gauß–Seidel-Verfahren ein, dessen Durchlaufreihenfolge mit der Konvektionsrichtung läuft, dann ist dafür unter Umständen ebenfalls das oben erwähnte Kriterium für einen robusten Glätter erfüllt. Diese Methode wird in [68, 79, 94] genauer vorgestellt und es wird diskutiert, wie im zweidimensionalen Fall Anordnungsreihenfolgen gefunden werden können, die Zyklen im Konvektionsfluß vermeiden. Damit werden Block-Gauß–Seidel-Glätter

geschaffen, bei denen die Abarbeitungsreihenfolge der Blöcke weitgehend freigestellt ist, innerhalb der Blöcke jedoch der Konvektionsrichtung auf Stromlinien folgt. In drei Raumdimensionen ist die Suche nach geeigneten Anordnungstechniken Gegenstand aktueller Forschung [17, 76, 79, 94].

Beschränkt man in variationellen Mehrgitter-Zyklen die Glättungsprozeduren für die einzelnen Level auf diejenigen Freihheitsgrade, die durch Gitterpunkte gegeben sind, die nicht auch zum nächstgröberen Gitter gehören, also auf sogenannte *Fein-ohne-Grobgitterpunkte*, so erhält man *Hierarchische Basis Mehrgitter-Verfahren* (HBMG) [12, 16, 129]. Sie ebnen uns den Übergang zur Welt der *Multiskalen-Methoden*. Aus Sicht von Unterraumkorrektur-Verfahren lassen sich die zu Hierarchische Basis Mehrgitter-Verfahren gehörenden Glättungsiterationen als Korrekturen ansehen, die lediglich bezüglich der Komplemente von Räumen gröberer Skalen relativ zu den Räumen nächstfeinerer Auflösungsstufe erfolgen. Es handelt sich daher um Unterraumkorrektur-Methoden, die im Unterschied zu den bisher erwähnten Multilevel-Verfahren auf *direkten Zerlegungen* der zugrundeliegenden Funktionenräume beruhen. Die Komplementräume werden bei Hierarchische Basis Mehrgitter-Verfahren von den Funktionen der oben angesprochenen Erzeugendensysteme aufgespannt, die in den jeweiligen Fein-ohne-Grobgitterpunkten eines Gitters verankert sind [56]. Durch sie kann hochfrequente Detail-Information dargestellt werden, die unter Umständen beim Übergang von einer feineren zur nächstgröbern Skala vernachlässigt werden kann (Grundidee der nichtverlustfreien Datenkompression). Faßt man all die komplementären Basisfunktionen unterschiedlicher Level zusammen, so erhält man Teilmengen der Erzeugendensysteme, die im Unterschied zu diesen tatsächlich Basen von $\mathcal{V}_0$ und $\mathcal{S}_0$ darstellen, sogenannte *hierarchische Basen* [56].

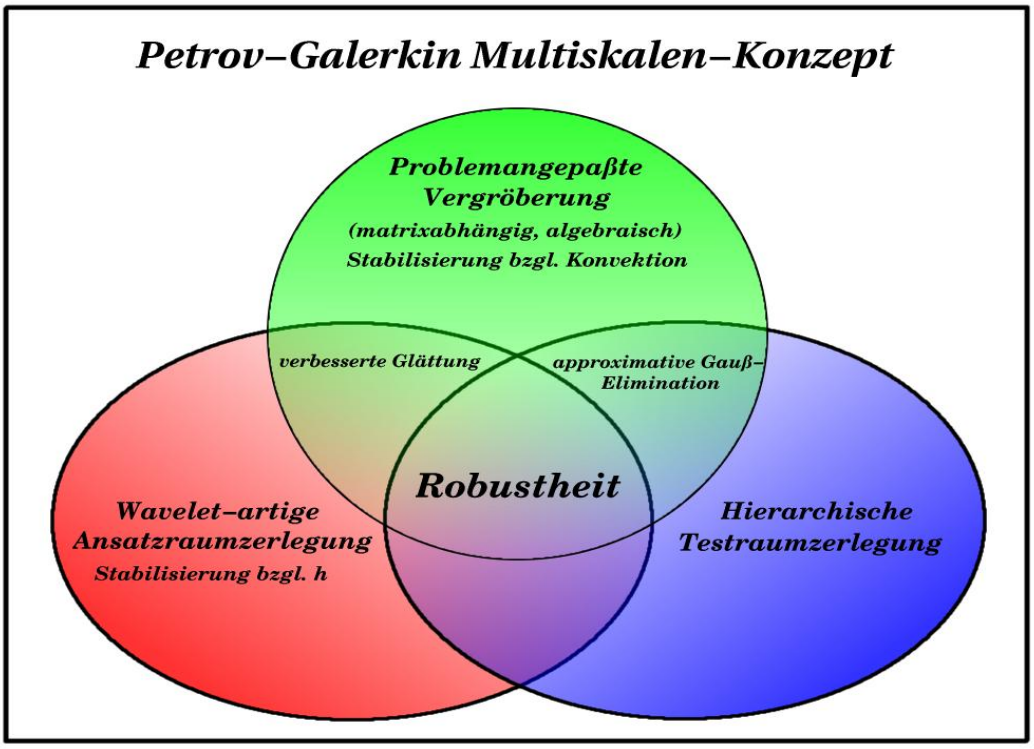
Die Hierarchische Basis Mehrgitter-Methode kann formal ebenfalls auf Gleichungssysteme angewendet werden, die mittels Finite-Differenzen entstanden sind, und eignet sich besonders für Probleme mit komplizierten Geometrien, deren Diskretisierung auf hochgradig nichtuniformen und adaptiv verfeinerten Gitter erfolgt. Der Rechenaufwand für einen Hierarchische Basis Mehrgitter-V-Zyklus ist auch in solchen Fällen stets proportional zur Anzahl n der vorhandenen Freiheitsgrade, was man wiederum leicht anhand der Block-Gauß–Seidel Interpretation über den Erzeugendensystem-Zugang erkennt. Im Gegensatz dazu verlangen gewöhnliche Mehrgitter-V-Zyklen bekanntlich einen geometrischen Abfall der Dimensionen der grobskaligeren Unterräume, damit die Arbeit auf gröberen Gittern nicht zu stark anwächst. Die im Finite-Elemente Kontext für symmetrisch positv definite Probleme entwickelte Konvergenztheorie benötigt als Voraussetzungen lediglich die Formregularität der einzelnen Elemente, jedoch nicht die im Falle variationeller Mehrgitter-Verfahren geforderte globale *Quasiuniformität* der Triangulierung. Die notwendigen Abschätzungen gründen auf der lokalen Elliptizität des Problems und verlangen keine zusätzlichen Regularitätsannahmen, die über die zur schwachen Formulierung gehörende minimale $\mathcal{H}^1(\Omega)$-Forderung hinausgehen. Man erhält damit für zweidimensionale Probleme *suboptimale* Konvergenzresultate, die für Gleichungssysteme, die durch Diskretisierungen bezüglich uniform verfeinerter Triangulierungen enstanden sind, noch eine logarithmische Abhängigkeit von der Maschenweite h_0 in sich tragen. Diese beobachtet man auch in numerischen Beispielen.

Da die klassische Hierarchische Basis Mehrgitter-Methode genauso wie klassisches Mehrgitter auf (bi-)linearen Interpolationen zur Definition der Prolongations- und Restriktionsabbildungen beruht, ist es nicht verwunderlich, daß sie für das Modellproblem (1.1) zu keinen robusten Lösern führt. Ihr Konvergenzverhalten hängt sowohl von der Problemgröße als auch von dem Konvektionsfeld $\vec{b}$ ab [13]. In jüngerer Zeit wurden verschiedene Versuche unternommen, die

Robustheit des Verfahrens durch Einführen verallgemeinerter hierarchischer Basen zu verbessern [14, 35, 96, 97]. Sie werden analog zu den mittels matrixabhängiger Prolongationen und Restriktionen definierten grobskaligen Basisfunktionen durch problemangepaßte Funktionen gebildet. Man nutzt hier eine enge Verbindung zwischen hierarchischen Basen und einer unvollständigen Gauß-Elimination, die implizit zu problemangepaßten und physikalisch sinnvollen Grobgitterdiskretisierungen führt [15]. Trotz der darüber erhaltenen stabilen Grobgitterprobleme, ist die aufgrund des hierarchischen Ansatzes geleistete schwächere Glättungsarbeit nicht ausreichend, um eine Stabilisierung hinsichtlich der Maschenweitenabhängigkeit der Verfahren zu erreichen. Unsere numerischen Beispiele zeigen, daß sogar im Fall exakter Korrekturen der jeweiligen Detail-Anteile sich mit zunehmender Störung die Konvergenz in der Regel weiter verschlechtert oder das Verfahren sogar divergiert.

Betrachtet man anstelle der hierarchischen Basis Wavelet-artige Zerlegungen, so lassen sich damit ebenfalls verallgemeinerte Hierarchische Basis Mehrgitter-Verfahren definieren [35, 92, 121, 122]. Es werden dabei lediglich die Basen der Komplementräume ausgetauscht. Mit Hilfe zusätzlicher Prolongationen und Restriktionen für die Komplemente können darüber entstehende Wavelet-basierte Multiskalen-Zyklen auch als gewöhnliche Mehrgitter-Verfahren interpretiert werden, die spezielle Glättungsiterationen, sogenannte *Multiskalen-Glätter*, benutzen. Im Fall der ungestörten Poissongleichung zeigen solche Multiskalen-Zyklen optimale Konvergenzeigenschaften [121]. Von einem theoretischen Standpunkt aus liegt dies an den verbesserten Stabilitätseigenschaften der entsprechenden Unterraumzerlegungen hinsichtlich der Maschenweiten [28, 91]. Aus praktischer Sicht ist der stärkere Einfluß der zugehörigen Glättungsiterationen dafür verantwortlich. Die Raten sind aber durchweg schlechter als bei klassischen Mehrgitter-Verfahren.

Für Konvektions-Diffusions Probleme ist es im allgemeinen nicht möglich, einen Multiskalen-Glätter zu konstruieren, der das im Zusammenhang mit Mehrgitter-Verfahren erwähnte Robustheitskriterium erfüllt. Aufgrund der mangelnden Flexibilität und Stärke der Multiskalen-Glätter bleibt daher als eine Möglichkeit zur Konstruktion robuster Verfahren noch die Wahl der Multiskalen-Zerlegungen selbst. Es ist sicherzustellen, daß man sowohl hinsichtlich der singulären Störung stabile Grobgitterprobleme als auch bezüglich der Maschenweite stabile Unterraumzerlegungen erhält. Gleichzeitig muß der Aspekt der approximativen Gauß-Elimination beachtet werden, der durch das Zusammenspiel matrixabhängiger Prolongationen und Restriktionen mit einer hierarchischen Basis Zerlegung gegeben ist. Um alle diese Forderungen zu erfüllen, gehen wir in unserem Ansatz von einem sehr allgemeinen Petrov–Galerkin-artigen Multiskalen-Konzept aus, bei dem die Zerlegungen auf der Ansatz- und Testseite unterschiedlich sind. Wir verwenden direkt problemabhängige Vergröberungsstrategien, die von robusten Mehrgitter-Techniken her bekannt sind, zusammen mit Wavelet-artigen und hierarchischen Multiskalen-Zerlegungen der Ansatz- und Testräume $\mathcal{V}_0$ and $\mathcal{S}_0$ bezüglich des feinsten Gitters. Die Kernidee bei unserer Konstruktion robuster Wavelet-basierter Multiskalen-Verfahren ist, jeweils einen der Komplementräume auf der Ansatz- oder Testseite hierarchisch zu wählen, um zusammen mit einer problemabhängigen Vergröberung auf der anderen Seite physikalisch sinnvolle Grobgitterdiskretisierungen und gleichzeitig einen approximativen Eliminationseffekt zu erhalten. Wir verwenden ohne Einschränkung eine hierarchische Zerlegung der Testseite. Die Komplementräume auf der entsprechend anderen Seite werden hingegen Wavelet-artig aufgespannt, was zu einer Stabilisierung des Verfahrens bezüglich der Abhängigkeit von der Maschenweite der Diskretisierung führt. Unsere Strategie veranschaulicht die nachfolgende Abbildung. Numerische Beispiele



zeigen, daß man durch eine solche Wahl zu verallgemeinerten Hierarchische Basis Mehrgitter-Verfahren mit robusten Konvergenzeigenschaften gelangt.

Wir geben nun kurz den Inhalt der restlichen Kapitel dieser Arbeit wieder. In Kapitel 2 diskutieren wir Schwierigkeiten hinsichtlich der Diskretisierung von Konvektions-Diffusions Gleichungen und geben einen Überblick über geeignete Finite-Differenzen und Finite-Elemente Techniken. Es werden insbesondere *gleichmäßig konvergente Methoden* betrachtet, die auf einer exponentiellen Anpassung des Diskretisierungsverfahrens an die Lösungen beruhen. Für diese ist der Diskretisierungsfehler unabhängig von $\vec{b}$, und es bestehen interessante Zusammenhänge mit den von uns in Kapitel 5 konstruierten problemangepaßten Multiskalen-Zerlegungen. Kapitel 3 stellt den allgemeinen Rahmen unseres Petrov–Galerkin Multiskalen-Konzepts vor. Nach einer Herleitung der zu lösenden linearen Gleichungssysteme und ihrer Darstellung ausgehend von einer Petrov–Galerkin-artigen Zweiskalen-Zerlegung führen wir gemäß [19] zwei äquivalente Multiskalen-Darstellungen ein, die *Standard-* und *Nichtstandardform*. Wir zeigen, daß sich beide Darstellungen zusammen mit der aus der nichtklassischen Mehrgitter-Sichtweise stammenden Multilevel-Erzeugendensystem-Darstellung in eine umfassendere Multiskalen-Erzeugendensystem-Darstellung einbetten lassen. In Kapitel 4 leiten wir anhand der Standard- und Nichtstandardform bezüglich Petrov–Galerkin-artiger Multiskalen-Zerlegungen sowohl multiplikative als auch additive Multiskalen-Methoden zur Lösung des Feingittersystems her. Dies geschieht stets nach dem Muster, daß wir klassische Verfahren (ILU, Gauß–Seidel, Jacobi) für die Multiskalen-transformierten Systeme betrachten und die im Vergleich zur Standardform effizientere Nichtstandardform als „Vehikel“ für deren Umsetzung benutzen. Wir diskutieren unterschiedliche Implementierungen und vergleichen sie mit Methoden, die aus der Literatur bekannt sind. Im darauffolgenden Kapitel 5 besprechen wir problemangepaßte Multiskalen-Zerlegungen für Konvektions-Diffusions Probleme. Ihnen

kommt eine Schlüsselrolle für die Konstruktion robuster Methoden zu. Nach der Diskussion problemangepaßter Räume grobskaliger Funktionen für die Ansatz- und Testseite erklären wir ihr Zusammenspiel mit einer hierarchischen und Wavelet-artigen Wahl der Basen der zugehörigen Komplementräume. Wir betrachten dabei zunächst den (trivialen) eindimensionalen Fall und zeigen, wie sich die dazu gehörenden Techniken im Rahmen eines klassischen Tensorprodukt-Ansatzes auf zweidimensionale Probleme übertragen lassen. Für nichtseparable Fälle schlagen wir die Verwendung AMG-basierter hierarchischer Zerlegungen zusammen mit AMG-basierten Wavelet-artigen Multiskalen-Zerlegungen vor. Die von uns betrachteten Stabilisierungen der AMG hierarchischen Basen führen zu völlig neuartigen Basisfunktionen der betreffenden Komplementräume, zu sogenannten *AMGlets*, die der Philosophie algebraischer Verfahren entsprechend auf rein algebraischem Wege konstruiert werden. In Kapitel 6 zeigen wir die numerischen Ergebnisse, die wir durch unsere Verfahren für vier unterschiedliche Klassen von Testproblemen mit zunehmendem Schwierigkeitsgrad erhalten. Wir betrachten zuerst Probleme mit konstanter achsenorientierter Konvektion sowie Konvektion in Diagonalrichtung. Danach untersuchen wir das Verhalten der Verfahren bei Problemen mit konstanter Konvektion in unterschiedlichen Winkeln relativ zur x-Achse. Wir betrachten schließlich noch eine Aufgabe mit nichtseparabler wirbelbehafteter Konvektion. In einem Anhang geben wir der Darstellung [125] folgend eine recht allgemeine Herleitung moderner Krylovraum-Löser und ihrer Konvergenzeigenschaften. Die Matrix T_0 des nach Diskretisierung enstehenden linearen Gleichungssystems ist bekanntlich mit zunehmendem Einfluß des konvektiven Anteils der kontinuierlichen Gleichung zunehmend nichtsymmetrisch und nichtnormal, so daß Konvergenzaussagen, die auf dem Spektrum des Operators basieren, unter Umständen wertlos sind [115]. Mit Hilfe von Konvergenzabschätzungen, die sich am *Numerischen Wertebereich* von T_0 orientieren [72], gelingt uns der Nachweis der Optimalität GMRES-artiger Krylovraum-Löser in einem etwas allgemeineren Kontext als in [108, 109].

Kapitel 2

Diskretisierungsverfahren für Konvektions-Diffusions Probleme

Die Schwierigkeiten bei der numerischen Behandlung von Konvektions-Diffusions Problemen beschränken sich nicht alleine auf das Lösen der resultierenden linearen Gleichungssysteme, sondern treten schon vorher beim Diskretisierungsprozeß auf. Bei Multiskalen-Verfahren zur Lösung der entstehenden linearen Gleichungssysteme setzt man sukzessive Probleme gröberer Auflösungsstufen zur Konvergenzbeschleunigung ein. Diese entstehen entweder durch die explizite Diskretisierung bezüglich gröberer Gitter oder mittels Prolongations- und Restriktionsabbildungen über das Galerkin-Produkt [62]. Daher ist klar, daß die Untersuchung geeigneter Diskretisierungsverfahren auch für die Konstruktion effektiver Multiskalen-Löser von großer Bedeutung ist.

Verwendet man zur Diskretisierung Standardmethoden, so begegnet man bei konvektionsdominierten Aufgaben zunächst einmal einem *Stabilitätsproblem*, das sich in einem unphysikalischen Verhalten der berechneten numerischen Lösungen äußert. Bekanntlich neigt sowohl für gewöhnliche Finite-Differenzen als auch für gewöhnliche Finite-Elemente Diskretisierungen die diskrete Näherungslösung u_h im Fall dominierender Konvektion und realistischer Maschenweiten h zu Oszillationen. (Da wir in diesem Kapitel immer nur ein Gitter oder eine Triangulierung gleichzeitig betrachten, kennzeichnen wir die zugehörigen diskreten Objekte durch einen skalenfreien Index h.) Oszillationen lassen sich insbesondere dort beobachten, wo die Lösung u zur kontinuierlichen Gleichung *Grenzschichten* ausbildet, können sich aber auch in Bereiche fortsetzen, in denen u glatt ist. Man verwendet zur Stabilisierung der resultierenden Gleichungssysteme gerne sogenannte *Upwind-Techniken*. Sie führen meist zu einfachen Verfahren, die unabhängig von der Konvektionsstärke stabil sind, und liefern zufriedenstellende Ergebnisse in Teilbereichen des gegebenen Gebiets Ω, in denen die asymptotische Struktur der Lösung nicht zu kompliziert ist.

Neben dem Wunsch nach stabilen und damit oszillationsfreien Verfahren kann man weitere Forderungen an geeignete Diskretisierungsverfahren stellen. Die Lösungen von Konvektions-Diffusions Gleichungen beschreiben oft Dichten oder Konzentrationen, die von Natur aus nichtnegativ sind. Daß diese Eigenschaft tatsächlich auf die kontinuierlichen Lösungen zutrifft, gewährleisten Maximumsprinzipien, denen die kontinuierlichen Gleichungen gehorchen. Invers-monotone Matrizen und insbesondere M-Matrizen besitzen Eigenschaften, die die Gültigkeit diskreter Maximumsprinzipien garantieren. Es ist natürlich, solche Forderungen auch an die Systemmatrizen der diskretisierten Probleme zu stellen.

Heutzutage beschäftigt sich eine wachsende Zahl von Arbeiten mit der Frage nach *gleichmäßig konvergenten Methoden*, die unabhängig von Art und Stärke der Konvektion genaue Näherungslösungen liefern. Einen Überblick findet man in [44, 85, 86, 102]. Zur Konstruktion solcher Verfahren haben sich mittlerweile zwei prinzipielle Ansätze herausgebildet. Beide versuchen, dem singulären Verhalten der Lösungen in besonderer Weise Rechnung zu tragen. Beim ersten Ansatz modifiziert man das Gitter entsprechend dem Verhalten der Lösungen, wozu natürlich das qualitative Verhalten der Lösungen *a priori* bekannt sein muß. Die zweite Technik versucht im Gegensatz dazu, ausgehend von einem Standardgitter den diskreten Operator dem Verhalten lokaler Lösungen anzupassen. Für mehrdimensionale Probleme existieren bislang aber für beide Ansätze theoretische Resultate nur in wenigen Spezialfällen. Wir sehen in Kapitel 5, daß für die zuletzt genannte Methode interessante Zusammenhänge mit Techniken bestehen, die bei unserer Konstruktion robuster Multiskalen-Löser eingesetzt werden.

In diesem Kapitel betrachten wir die mittels des *Störungsparameters* $\varepsilon = 1/||\vec{b}||_\infty$ skalierte Konvektions-Diffusions Gleichung als das *singulär gestörte Problem*. Damit fassen wir anders als in der Einleitung, jedoch wie in vielen Arbeiten zur numerischen Analysis singulär gestörter Aufgaben üblich, den Diffusionsterm anstatt des Konvektionsterms als singuläre Störung auf. Diese Ansicht widerspricht oft der physikalischen Realität, da in vielen Anwendungen gerade die Konvektionsgeschwindigkeit stark variiert, die Diffusionskoeffizienten des fließenden Mediums aber von vornherein feststehen.

2.1 Finite-Differenzen Verfahren

2.1.1 Upwind-Verfahren in 1D

Wir betrachten zunächst das ungestörte ($\varepsilon = 1$) lineare Randwertproblem

$$Lu := -u'' + b(x)u' + c(x)u = f(x), \quad u(0) = u(1) = 0, \tag{2.1}$$

auf dem beschränkten Intervall $\Omega =]0,1[$ mit hinreichend glatten Koeffizientenfunktionen $b(x)$, $c(x)$, wobei $c(x) \geq 0$ für alle $x \in \bar{\Omega}$ sei (dies sichert die eindeutige Lösbarkeit), sowie glatter rechter Seite $f(x)$. Es seien $\Omega_h := \{x_i \in \Omega : x_i := ih \text{ für } 0 < i < 1/h\}$ ein *äquidistantes Gitter* der Maschenweite h, $N := 1/h$ und für $i = 1, ..., N-1$ $b_i := b(x_i)$, $c_i := c(x_i)$ sowie $f_i := f(x_i)$ Auswertungen der Koeffizientenfunktionen und rechten Seite in den betrachteten Gitterpunkten. Bezeichnen die Unbekannten u_i Approximationen von u in den Punkten von Ω_h und sind u_0 sowie u_N die Werte an den Rändern $x_0 = 0$ und $x_N = 1$, so führt das klassische Finite-Differenzen Verfahren

$$\begin{aligned} -D^+D^-u_i + b_iD^0u_i + c_iu_i &= f_i, \quad \text{für } i = 1, ..., N-1, & (2.2)\\ u_0 = u_N &= 0 & (2.3) \end{aligned}$$

mit der *zentralen Differenz* D^0 sowie den *Vorwärts-* und *Rückwärtsdifferenzen* D^+, D^-

$$(D^0u)(x) := \frac{u(x+h) - u(x-h)}{2h}, \quad (D^\pm u)(x) := \frac{\pm u(x \pm h) \mp u(x)}{h}$$

unter Einbeziehung der Randbedingungen auf ein lineares Gleichungssystem der Gestalt

$$L_hu_h = f_h \qquad (=: R_hf = R_hLu). \tag{2.4}$$

Dabei sind $u_h := [u_0; ...; u_N]$, $f_h := [f_0; ...; f_N]$ *Gitterfunktionen* und R_h die triviale *Restriktion* von $\mathcal{C}([0,1])$ auf den Raum der Gitterfunktionen. (Wir verwenden hier und im folgenden oft eine Notation des Programmiersystems Matlab, um Elemente von (Block-) Vektoren oder (Block-) Matrizen zusammenzufassen.) Die klassische Konvergenztheorie für Finite-Differenzen Verfahren basiert auf den Begriffen *Konsistenz* und *Stabilität*. Die *Konsistenz* beschreibt die Nähe des diskreten Problems zum kontinuierlichen. Die *Stabilität* drückt die stetige Abhängigkeit der diskreten Lösung von der diskreten rechten Seite aus.

Definition 1 (Konsistenz, Stabilität, Konvergenz)
In das diskrete Problem (2.4) seien die Randbedingungen miteinbezogen, und $||\cdot||$ sei eine gewählte Norm. Ein zugehöriges Schema heißt

i) konsistent von der Ordnung k bezüglich $||\cdot||$ *für eine Klasse hinreichend glatter Funktionen u, wenn gilt*
$$||L_h R_h u - R_h L u|| \leq C h^k,$$
wobei C und k unabhängig von h und u sind,

ii) stabil mit Stabilitätskonstante C_{st} bezüglich $||\cdot||$, *wenn für alle Gitterfunktionen u_h gilt*
$$||u_h|| \leq C_{st} ||L_h u_h||,$$
wobei C_{st} unabhängig von h und u_h ist,

iii) konvergent von der Ordnung k bezüglich $||\cdot||$, *wenn gilt*
$$||u_h - R_h u|| \leq C h^k,$$
wobei C und k unabhängig von h, u und u_h sind.

Eine sinnvolle Norm, insbesondere auch für die Behandlung von singulär gestörten Problemen, ist die diskrete Maximumsnorm

$$||u||_{\infty,d} := \max_{i=0,...,N} |u(x_i)|, \tag{2.5}$$

die wir bei Finite-Differenzen Techniken durchweg betrachten werden. Eine interessante Diskussion unterschiedlicher Normen hinsichtlich der Lösungen von Konvektions-Diffusions Problemen findet man in [44, 71]. Wegen der Gleichung

$$L_h(R_h u - u_h) = L_h R_h u - f_h = L_h R_h u - R_h L u$$

erhält man anhand der obigen Definitionen, daß Konsistenz und Stabilität zusammen die Konvergenz eines Verfahrens implizieren. Bekanntlich ist das klassische Finite-Differenzen Verfahren (2.2), (2.3) im Fall der ungestörten Gleichung (2.1) konvergent von der Ordnung 2 bezüglich der diskreten Maximumsnorm [60].
Wir betrachten nunmehr das singulär gestörte Problem

$$Lu := -\varepsilon u'' + b(x)u' + c(x)u = f(x), \quad u(0) = u(1) = 0, \tag{2.6}$$

ebenfalls auf $\Omega =]0,1[$ mit $0 < \varepsilon \ll 1$ unter der zusätzlichen Voraussetzung, daß die Koeffizientenfunktion $b(x)$ keinem Vorzeichenwechsel unterliegt und *Umkehrpunkte*, das heißt Punkte x mit $b(x) = 0$, ausgeschlossen sind. Schon einfache Beispiele, wie etwa

$$-\varepsilon u'' - u' = 0, \quad u(0) = 0, \quad u(1) = 1$$

(die Transformation $v(x) := u(x) - x$ würde ein äquivalentes Problem für v mit homogenen Randwerten ergeben), mit der exakten Lösung $u(x) = (1 - \exp(-x/\varepsilon))/(1 - \exp(-1/\varepsilon))$ zeigen eine Schwierigkeit, die sich durch Verwenden des Verfahrens (2.2), (2.3) ergibt. Die zugehörige diskrete Lösung lautet

$$u_i = \frac{1 - r^i}{1 - r^N} \quad \text{mit } r = \frac{2\varepsilon - h}{2\varepsilon + h}$$

und oszilliert im Falle $h > 2\varepsilon$, obwohl die exakte Lösung monoton ist. Ursache hierfür ist, daß sich aufgrund der singulären Störung am Rand $x = 0$ eine *Grenzschicht* befindet, in der sich mit kleiner werdendem ε die Lösung u extrem schnell ändert. Mit der zentralen Differenz D^0 zur Approximation der Ableitung erster Ordnung erhält man als Diskretisierungsmatrix L_h für die allgemeine Aufgabe (2.6) dann eine M-Matrix und damit ein stabiles Verfahren (siehe dazu das M-Matrix-Kriterium in [102] sowie die Diskussion in [44]), falls $h < h_0(\varepsilon) := 2\varepsilon/||b||_{\infty,d}$ ist. Hieraus folgt die Nichtpositivität der Außerdiagonalelemente von L_h und aufgrund der irreduziblen Diagonaldominanz der Matrix damit auch deren M-Matrix-Eigenschaft. Wegen $h_0(\varepsilon) \longrightarrow 0$ für $\varepsilon \to 0$ schließen wir umgekehrt, daß klassische Finite-Differenzen Verfahren auf äquidistanten Gittern hinsichtlich singulär gestörter Probleme in dem Sinne scheitern, daß sie nur bei sehr kleinen Schrittweiten zufriedenstellende Resultate liefern.

Einfache Upwind-Verfahren

Einen möglichen Ausweg bieten *Upwind-Verfahren*, die einseitige Differenzen zur Approximation der Ableitung erster Ordnung heranziehen, um so das „richtige“ Vorzeichen in der Diskretisierungsmatrix zu erzwingen.

Schema 1 (Einfaches Upwind-Verfahren)

$$\begin{aligned} -\varepsilon D^+D^-u_i + b_i D^{\aleph}u_i + c_i u_i &= f_i, \quad \text{für } i = 1, ..., N-1, && (2.7)\\ u_0 = u_N &= 0, && (2.8) \end{aligned}$$

wobei

$$D^{\aleph} := \begin{cases} D^+, & \text{für } b < 0, \\ D^-, & \text{für } b > 0. \end{cases}$$

Der Konvektionsanteil $b(x)u'$ in der kontinuierlichen Differentialgleichung bestimmt je nach Vorzeichen von b eine *Flußrichtung* für den beschriebenen Transportvorgang. Der Ausdruck *Upwind* bedeutet, daß die Finite-Differenzen Approximation des konvektiven Anteils in den einzelnen Gitterpunkten jeweils durch die stromaufwärts davon gelegenen Gitterpunkte bestimmt wird. Dies entspricht auch der Vorstellung, die man hinsichtlich des „Informationsflußes“ hat.

Satz 1 (Eigenschaften des einfachen Upwind-Verfahrens)

i) Die Diskretisierungsmatrix für das einfache Upwind-Verfahren ist eine M-Matrix. Das Verfahren ist in der diskreten Maximumsnorm gleichmäßig stabil bezüglich ε:

$$||u_h||_{\infty,d} \leq C_{st}||L_h u_h||_{\infty,d}$$

ii) Außerhalb der Grenzschicht (das heißt in $[0, 1-\delta]$ *bzw.* $[\delta, 1]$ *mit* $\delta > 0$*, fest) konvergiert das Verfahren in erster Ordnung gleichmäßig bezüglich* ε *in der diskreten Maximumsnorm. Innerhalb der Grenzschicht liegt im allgemeinen keine Konvergenz vor.*

Beweis: Siehe [102]

□

Um ein hinsichtlich des Störungsparameters ε gleichmäßig stabiles Verfahren zu konstruieren, müssen wir auf diesem Wege also eine geringere Genauigkeit in Kauf nehmen. Sie drückt sich in einer Abnahme der Konsistenz des Verfahrens aus. Für die weitere Diskussion nehmen wir $b(x) \geq \underline{b} > 0$ an. Der Fall $b(x) \leq \underline{b} < 0$ kann analog behandelt werden.

Allgemeine Upwind-Verfahren

Wir können durch einfaches Umschreiben des zweiten Terms in (2.7) das zugehörige Schema auch als Zentrale-Differenzen Verfahren für eine Gleichung mit modifiziertem Diffusionskoeffizienten interpretieren:

$$(2.7) \Longleftrightarrow -(\varepsilon + \frac{b_i h}{2})D^+D^-u_i + b_i D^0 u_i + c_i u_i = f_i, \quad \text{für } i = 1, ..., N-1. \tag{2.9}$$

Fügt man also eine geeignete Menge *künstlicher Diffusion (Viskosität)* zum ursprünglichen Problem hinzu, so ergibt das klassische Finite-Differenzen Verfahren (2.2), (2.3) für das auf diese Weise *modifizierte* Problem eine stabile Diskretisierung. Es ist möglich, das eben beschriebene Vorgehen über eine Funktion σ, die in Abhängigkeit von b, h und ε die Diffusion des modifizierten Problems beschreibt, in allgemeinerer Weise darzustellen:

Schema 2 (Allgemeines Upwind-Verfahren mit künstlicher Diffusion)

$$-\varepsilon\sigma(q(x_i))D^+D^-u_i + b_i D^0 u_i + c_i u_i = f_i, \quad \text{für } i = 1, ..., N-1, \tag{2.10}$$

$$u_0 = u_N = 0, \tag{2.11}$$

wobei $q(x) = \frac{b(x)h}{2\varepsilon}$ *ist.*

Im Fall $\sigma(q) = 1 + q$ erhält man gerade das einfache Upwind-Schema 1. Wir notieren als Eigenschaften:

Satz 2 (Eigenschaften des allgemeinen Upwind-Verfahrens)
Es gelte $\sigma(q) > q$.

i) Die zum Schema 2 gehörende Diskretisierungsmatrix ist eine M-Matrix und das sich damit ergebende Verfahren bezüglich ε *gleichmäßig stabil in der diskreten Maximumsnorm.*

ii) Es gelte zusätzlich $|\sigma(q) - 1| \leq \min\{q, mq^2\}$ *mit einer Konstanten* $m > 0$*. Die Konvergenzordnung beträgt für festes* ε *dann zwei. Sie beträgt eins gleichmäßig bezüglich* ε *in Bereichen außerhalb der Grenzschicht. Diese Aussagen beziehen sich alle auf die diskrete Maximumsnorm.*

Beweis: Siehe [102]

□

Man kann versuchen, solche Upwind-Verfahren mit künstlicher Diffusion auch für höherdimensionale Probleme einzusetzen, doch ist das Bestimmen der richtigen Menge an hinzuzufügender Diffusion im allgemeinen recht schwierig, da zuviel Viskosität ein „Verschmieren“ der Grenzschichten in der berechneten Lösung bewirkt.

2.1.2 Gleichmäßig konvergente Verfahren in 1D

Aufgrund des Verhaltens der Lösungen innerhalb der Grenzschicht sind die bislang vorgestellten Verfahren im allgemeinen nicht bezüglich ε gleichmäßig konvergent in der diskreten Maximumsnorm. Schlimmer noch: Man beobachtet, daß in vielen Fällen für festes ε der Fehler zur exakten Lösung mit zunehmend feiner werdenden Gittern solange wächst, bis die Maschenweite h die gleiche Größenordnung wie der Diffusionsparameter ε erreicht. Erst dann nimmt der Fehler bei weiterer Verfeinerung wieder ab, wie man es von einem sinnvollen Verfahren erwartet. Für die Konstruktion von Verfahren, die gleichmäßig gut für alle Werte von ε funktionieren sollen, haben sich mittlerweile zwei prinzipielle Herangehensweisen herausgebildet. Bei der ersten Methode modifiziert man das Gitter entsprechend dem Verhalten der Lösungen ("fitted mesh methods"). Insbesondere müssen dazu im Grenzschichtgebiet extrem kleine Maschenweiten gewählt werden, was natürlich die Kenntnis des qualitativen Verhaltens der Lösungen voraussetzt. Es gibt verschiedene Ansätze mit unterschiedlich komplizierten Gitterkonstruktionen. Viele Ergebnisse und Literaturhinweise findet man in [44, 85]. Um bezüglich ε gleichmäßig konvergente Verfahren zu konstruieren, reichen meist jedoch stückweise uniforme Gitter aus, wie Shishkin unlängst gezeigt hat [85]. Die zweite Technik versucht, den diskreten Operator auf einem gewöhnlichen Gitter dem Verhalten lokaler Lösungen anzupassen ("fitted operator methods"). Solche problemangepaßten Diskretisierungen spielen auch für die von uns betrachteten Multiskalen-basierten Lösungsverfahren der resultierenden linearen Gleichungssysteme eine wichtige Rolle, wie wir in Kapitel 5 sehen werden. Wir stellen daher nur die zuletzt genannte Methode vor.

Ausgehend von dem verallgemeinerten Upwind-Schema 2 kann man die folgende notwendige Bedingung für ein auf dem gesamten Intervall bezüglich ε gleichmäßig konvergentes Verfahren zeigen.

Lemma 1 (Notwendige Bedingung für gleichmässig konvergente Verfahren)
Schema 2 sei gleichmäßig konvergent bezüglich ε in der diskreten Maximumsnorm. Es seien $n \in \mathbb{N}$ und $\rho = \frac{h}{\varepsilon}$ fest. Dann gilt mit $q(x) = \frac{b(x)h}{2\varepsilon}$

$$\lim_{h \to 0} \sigma(q(x_{N-n})) = q(1) \coth q(1).$$

Beweis: Siehe [102]

□

Schema 3 (Il'in–Allen–Southwell Schema)
Die aufgrund der obigen notwendigen Bedingung natürliche Wahl

$$\sigma(q(x)) = q(x) \coth q(x)$$

im verallgemeinerten Upwind-Schema 2 führt zum Il'in–Allen–Southwell Schema.

Diese problemangepaßte Diskretisierungstechnik wurde von Allen und Southwell in [2] vorgeschlagen. Der erste Beweis, daß man damit ein bezüglich ε gleichmäßig konvergentes Verfahren erhält, stammt von Il'in [74]. Im Grenzübergang $h/\varepsilon \longrightarrow \infty$ ergibt sich wegen $\coth z \longrightarrow \pm 1$ für $z \to \pm\infty$ gerade das einfache Upwind-Schema 1.

Satz 3 (Eigenschaften des Il'in–Allen–Southwell Schemas)

i) Das Il'in–Allen–Southwell Schema ist bezüglich ε gleichmäßig stabil und für festes ε konsistent von der Ordnung 2 in der diskreten Maximumsnorm.

ii) Das Il'in–Allen–Southwell Schema ist von erster Ordnung gleichmäßig konvergent bezüglich ε in der diskreten Maximumsnorm:

$$||u - u_h||_{\infty,d} \leq Ch \quad \text{mit } C \neq C(h, \varepsilon).$$

Beweis: Siehe [102]

□

Wir skizzieren nun eine weitere Möglichkeit zur Herleitung des Il'in–Allen–Southwell Verfahrens. Betrachtet man den zu $Mw := -\varepsilon w'' + bw'$ formal adjungierten Operator $M^*v := -\varepsilon v'' - (bv)'$, so gilt für hinreichend glatte Funktionen mit homogenen Randwerten

$$\int_0^1 (Mv)w dx = \int_0^1 v(M^*w)dx.$$

Es seien weiterhin g_i für $i = 1, ..., N-1$ die *lokalen Greenschen Funktionen* zu M^* bezüglich der Gitterpunkte x_i mit Träger $\text{supp}(g_i) = [x_{i-1}, x_{i+1}]$. Für diese gilt also

$$\begin{aligned} 0 &= M^* g_i \quad \text{in }]x_{i-1}, x_i[\cup]x_i, x_{i+1}[, \\ 0 &= g_i(x_{i-1}) = g_i(x_{i+1}), \\ 1 &= \varepsilon(g_i'(x_i - 0) - g_i'(x_i + 0)). \end{aligned}$$

Wegen

$$\int_{x_{i-1}}^{x_{i+1}} (Mu)g_i dx = \int_{x_{i-1}}^{x_{i+1}} (f - cu)g_i dx$$

erhält man nach partieller Integration die Identität

$$-\varepsilon g_i'(x_{i-1})u(x_{i-1}) + u(x_i) + \varepsilon g_i'(x_{i+1})u(x_{i+1}) = \int_{x_{i-1}}^{x_{i+1}} (f - cu)g_i dx.$$

Das damit entstehende Verfahren ist im Falle $c = 0$ offensichtlich *exakt*, das heißt die berechnete Lösung stimmt in den Gitterpunkten mit der exakten Lösung überein. (Dies gilt ebenso im allgemeinen Fall, falls man statt M den gesamten Operator L aus (2.6) in obiger Herleitung verwendet.) Man kann das Schema aber nicht direkt benutzen, da die g_i im Fall variabler Koeffizientenfunktionen im allgemeinen nicht bekannt sind. Ersetzt man die Funktionen b und f auf den Intervallen $]x_{i-1}, x_{i+1}[$ durch die konstanten Werte b_i und f_i, so daß die lokalen Greenschen Funktionen dafür bestimmt werden können, so erhält man unter Verwendung der Quadraturformel

$$\int_{x_{i-1}}^{x_{i+1}} (f - cu)g_i dx \approx (f_i - c_i u_i) \int_{x_{i-1}}^{x_{i+1}} g_i dx$$

und der Abkürzung $\rho_i = \frac{b_i h}{\varepsilon}$ das Schema

$$-\frac{e^{\rho_i} - 1}{e^{\rho_i} - e^{-\rho_i}} u_{i-1} + u_i - \frac{1 - e^{-\rho_i}}{e^{\rho_i} - e^{-\rho_i}} u_{i+1} = (f_i - c_i u_i) \frac{h}{b_i} \frac{e^{\rho_i} - 1}{e^{\rho_i} + 1}.$$

Dieses stimmt, wie man nach Umformen sieht, mit dem Il'in–Allen–Southwell Schema überein. Hinsichtlich dieser Herleitung des Il'in–Allen–Southwell Verfahrens bieten sich sofort die folgenden beiden Verbesserungsmöglichkeiten an. Man kann unterschiedliche Konstanten zur Approximation der Funktionen b und f auf den Intervallen $]x_{i-1}, x_i[$ und $]x_i, x_{i+1}[$ benutzen, zum Beispiel $b_i := [b(x_{i-1}) + b(x_i)]/2$ auf $]x_{i-1}, x_i[$. Mit Hilfe einer verbesserten Quadraturformel ergibt sich so genau das *El–Mistikawy–Werle Schema*, das in zweiter Ordnung gleichmäßig konvergent bezüglich ε in der diskreten Maximumsnorm ist [102]. Anstatt des vereinfachenden Operators M kann man auch den ursprünglichen Operator L aus (2.6) verwenden. Die darüber entstehenden Verfahren heißen *Verfahren mit vollständigem exponentiellen Fitting*, da sie alle Terme des zugrundeliegenden Differentialoperators berücksichtigen. Weitere Möglichkeiten zur Herleitung des Il'in Verfahrens findet man in [101]. Wir zeigen darüberhinaus in Kapitel 5, daß es möglich ist, das Il'in Verfahren auch aus Sicht eines operatorabhängigen Vergröberungsprozeßes zu verstehen, der einer Multilevel-basierten Lösung der resultierenden linearen Gleichungssysteme dient.

2.1.3 Upwind-Verfahren in 2D

Das Lösungsverhalten von Konvektions-Diffusions Gleichungen in zwei Raumdimensionen ist wesentlich komplizierter als im eindimensionalen Fall. Neben *Einström-* und *Ausströmrändern*, an denen sich *exponentielle Grenzschichten* entwickeln können, treten im allgemeinen auch *charakteristische Ränder* auf, an denen die Lösungen *parabolische Grenzschichten* ausbilden. Parabolische Schichten können ebenfalls im Inneren des Gebiets entlang *Subcharakteristiken* entstehen, die von Randpunkten des Einströmrandes ausgehen und an denen die rechte Seite Unstetigkeiten aufweist. Als *Subcharakteristiken* bezeichnet man hierbei die Charakteristiken des *reduzierten Problems*, das man erhält, wenn man den diffusiven Anteil des Operators auf Null setzt (d.h. $\varepsilon = 0$) und auf die Randbedingungen am Einströmrand verzichtet. Dieses ist dann rein hyperbolischer Natur. Weiter führen im Fall wirbelbehafteter Konvektion *geschlossene Subcharakteristiken* zu ernsten Schwierigkeiten. Für eine präzise Darstellung der auftretenden Phänomene verweisen wir auf [50, 102].

Wir betrachten das zweidimensionale Konvektions-Diffusions Problem

$$\begin{aligned} Lu := -\varepsilon \triangle u + \vec{b}(x,y) \cdot \nabla u + c(x,y)u &= f \quad \text{in } \Omega =]0,1[^2, && (2.12)\\ u &= 0 \quad \text{auf } \partial\Omega, && (2.13) \end{aligned}$$

mit hinreichend glatten Koeffizientenfunktionen $\vec{b}(x,y)$ und $c(x,y)$ sowie glatter rechter Seite $f(x,y)$ auf $\bar{\Omega}$. Hierbei ist $\vec{b}(x,y)$ ein zweidimensionales Vektorfeld. Es gelte ferner $c(x,y) \geq 0$ auf $\bar{\Omega}$, was unter entsprechenden Regularitätsannahmen die Existenz einer eindeutig bestimmten klassischen Lösung $u \in \mathcal{C}(\bar{\Omega}) \cap \mathcal{C}^2(\Omega)$ sichert [102]. Die Lösung u zu (2.12), (2.13) genügt bekanntlich einem *Maximumsprinzip*, aus dem das nachstehende *Vergleichsprinzip* folgt [95]:

$$Lu \geq 0 \quad \text{in } \Omega \quad \text{und} \quad u \geq 0 \quad \text{auf } \partial\Omega \quad \Longrightarrow \quad u \geq 0 \quad \text{in } \bar{\Omega}.$$

Man sagt auch, der Operator L ist *invers monoton*. Die diskrete Entsprechung hierzu sind *invers monotone Matrizen*. Eine Matrix L_h heißt *invers monoton*, wenn $L_h^{-1} \geq 0$, das heißt $(L_h^{-1})_{i,j} \geq 0$ für alle möglichen Indizes i, j gilt. Diese Eigenschaft ist etwas schwächer als die, eine M-Matrix zu sein. Bei M-Matrizen wird zusätzlich gefordert, daß alle Außerdiagonalelemente der Matrix kleiner oder gleich Null sind [60]. Wie man leicht sieht, erfüllen invers monotone Matrizen ein *diskretes Vergleichsprinzip*:

$$L_h x \geq 0 \quad \Longrightarrow \quad x \geq 0,$$

wobei x einen entsprechend dimensionierten reellen Vektor darstellt und die letzte Ungleichung komponentenweise zu lesen ist. Die durch Konvektions-Diffusions Gleichungen beschriebenen Größen stellen oft Dichten oder Konzentrationen dar, die von Natur aus nichtnegativ sind. Ausgehend von einer Situation, in der keine Quellen oder Senken vorhanden sind, das heißt $f = 0$, folgt aus dem Vergleichsprinzip, daß eine Lösung u der homogenen Gleichung $Lu = 0$ mit nichtnegativen Randwerten auch im Innern des Gebietes nichtnegativ ist. Diskretisierungen, die als Systemmatrizen keine invers monotone Matrizen ergeben, liefern entgegen der physikalischen Erwartung dann aber im allgemeinen Lösungen, die negative Werte annehmen. Eine physikalisch sinnvolle Diskretisierung sollte daher eine invers monotone Matrix ergeben. Invers monotone Diskretisierungsmatrizen ergeben sich zum Beispiel durch die Erweiterung des einfachen Upwind-Schemas 1 auf die zweidimensionale Situation. Im Fall, daß der Reaktionsterm verschwindet, das heißt $c = 0$, erhalten wir

Schema 4 (Einfaches Upwind-Verfahren in 2D)
Der zum einfachen Upwind-Verfahren in zwei Raumdimensionen gehörende diskrete Operator lautet für $c = 0$ lokal in Sternnotation: $(L_{h,(i,j)})_* =$

$$-\frac{\varepsilon}{h^2}\begin{bmatrix} \cdot & 1 & \cdot \\ 1 & -4 & 1 \\ \cdot & 1 & \cdot \end{bmatrix} + \frac{1}{h}\begin{bmatrix} \cdot & \cdot & \cdot \\ -b^+_{1,(i,j)} & |b_{1,(i,j)}| & b^-_{1,(i,j)} \\ \cdot & \cdot & \cdot \end{bmatrix} + \frac{1}{h}\begin{bmatrix} \cdot & b^-_{2,(i,j)} & \cdot \\ \cdot & |b_{2,(i,j)}| & \cdot \\ \cdot & -b^+_{2,(i,j)} & \cdot \end{bmatrix}, \qquad (2.14)$$

wobei $b_{1,(i,j)} := b_1(ih, jh)$, $b^+_{1,(i,j)} := \max(b_{1,(i,j)}, 0)$ sowie $b^-_{1,(i,j)} := \min(b_{1,(i,j)}, 0)$ sind und die analogen Bezeichnungen für die Ausdrücke bezüglich $b_2(x, y)$ gelten.

Das resultierende Verfahren ist konsistent von erster Ordnung und bezüglich ε gleichmäßig stabil in der diskreten Maximumsnorm. Es führt zu invers monotonen Diskretisierungsmatrizen, verschmiert aber im allgemeinen die Grenzschichten ähnlich wie im eindimensionalen Fall und kann sogar zu großen Ungenauigkeiten führen. In [26] wird etwa für das Beispiel einer wirbelbehafteten Strömung in einem Kreisring gezeigt, daß der Diskretisierungsfehler überall von der Ordnung $\mathcal{O}(1)$ ist außer am Rand des Gebiets.

Wie bereits erwähnt wurde, lassen sich Upwind-Verfahren alternativ auch dadurch konstruieren, daß man ein gewöhnliches Finite-Differenzen Verfahren auf eine Gleichung anwendet, deren diffusiver Anteil künstlich erhöht wurde. Das Hinzufügen der künstlichen Diffusion kann hierbei isotrop oder anisotrop erfolgen. Eine weitere Möglichkeit besteht darin, ausgehend von einem sinnvollen Upwind-Schema im Eindimensionalen, dies mittels einer Tensorprodukt-Konstruktion auf den zweidimensionalen Fall zu erweitern. Die Lösungen, die man durch diesen Ansatz erzielt, verschmieren oft aber in besonderer Weise die Grenzschichtbereiche der exakten Lösung. Dies liegt daran, daß aufgrund der Tensorprodukt-Struktur der so gebildeten Verfahren unter Umständen Information nicht entlang der Subcharakteristiken sondern transversal dazu übertragen wird ("crosswind-diffusion"). Um allgemeinere Gebiete behandeln zu können, ist entweder eine Randanpassung notwendig, beispielsweise mittels einer Shortley–Weller Diskretisierung [65], oder man kann dazu übergehen, Dreiecksgitter zu verwenden. Für beliebige Dreiecksgitter sind die bisher beschriebenen Finiten-Differenzen nicht mehr direkt anwendbar. Die Diskretisierung des diffusiven Anteils erfolgt deshalb meist mittels einer Finite-Volumen Technik [17]. Der konvektive Anteil wird durch Festlegung eines sogenannten *Upwind-Dreiecks* über eine einseitige Differenz approximiert. Hieraus resultiert der Upwind-Charakter des Verfahrens. Für eine genauere Darstellung verweisen wir auf [94, 102].

2.1.4 Gleichmäßig konvergente Verfahren in 2D

Analog zum eindimensionalen Fall läßt sich eine notwendige Bedingung für gleichmäßig konvergente Verfahren in zwei Raumdimensionen finden. Wir erhalten als Erweiterung des Il'in–Allen–Southwell Verfahrens für zweidimensionale Probleme im Fall, daß der Reaktionsterm verschwindet:

Schema 5 (Il'in–Allen–Southwell Verfahren in 2D)
Der zum Il'in–Allen–Southwell Schema in zwei Raumdimensionen gehörende diskrete Operator lautet für $c = 0$ lokal in Sternnotation: $(L_{h,(i,j)})_* =$

$$-\frac{\varepsilon}{h}\begin{bmatrix} \cdot & 1 & \cdot \\ 1 & -4 & 1 \\ \cdot & 1 & \cdot \end{bmatrix} + \frac{b_{1,(i,j)}}{2}\begin{bmatrix} \cdot & \cdot & \cdot \\ -1-p & 2p & 1-p \\ \cdot & \cdot & \cdot \end{bmatrix} + \frac{b_{2,(i,j)}}{2}\begin{bmatrix} \cdot & 1-q & \cdot \\ \cdot & 2q & \cdot \\ \cdot & -1-q & \cdot \end{bmatrix}, \tag{2.15}$$

mit den lokalen Upwind-Parametern

$$p = \coth\left(\frac{b_{1,(i,j)}h}{2\varepsilon}\right) - \frac{2\varepsilon}{b_{1,(i,j)}h} \quad \text{und} \quad q = \coth\left(\frac{b_{2,(i,j)}h}{2\varepsilon}\right) - \frac{2\varepsilon}{b_{2,(i,j)}h}.$$

Im Fall $\vec{b}(x,y) = (b_1(x), b_2(y)) \geq (\underline{b}_1, \underline{b}_2) > (0,0)$ konvergiert das Il'in–Allen–Southwell Verfahren unter gewissen Kompatibilitätsbedingungen, die an die rechte Seite zu stellen sind, gleichmäßig bezüglich ε in der diskreten Maximumsnorm und hat eine Konvergenzordnung, die nahe 0.5 liegt [102].

2.2 Finite-Elemente Verfahren

2.2.1 Upwind-Verfahren in 1D

Elliptische Randwertaufgaben können über ihre *schwache Formulierung* auch als Variationsprobleme in einem Hilbertraum $\mathcal{V}$ dargestellt werden:

$$\text{Finde } u \in \mathcal{V} \; : \quad a(u,v) = f(v) \quad \forall v \in \mathcal{V}. \tag{2.16}$$

Hierbei ist $a(\cdot\,,\cdot)$ eine zum Differentialoperator gehörende Bilinearform auf $\mathcal{V} \times \mathcal{V}$ und $f(\cdot)$ ein durch die rechte Seite der Differentialgleichung gegebenes stetiges lineares Funktional auf $\mathcal{V}$ [22]. Für Randwertaufgaben zweiter Ordnung ist der Hilbertraum $\mathcal{V}$ für gewöhnlich ein Teilraum des Sobolevraums $\mathcal{H}^1(\Omega)$. Im Fall homogener Dirichletscher Randbedingungen gilt $\mathcal{V} = \mathcal{H}_0^1(\Omega)$. Der Satz von Lax–Milgram sichert unter den Voraussetzungen

$$\begin{array}{rcll} \alpha||v||_{\mathcal{V}}^2 & \leq & a(v,v) & \forall v \in \mathcal{V} \qquad (\mathcal{V}\text{-}\textit{Elliptizität, Koerzivität}), \\ a(v,w) & \leq & \beta||v||_{\mathcal{V}}||w||_{\mathcal{V}} & \forall v,w \in \mathcal{V} \qquad (\textit{Stetigkeit}) \end{array}$$

die Existenz einer eindeutigen Lösung zu (2.16). Die Konstanten α und β werden dabei durch den Differentialoperator und die Koeffizientenfunktionen bestimmt. Die Lösung hängt stetig von den Daten ab, die Konstante in der zugehörigen Abschätzung wird aber mit abnehmender $\mathcal{V}$-Elliptizität des Problems zunehmend schlechter. Die Wahl endlichdimensionaler Räume $\mathcal{V}_h$, die den kontinuierlichen Lösungsraum $\mathcal{V}$ approximieren, führt zu endlichdimensionalen Variationsproblemen, die wiederum nach dem Satz von Lax–Milgram eindeutig lösbar sind

und über lineare Gleichungssysteme dargestellt werden können. Die zugehörige Matrix heißt *Steifigkeitsmatrix*. Dieses Vorgehen entspricht genau dem *Ritz–Galerkin-Verfahren*. Es heißt *konform*, falls die approximierenden Räume $\mathcal{V}_h$ Unterräume von $\mathcal{V}$ sind, was nicht notwendig der Fall sein muß. Als *Finite-Elemente Verfahren* bezeichnet man Ritz–Galerkin-Verfahren, bei denen $\mathcal{V}_h$ von stückweise polynomialen Basisfunktionen mit kompaktem Träger aufgespannt wird. Die zugehörige Steifigkeitsmatrix ist dann im Unterschied zu allgemeinen Ritz–Galerkin-Verfahren schwachbesetzt.

Wir betrachten wieder das singulär gestörte Problem (2.6), nunmehr unter den Voraussetzungen $b(x) \geq \underline{b} > 0$ und $c(x) - \frac{b'(x)}{2} \geq \underline{c} > 0$ für alle $x \in \bar{\Omega}$, wobei letztere die Voraussetzungen des Satzes von Lax–Milgram sichert. Wir definieren als zugehörige Bilinearform $a(v,w) := \varepsilon(v',w') + (bv' + cv, w)$ und betrachten als approximativen Lösungsraum $\mathcal{V}_h$ den Raum der stückweise linearen Funktionen bezüglich eines äquidistanten Gitters Ω_h. Eine Basis von $\mathcal{V}_h$ bilden lineare Splines, die sogenannten *Hutfunktionen*, die für $i = 1, ..., N-1$ $\varphi_i(x_j) = \delta_{i,j}$ erfüllen. Berechnet man in einem damit gebildeten Finite-Elemente Ansatz die auftretenden Integrale über die Mittelpunktsregel, so kommt man zu dem Schema

$$\begin{aligned} -\varepsilon D^+D^-u_i + \frac{1}{2}(b_{i+\frac{1}{2}}D^+u_i + b_{i-\frac{1}{2}}D^-u_i) + \frac{1}{2}(c_{i+\frac{1}{2}} + c_{i-\frac{1}{2}})u_i &= \frac{1}{2}(f_{i+\frac{1}{2}} + f_{i-\frac{1}{2}}), \\ u_0 = u_N &= 0 \end{aligned}$$

für $i = 1, ..., N-1$. Es ähnelt sehr dem Verfahren, das man mit Hilfe des klassischen Finite-Differenzen Schemas (2.2), (2.3) erhält, und stimmt mit diesem im Fall konstanter Koeffizienten sogar überein. Klassische Finite-Elemente Verfahren scheitern daher offensichtlich ebenfalls bei singulär gestörten Problemen.

Upwind Finite-Elemente Verfahren kann man für singulär gestörte Aufgaben über einen Petrov–Galerkin-Ansatz konstruieren, bei dem bekanntlich unterschiedliche Ansatz- und Testräume $\mathcal{V}_h$ und $\mathcal{S}_h$ in der schwachen Formulierung des Problems benutzt werden [8]. Verwendet man lineare Ansatz- und quadratische Testfunktionen (für deren genaue Definition siehe [102]), so erhält man

Schema 6 (Upwind-Verfahren durch einen Petrov–Galerkin-Ansatz)

$$\begin{aligned} -\varepsilon D^+D^-u_i + (\bar{b}_{i+\frac{1}{2}}D^+u_i + \bar{b}_{i-\frac{1}{2}}D^-u_i) + (\bar{c}_{i+\frac{1}{2}} + \bar{c}_{i-\frac{1}{2}})u_i &= (\bar{f}_{i+\frac{1}{2}} + \bar{f}_{i-\frac{1}{2}}), \\ u_0 = u_N &= 0 \end{aligned}$$

für $i = 1, ..., N-1$, wobei $\bar{q}_{i\pm\frac{1}{2}} := (\frac{1}{2} \mp \kappa_{i\pm\frac{1}{2}})q_{i\pm\frac{1}{2}}$ für $q = b, c, f$ ist und die $\kappa_{i\pm\frac{1}{2}}$ lokale Upwind-Parameter bezeichnen, mittels derer die linearen Ansatzfunktionen mit einer quadratischen Funktion zu den Testfunktionen linearkombiniert werden.

Schema 6 läßt sich ebenfalls in der gefitteten Form (2.10) schreiben.

Für eine elementare Fehleranalyse, die die Schwierigkeiten von Konvektions-Diffusions Problemen im Zusammenhang mit einfachen Finite-Elemente Verfahren weiter verdeutlichen soll, betrachten wir zunächst wieder den Galerkin-Fall. Wir wählen dazu die problemspezifische ε-gewichtete $\mathcal{H}^1(\Omega)$-Norm

$$||u||_\varepsilon^2 := \varepsilon|u|_1^2 + ||u||^2, \tag{2.17}$$

die der Energienorm des zugehörigen symmetrischen Problems entspricht. Es gelten dann die folgenden Koerzivitäts- und Stetigkeitsabschätzungen gleichmäßig bezüglich ε:

$$\alpha||v||_\varepsilon^2 \leq a(v,v) \qquad \forall v \in \mathcal{V}, \tag{2.18}$$

$$|a(v,w)| \leq \beta||v||_\varepsilon||w||_1 \quad \forall v, w \in \mathcal{V}. \tag{2.19}$$

Die Abschätzung des Diskretisierungsfehlers durch den Approximationsfehler (Céa Lemma) ergibt nun die bezüglich ε gleichmäßig geltende Abschätzung

$$||u - u_h||_\varepsilon \leq C \inf_{v_h \in \mathcal{V}_h} ||u - v_h||_1 \quad \text{mit } C \neq C(h, \varepsilon). \tag{2.20}$$

Da für polynomiale Finite-Elemente Räume $\mathcal{V}_h$ und Probleme mit Grenzschichten im allgemeinen jedoch $\inf_{v_h \in \mathcal{V}_h} ||u - v_h||_1 \longrightarrow \infty$ für $\varepsilon \to 0$ gilt, deutet dieser Sachverhalt zusammen mit Abschätzung (2.20) darauf hin, daß man für die ε-gewichtete $\mathcal{H}^1(\Omega)$-Norm mit polynomialen Finite-Elemente Räumen keine bezüglich ε gleichmäßig konvergenten Verfahren erhält. Ähnliches gilt auch für entsprechende Petrov–Galerkin-Verfahren [102].

2.2.2 Gleichmäßig konvergente Verfahren in 1D

Wir wenden uns nun der Konstruktion bezüglich ε gleichmäßig konvergenter Verfahren zu und untersuchen zunächst der Einfachheit halber den Fall $b = const > 0$ und $c = 0$. Ausgehend von einem äquidistanten Gitter Ω_h verwenden wir nun im Unterschied zu den bisher betrachteten polynomialen Finite-Elemente Räumen in einem gewöhnlichen Galerkin-Ansatz den Raum $\mathcal{V}_h$, aufgespannt von sogenannten *L-Splines* φ_i, die der Lösung exponentiell angepaßt sind. Sie werden für $i = 1, ..., N-1$ definiert über die lokalen homogenen Probleme

$$\begin{aligned} -\varepsilon\varphi_i'' + b\varphi_i' &= 0 \qquad \text{auf jedem offenen Teilintervall,} \\ \varphi_i(x_j) &= \delta_{ij} \qquad \text{in allen Gitterpunkten.} \end{aligned}$$

Wir möchten zeigen, daß man für ein damit gebildetes Finite-Elemente Verfahren eine bezüglich ε gleichmäßig geltende Fehlerabschätzung in der $||\cdot||_\varepsilon$-Norm erhält. Wir benutzen dazu die Interpolationsabschätzungen [102]

$$||u - u^I||_\infty \leq Ch \quad \text{und} \quad ||u - u^I||_\varepsilon \leq Ch^{\frac{1}{2}},$$

wobei die Interpolanten $u^I \in \mathcal{V}_h$ über die Bedingung $u^I(x_i) = u(x_i)$ in allen Gitterpunkten erklärt sind. Man benötigt ferner noch eine spezielle Abschätzung der $\mathcal{L}^1(\Omega)$-Norm der Ableitungen von Funktionen aus $\mathcal{V}_h$ durch deren $\mathcal{L}_2(\Omega)$-Norm:

$$||v_h'||_{\mathcal{L}^1} \leq Ch^{-\frac{1}{2}}\varepsilon^{\frac{1}{2}}||v_h'||.$$

Für einen Beweis verweisen wir ebenfalls auf [102]. Wegen $u^I - u_h \in \mathcal{V}_h$ folgt aus der gleichmässigen Koerzivitätsabschätzung (2.18) und der Orthogonalität des Fehlers

$$\begin{aligned} \alpha||u - u_h||_\varepsilon^2 \leq a(u - u_h, u - u_h) &= a(u - u_h, u - u^I) + a(u - u_h, u^I - u_h) \\ &= a(u - u_h, u - u^I) + 0, \end{aligned}$$

und es genügt, den Term $a(u - u_h, u - u^I)$ abzuschätzen. Wir erhalten für dessen diffusiven Anteil mit Hilfe der Cauchy–Schwarzschen Ungleichung sowie der zweiten Interpolationsabschätzung

$$\begin{aligned} \varepsilon|((u - u_h)', (u - u^I)')| &\leq \varepsilon^{\frac{1}{2}}|u - u_h|_1\varepsilon^{\frac{1}{2}}|u - u^I|_1 \\ &\leq C(c_1)\varepsilon|u - u^I|_1^2 + c_1\varepsilon|u - u_h|_1^2 \\ &\leq C(c_1)Ch + c_1||u - u_h||_\varepsilon^2. \end{aligned}$$

Mit partieller Integration folgt $(b(u-u^I)', u-u^I) = 0$. Daher ergibt sich für den konvektiven Anteil, indem wir zusätzlich die erste Interpolationsabschätzung sowie die Ungleichung für die Ableitungen verwenden,

$$\begin{aligned}
|(b(u-u_h)', u-u^I)| &= |(b(u^I-u_h)', u-u^I)| \\
&\leq C||(u^I-u_h)'||_{\mathcal{L}^1}||u-u^I||_\infty \\
&\leq Ch^{-\frac{1}{2}}\varepsilon^{\frac{1}{2}}||(u^I-u_h)'|| \cdot Ch \\
&\leq Ch^{\frac{1}{2}}\varepsilon^{\frac{1}{2}}(|u^I-u|_1 + |u-u_h|_1) \\
&\leq c_2\varepsilon|u-u_h|_1^2 + C(c_2)h + Ch^{\frac{1}{2}}\varepsilon^{\frac{1}{2}}|u^I-u|_1 \\
&\leq c_2||u-u_h||_\varepsilon^2 + C(c_2)h + Ch^{\frac{1}{2}}h^{\frac{1}{2}}.
\end{aligned}$$

Durch Wahl hinreichend kleiner Konstanten c_1 und c_2, was unabhängig von ε möglich ist, erhalten wir mit Hilfe der soeben gezeigten Abschätzungen die Fehlerabschätzung

$$||u-u_h||_\varepsilon \leq Ch^{\frac{1}{2}} \quad \text{mit } C \neq C(h,\varepsilon). \tag{2.21}$$

Im allgemeinen Fall $b \neq const$ und $c \neq 0$ approximiert man die Funktionen b, c und f auf den Teilintervallen, etwa durch $\bar{b} := [b(x_{i-1}) + b(x_i)]/2$ auf $]x_{i-1}, x_i[$, und erhält eine modifizierte Bilinearform $\bar{a}_h(\cdot\,,\cdot)$ zu einem modifizierten Differentialoperator mit „eingefrorenen" Koeffizienten $\bar{L}$ und somit auf analoge Weise als Ansatzfunktionen $\bar{L}$-Splines. Man kann dann Abschätzung (2.21) auch im allgemeinen Fall zeigen. Es gilt

Satz 4 (Gleichmässige Fehlerabschätzung für die $||\cdot||_\varepsilon$-Norm)
Für ein Galerkin-Verfahren bezüglich der modifizierten Bilinearform $\bar{a}_h(\cdot\,,\cdot)$ und rechten Seite $\bar{f}$, das mittels $\bar{L}$-Splines gebildet wird, gilt hinsichtlich der $||\cdot||_\varepsilon$-Norm die optimale Fehlerabschätzung

$$||u-u_h||_\varepsilon \leq Ch^{\frac{1}{2}} \quad \text{mit } C \neq C(h,\varepsilon).$$

Beweis: Siehe [102]

□

Um ein in der diskreten Maximumsnorm $||\cdot||_{\infty,d}$ bezüglich ε gleichmäßig konvergentes Finite-Elemente Verfahren zu entwickeln, ist es günstig, einen Testraum zu wählen, der die *diskreten Greenschen Funktionen* zu dem formal adjungierten Problem enthält. Definiert man für $i = 1, ..., N-1$ die diskrete Greensche Funktion $G_i(x)$ zum Gitterpunkt x_i durch

$$\bar{a}_h(w, G_i) = w(x_i) \quad \forall w \in \mathcal{H}_0^1(\Omega),$$

so erhält man unter der Annahme $G_i \in \mathcal{S}_h$ als Darstellung der Fehler in den Gitterpunkten

$$(u-u_h)(x_i) = \bar{a}_h(u-u_h, G_i) = (\bar{a}_h - a)(u, G_i) + (f-\bar{f}, G_i). \tag{2.22}$$

Wir machen im folgenden einen Petrov–Galerkin-Ansatz, wobei für die Basisfunktionen des Ansatzraums lediglich $\varphi_i(x_j) = \delta_{i,j}$ in allen Gitterpunkten gelte. Man kann zeigen, daß die G_i in dem Raum enthalten sind, der durch $\bar{L}^*$-Splines ψ_i für $i = 1, ..., N-1$ aufgespannt wird. Diese werden definiert über die lokalen homogenen adjungierten Probleme

$$\begin{aligned}
-\varepsilon\psi_i'' - \bar{b}\psi_i' + \bar{c}\psi_i &= 0 \qquad \text{auf jedem offenen Teilintervall,} \\
\psi_i(x_j) &= \delta_{ij} \qquad \text{in allen Gitterpunkten.}
\end{aligned}$$

Den Beweis der folgenden Fehlerabschätzung erhält man nunmehr aus der Darstellung des Fehlers (2.22) zusammen mit den bezüglich ε gleichmäßig geltenden a priori Abschätzungen $||u||_\infty \leq C$, $||G_i||_\infty \leq C$ sowie $||u'||_{\mathcal{L}^1} \leq C$ [102].

Satz 5 (Gleichmässige Fehlerabschätzung für die $||\cdot||_{\infty,d}$-Norm)
Es seien $\bar{b}, \bar{c}$ und $\bar{f}$ Approximationen der Ordnung $\mathcal{O}(h)$ an b, c und f. Für eine Petrov–Galerkin-Diskretisierung bezüglich der modifizierten Bilinearform $\bar{a}_h(\cdot\,,\cdot)$ mit $\bar{L}^*$-Splines als Testfunktionen gilt in der diskreten Maximumsnorm die Abschätzung

$$||u - u_h||_{\infty,d} \leq Ch \quad \text{mit } C \neq C(h, \varepsilon).$$

Die zugehörige Steifigkeitsmatrix ist unter den gemachten Voraussetzungen unabhängig von der Wahl der Ansatzfunktionen.

Beweis: Siehe [102]

□

Die Steifigkeitsmatrix eines mit $\bar{L}$-Splines als Ansatzfunktionen gebildeten Petrov–Galerkin-Verfahrens ist umgekehrt unabhängig von der Wahl der Testfunktionen. (Allerdings sind die Diskretisierungen der rechten Seite durchaus verschieden.) Deshalb gilt obiges Resultat auch für den ersten Finite-Elemente Ansatz dieses Abschnitts. Verwendet man $\bar{L}$-Splines als Ansatzfunktionen, so kann man die gleichen Abschätzungen auch bezüglich der uniformen Maximumsnorm $||\cdot||_\infty$ zeigen [102].

2.2.3 Upwind-Verfahren in 2D

Wir betrachten das zweidimensionale Konvektions-Diffusions Problem (2.12), (2.13) mit hinreichend glatten Koeffizientenfunktionen $\vec{b}, c$ und rechter Seite f und nehmen weiter an, daß $c(x,y) \geq 0$ sowie $c(x,y) - \frac{1}{2}\nabla \cdot \vec{b}(x,y) \geq \underline{c} > 0$ auf $\bar{\Omega}$ gelten. Definieren wir als Lösungsraum $\mathcal{V}$ den Sobolevraum $\mathcal{H}_0^1(\Omega)$ und diesbezüglich die Bilinearform

$$a(v,w) := \varepsilon(\nabla v, \nabla w) + (\vec{b} \cdot \nabla v, w) + (cv, w) \tag{2.23}$$

für $v, w \in \mathcal{V}$, so erhalten wir wie im Eindimensionalen eine schwache Formulierung des Randwertproblems gemäß (2.16). Als erfolgreiche Upwind Finite-Elemente Ansätze hierfür erweisen sich beispielsweise Verfahren, die ausgehen von einer *Dreieckszerlegung vom schwach spitzen Typ* und eine *Sekundärzerlegung* verwenden. Die so entstehenden invers monotonen Verfahren konvergieren bezüglich der ε-gewichteten $\mathcal{H}^1(\Omega)$-Norm in erster Ordnung [60, 102]. Strebt man Verfahren höherer Ordnung an, so ist es unrealistisch zu versuchen, die inverse Monotonie des diskreten Problems aufrecht zu erhalten. Schon für die Diskretisierung des gewöhnlichen Laplace-Operators mit quadratischen Finiten Elementen kann nämlich die inverse Monotonie nur noch in speziellen geometrischen Situationen nachgewiesen werden. Die *Stromliniendiffusions-Methode*, die auf dem Hinzuaddieren künstlicher Diffusion zum ursprünglichen Problem in Richtung der Subcharakteristiken beruht, führt unter Aufgabe der inversen Monotonie der Diskretisierung zu Upwind-Verfahren höherer Ordnung. Wir wollen ihre Grundidee im folgenden kurz skizzieren.

Es sei $\mathcal{V}_h := \{v_h \in \mathcal{V} : v_h|_T \in \mathcal{P}_k(T) \quad \forall T \in \mathcal{T}_h\} \subset \mathcal{V}$ ein konformer Finite-Elemente Raum bestehend aus stückweise polynomialen, stetigen Funktionen zu einer regulären, quasiuniformen Triangulierung $\mathcal{T}_h$ von Ω. Hierbei ist $\mathcal{P}_k(T)$ der Raum der Polynome vom Grad k auf

dem Element $T \in \mathcal{T}_h$. Die Grundidee der Stromliniendiffusions-Methode beruht auf der Annahme, daß die kontinuierliche Lösung u lokal regulär ist und auf den einzelnen Elementen $T \in \mathcal{T}_h$ im Sinne von $\mathcal{L}_2(T)$ der Differentialgleichung genügt:

$$-\varepsilon \triangle u + \vec{b} \cdot \nabla u + cu = f \quad \forall T \in \mathcal{T}_h.$$

Diese lokalen Ausdrücke werden nun mit dem Ergebnis der Anwendung des konvektiven Anteils auf die Basisfunktionen getestet und mit elementweisen Upwind-Parametern δ_T gewichtet zur gewöhnlichen schwachen Form (2.23) aufaddiert. Man erhält mit

$$\begin{aligned} a_h^{SD}(v,w) &:= a(v,w) + \sum_{T \in \mathcal{T}_h} \delta_T(-\varepsilon \triangle v + \vec{b} \cdot \nabla v + cv, \vec{b} \cdot \nabla w)_T, \\ f_h^{SD}(w) &:= (f,w) + \sum_{T \in \mathcal{T}_h} \delta_T(f, \vec{b} \cdot \nabla w)_T, \end{aligned} \tag{2.24}$$

daß für die (hinreichend glatt vorausgesetzte) Lösung u dann $a_h^{SD}(u, v_h) = f_h^{SD}(v_h)$ für alle $v_h \in \mathcal{V}_h$ gilt. Wir definieren daher:

Schema 7 (Stromliniendiffusions-Methoden)
Als Stromliniendiffusions-Methoden bezeichnet man Verfahren, die sich mittels der diskreten schwachen Formulierung

$$\text{Finde } u_h \in \mathcal{V}_h : \quad a_h^{SD}(u_h, v_h) = f_h^{SD}(v_h) \quad \forall v_h \in \mathcal{V}_h \tag{2.25}$$

ergeben.

Eine Stromliniendiffusions-Methode kann auch als Petrov–Galerkin-Verfahren angesehen werden, wenn man als Testfunktionen die in Stromrichtung modifizierten Basisfunktionen $w_h + \delta_T \vec{b} \cdot \nabla w_h$ verwendet. Für stückweise lineare Elemente ($\Rightarrow \triangle v_h|_T = 0$) schreibt sich die Bilinearform (2.24) unter den Annahmen $\vec{b} = (b_1, b_2) = const$ und $c = 0$ mit $b = |\vec{b}|$ und dem Gradienten in Stromrichtung $\partial_{\vec{b}} := \frac{\vec{b}}{b} \cdot \nabla$ als

$$a_h^{SD}(v_h, w_h) = \varepsilon(\nabla v_h, \nabla w_h) + \sum_{T \in \mathcal{T}_h} (\delta_T b^2 (\partial_{\vec{b}} v_h, \partial_{\vec{b}} w_h)_T + (\vec{b} \cdot \nabla v_h, w_h),$$

was man als lokales Hinzufügen von zusätzlicher Diffusion der Stärke $\delta_T b^2$ in Stromrichtung interpretieren kann; daher der Name *Stromliniendiffusions-Methode*. Eine im Vergleich zur ε-gewichteten $\mathcal{H}^1(\Omega)$-Norm stärkere Norm, die eine Kontrolle des Gradienten in Stromrichtung ermöglich, ist auf der Basis der Bilinearform (2.24) gegeben durch

$$||v||_{SD} := \left[\varepsilon |v|_1^2 + c_0 ||v||^2 \sum_{T \in \mathcal{T}_h} \delta_T ||b \cdot \nabla v||_{\mathcal{L}_2(T)}^2 \right]^{\frac{1}{2}}.$$

Durch geeignete Wahl der Stromliniendiffusions-Paramter δ_T kann man dafür die Fehlerabschätzung

$$||u - u_h||_{SD} \leq C(\varepsilon^{\frac{1}{2}} + h^{\frac{1}{2}}) h^k |u|_{k+1} \tag{2.26}$$

zeigen. Im Fall dominanter Konvektion sieht man hieraus durch den Vergleich mit entsprechenden Interpolationsabschätzungen, daß der $\mathcal{L}_2(\Omega)$-Fehler des Gradienten der diskreten Näherungslösung in Stromrichtung optimal, der $\mathcal{L}_2(\Omega)$-Fehler selbst jedoch um die Ordnung $\frac{1}{2}$ kleiner als optimal ist. Abschätzung (2.26) sichert allerdings nicht die bezüglich ε gleichmäßige Konvergenz des Verfahrens, da die Halbnorm $|u|_{k+1}$ im allgemeinen von negativen Potenzen von ε abhängt.

2.2.4 Gleichmäßig konvergente Verfahren in 2D

Die Konstruktion von nachweislich bezüglich ε gleichmäßig konvergenten Finite-Elemente Verfahren ist für zweidimensionale Konvektions-Diffusions Probleme auf der Basis von Standardgittern bislang nur in speziellen Fällen gelungen. Einen Überblick der Entwicklungen findet man in [39, 71, 102]. Solche gleichmäßig konvergenten Verfahren werden zumeist über einen Galerkin- oder Petrov–Galerkin-Ansatz mittels partiell oder vollständig gefitteter Basisfunktionen gebildet, die sich über Tensorprodukte eindimensionaler exponentiell angepaßter Splines auf einem quadratischen Tensorprodukt-Gitter berechnen lassen.

Wir betrachten im folgenden wieder die Randwertaufgabe (2.12), (2.13), nun unter den Voraussetzungen $\vec{b}(x,y) = (b_1(x), b_2(y)) \geq (\underline{b}_1, \underline{b}_2) > (0,0)$, $c(x,y) - \frac{1}{2}\nabla \cdot \vec{b}(x,y) \geq \underline{c} > 0$ sowie $c(x,y) \geq 0$ auf $\bar{\Omega}$. Damit ist eine additive Zerlegung des Konvektions-Diffusions Anteils von L in die Richtungsanteile $L_x^{KD} := -\varepsilon\partial_{xx} + b_1(x)\partial_x$ und $L_y^{KD} := -\varepsilon\partial_{yy} + b_2(y)\partial_y$ formal möglich. Man spricht in diesem Zusammenhang daher auch von *separabler* Konvektion. Die Koeffizientenfunktionen und die rechte Seite seien ferner glatt, und f genüge zusätzlich der Kompatibilitätsbedingung $f(0,0) = f(1,0) = f(0,1) = f(1,1) = 0$. Es sei $\Omega_h := \{(x_i^h, y_j^h) \in \Omega : x_i^h := ih \text{ und } y_j^h := jh \text{ für } 0 < i,j < 1/h\}$ ein Tensorprodukt-Gitter mit Maschenweite h. Eine Möglichkeit zur Konstruktion eines bezüglich ε gleichmäßig konvergenten Verfahrens besteht darin, nach stückweise konstanter Approximation der Funktionen $b_1(x)$ und $b_2(y)$ als Ansatz- und Testfunktionen die Tensorprodukte eindimensionaler $\bar{L}$-Splines φ_i^x und φ_j^y der jeweiligen Richtungsanteile L_x^{KD} und L_y^{KD} des Konvektions-Diffusions Anteils von L zu wählen. Der Raum $\mathcal{V}_h$, der von den so definierten exponentiell angepaßten Basisfunktionen $\varphi_{(i,j)}(x,y) := \varphi_i^x(x)\varphi_j^y(y)$ aufgespannt wird, besteht dann aus stetigen Funktionen und wir erhalten ein konformes Finite-Elemente Verfahren. Man kann zeigen, daß für das über die Bilinearform $\bar{a}_h(v,w) := \int_\Omega [\varepsilon\nabla v \cdot \nabla w + (\vec{b} \cdot \nabla v)w + \bar{c}vw]\, dxdy$ mit einer entsprechenden Approximation von $c(x,y)$ und einer Quadratur der rechten Seite entstehende Galerkin-Verfahren die Fehlerabschätzung (2.21) wie im eindimensionalen Fall gilt. Der Beweis verläuft analog zum eindimensionalen Fall. Für den Nachweis der entsprechenden Interpolations- und Gradientenabschätzungen benötigt man jedoch die oben vorausgesetzte Kompatibilitätsforderung an f. Galerkin-Verfahren, die Tensorprodukte von eindimensionalen $\bar{L}^*$-Splines verwenden, führen zu ähnlichen oder je nach Quadraturformel sogar identischen Diskretisierungen und damit ebenfalls zu gleichmäßig konvergenten Methoden. In den numerischen Beispielen in [71] sieht man, daß Petrov–Galerkin-Verfahren, die bilineare Funktionen als Ansatz- beziehungsweise Testfunktionen sowie Tensorprodukte eindimensionaler $\bar{L}^*$-Splines als Test- respektive eindimensionaler $\bar{L}$-Splines als Ansatzfunktionen benutzen, ebenfalls erfolgreich und dem Galerkin-Verfahren sogar überlegen sind. Überraschenderweise führt jedoch die Wahl von Ansatzfunktionen mittels $\bar{L}$-Splines zusammen mit der Wahl von Testfunktionen über $\bar{L}^*$-Splines zu Instabilitäten.

Im Fall nichtseparabler Konvektion ergeben sich in der Regel nichtkonforme Verfahren, da die über Tensorprodukte gebildeten Funktionen $\varphi_{(i,j)}$ dann entlang der Zellgrenzen im Innern ihres Trägers im allgemeinen unstetig sind. Unter sehr restriktiven Voraussetzungen an $\vec{b}$, so daß man die gleichen a priori Abschätzungen für die partiellen Ableitungen der kontinuierlichen Lösungen wie im separablen Fall erhält, gelten die gleichen Fehlerabschätzungen wie dort [102].

Kapitel 3

Multiskalen-Transformation linearer Gleichungssysteme

Wie wir im vorangehenden Kapitel 2 gesehen haben, führen die kontinuierlichen Konvektions-Diffusions Probleme nach einer geeigneten Diskretisierung, zum Beispiel durch einen Petrov–Galerkin-Ansatz mit einem Finite-Elemente Ansatzraum $\mathcal{V}_0$ und Testraum $\mathcal{S}_0$ bezüglich eines Gitters Ω_0, zu großen schwachbesetzten linearen Gleichungssystemen der Form

$$T_0 u_0 = f_0. \tag{3.1}$$

Wir gehen nun also zu einer skalenbehafteten Indizierung über. Wir interessieren uns für Multiskalen-Verfahren zur effizienten Lösung der Systeme (3.1) und untersuchen dazu in diesem Kapitel zunächst, wie sich solche linearen Gleichungssysteme unter den Multiskalen-Zerlegungen der Tiefe lt

$$\mathcal{V}_0 = \bigoplus_{k=1}^{lt} \mathcal{W}_k \oplus \mathcal{V}_{lt}, \tag{3.2}$$

$$\mathcal{S}_0 = \bigoplus_{k=1}^{lt} \mathcal{F}_k \oplus \mathcal{S}_{lt} \tag{3.3}$$

transformieren. Es wird dabei vorausgesetzt, daß die Zerlegungen (3.2) und (3.3) rekursiv für $k = 1, ..., lt$ über Zweiskalen-Zerlegungen $\mathcal{V}_{k-1} = \mathcal{W}_k \oplus \mathcal{V}_k$ und $\mathcal{S}_{k-1} = \mathcal{F}_k \oplus \mathcal{S}_k$ der Räume $\mathcal{V}_{k-1}$ und $\mathcal{S}_{k-1}$ in Unterräume mit Detail- und grobskaligen Anteilen entstehen. Entgegen der zum Beispiel bei geometrischen Mehrgitter-Verfahren üblichen Notation [62] werden die gröberen Räume mit größeren Indizes, die feineren Räume mit kleineren Indizes versehen.

Nach einer genauen Untersuchung der Darstellung des diskretisierten Problems auf dem feinsten Level $k = 0$ studieren wir zunächst das Transformationsverhalten von (3.1) im Fall der einfachen Zerlegungen $\mathcal{V}_0 = \mathcal{W}_1 \oplus \mathcal{V}_1$ und $\mathcal{S}_0 = \mathcal{F}_1 \oplus \mathcal{S}_1$, für die $lt = 1$ ist. Ist die Zerlegungstiefe $lt > 1$, so wird eine natürliche Darstellung von (3.1) relativ zu den Zerlegungen (3.2) und (3.3) durch den zugehörigen Basiswechsel induziert. Man erhält als Systemmatrix des entsprechend transformierten linearen Gleichungssystems die sogenannte *Standardform* von T_0 [19]. Aus algorithmischer Sicht ist das Arbeiten im Standardsystem, das heißt ausgehend von der Multiskalen-Basis und der Standardform, meist nicht zu empfehlen. Dies gilt insbesondere, wenn man an multiplikativen Verfahren zur Lösung von (3.1) interessiert ist.

Die sogenannte *Nichtstandardform* ist eine alternative Multiskalen-Darstellung des Feingitteroperators, die über den rekursiven Aufbau der Zerlegungen (3.2) und (3.3) definiert ist. Mit Hilfe der Nichtstandardform lassen sich effiziente Algorithmen im betrachteten Multiskalen-Kontext formulieren [19, 48, 49], die äquivalent zu entsprechenden mittels der Standardform gebildeten Verfahren sind. Dies ist jedoch erst Gegenstand von Kapitel 4. Im letzten Unterkapitel des jetzigen Kapitels zeigen wir, daß die zur Standardform gehörenden Multiskalen-Transformationen jeweils auch als Produkte von Erzeugendensystem-Transformationen mit geeigneten Blockdiagonalmatrizen darstellbar sind. Die Erzeugendensystem-Transformationen gehen dabei aus einer Multilevel-nodalen Darstellung hervor, die zu einer nichtklassischen Mehrgitter-Sichtweise führt [55, 56]. Dies erlaubt es, die von uns vorgeschlagenen Multiskalen-Methoden auch auf der Basis der dort vorgestellten Erzeugendensysteme zu interpretieren. Wir zeigen schließlich, daß es möglich ist, die Standard- und Nichtstandardform eines Operators sowie auch seine klassische Erzeugendensystem-Darstellung in eine allgemeinere Multiskalen-Erzeugendensystem-Darstellung einzubetten.

Erfolgt die Diskretisierung der kontinuierlichen Randwertaufgabe durch eine Finite-Differenzen oder Finite-Volumen Methode, so kann T_0 im allgemeinen nicht mehr direkt als Steifigkeitsmatrix interpretiert werden, die aus einem Petrov–Galerkin-Verfahren stammt. Die zu den Zerlegungen (3.2) und(3.3) gehörenden Multiskalen-Transformationen sind dann als algebraische Transformationen der Ansatz- und Testseite anzusehen. Sind die Diskretisierung und die Räume $\mathcal{V}_0$ und $\mathcal{S}_0$ jedoch miteinander verträglich, so läßt sich T_0 als Petrov–Galerkin-Diskretisierung eines leicht gestörten kontinuierlichen Operators interpretieren, und die algebraischen Transformationen erhalten über diese Interpretation dann einen Sinn. Wir sprechen in diesen Fällen etwas vereinfachend ebenfalls von T_0 als der Steifigkeitsmatrix zu den Ansatz- und Testräumen $\mathcal{V}_0$ und $\mathcal{S}_0$.

Im folgenden werden wir die entsprechenden Definitionen und Begriffe zunächst für den Galerkin-Fall einführen, bei dem wir davon ausgehen, daß sowohl Ansatz- und Testraum als auch deren Zerlegungen gleich sind. Die Verallgemeinerung auf den Petrov–Galerkin-Fall ist danach einfach zu bewerkstelligen. Wir werden dabei durch untere Indizes $_\mathcal{V}$ und $_\mathcal{S}$ ansonsten gleichbezeichnete Größen der Ansatz- und Testseite unterscheiden. Geht aus dem Zusammenhang klar hervor, auf welche der beiden Seiten sich ein Ausdruck bezieht, so wird bisweilen auf diese Kennzeichnung verzichtet. Zur Darstellung von (Block-) Vektoren und (Block-) Matrizen verwenden wir gelegentlich die dafür im Programmiersystem Matlab vorgesehene Notation $[\cdot\,,\cdot]$ für die Zusammenfassung passender (Block-) Zeilen sowie $[\cdot\,;\cdot]$ für passende (Block-) Spalten.

3.1 Feingitterdarstellungen des diskretisierten Problems

3.1.1 Galerkin-Ansatz

Es sei $\mathcal{V}_0 \subset \mathcal{H}_0^1(\Omega)$ ein endlichdimensionaler Funktionenraum, der sowohl als Ansatz- als auch als Testraum zur Diskretisierung des betrachteten Randwertproblems mittels eines Finite-Elemente Ansatzes diene. Wir definieren unterschiedliche vektorielle Darstellungsmöglichkeiten von Funktionen aus $\mathcal{V}_0$ sowie die Massen- und Steifigkeitsmatrix, die bei der Herleitung des linearen Gleichungssystems (3.1) aus der schwachen Formulierung auftreten.
Es sei dazu $\Omega_0 := \{x_i^0 \in \Omega : i = 1, ..., N_0\}$ ($N_0 = \dim(\mathcal{V}_0)$) die Menge der Gitterpunkte eines feinen Gitters (dieses werde ebenfalls mit Ω_0 bezeichnet), in denen nodale Basisfunktionen der Basis $\{\varphi_{i,\mathcal{V}}^0 : i = 1, ..., N_0\}$ des Ansatz- und Testraums $\mathcal{V}_0$ verankert seien. Der Begriff der

nodalen Basis wird hier etwas allgemeiner aufgefaßt als es sonst üblich ist [22]: Wir nehmen an, daß es eine eineindeutige Zuordnung zwischen der Menge der Gitterpunkte und der Menge der Basisfunktionen gibt. Wir fordern weiter, daß die Basisfunktionen $\varphi^0_{i,\mathcal{V}}$ in dem Sinne lokal sind, daß ihre Träger jeweils in einem kleinen Bereich um diejenigen Gitterpunkte zentriert sind, denen sie zugehören. Die Knotenbasisdarstellung der Lösung $u_0(x) = \sum_{i=1}^{N_0} u_0(x_i^0)\varphi^0_{i,\mathcal{V}}(x)$ des diskretisierten Problems ist gegeben durch den Vektor

$$u_0^{KB_\mathcal{V}} := (u_0(x_i^0))_{i=1,\ldots,N_0},$$

das heißt, man identifiziert $u_0 \in \mathcal{V}_0$ mit dem Vektor $u_0^{KB_\mathcal{V}} \in \mathbb{R}^{N_0}$. Die Momentendarstellung von f_0, der $\mathcal{L}_2(\Omega)$-Projektion der rechten Seite f nach $\mathcal{V}_0$, ist bezüglich obiger Basis definiert durch den Vektor

$$f_0^{MOM_\mathcal{V}} := ((f, \varphi^0_{i,\mathcal{V}}))_{i=1,\ldots,N_0}.$$

Der Zusammenhang zwischen Momentendarstellung und nodaler Darstellung wird durch die Massenmatrix

$$M_0^{KB_\mathcal{V}\to MOM_\mathcal{V}} := ((\varphi^0_{j,\mathcal{V}}, \varphi^0_{i,\mathcal{V}}))_{i,j}$$

hergestellt. Wie man leicht nachrechnet, gilt

$$u_0^{MOM_\mathcal{V}} = M_0^{KB_\mathcal{V}\to MOM_\mathcal{V}} u_0^{KB_\mathcal{V}}.$$

Die Massenmatrix transformiert als Matrixdarstellung der Identität über die schwache Formulierung die nodale Darstellung einer Funktion in die zugehörige Momentendarstellung. Die Steifigkeitsmatrix ist definiert als

$$T_0^{KB_\mathcal{V}\to MOM_\mathcal{V}} := (a(\varphi^0_{j,\mathcal{V}}, \varphi^0_{i,\mathcal{V}}))_{i,j}.$$

Durch Einsetzen von u_0 in eine entsprechende diskrete schwache Form, die mittels $\mathcal{V}_0$ gebildet wird, erhalten wir dann wegen

$$a(\sum_{j=1}^{N_0} u_0(x_j^0)\varphi^0_{j,\mathcal{V}}, \varphi^0_{i,\mathcal{V}}) = (f, \varphi^0_{i,\mathcal{V}}), \quad (i = 1, ..., N_0),$$

das lineare Gleichungssystem

$$T_0^{KB_\mathcal{V}\to MOM_\mathcal{V}} u_0^{KB_\mathcal{V}} = f_0^{MOM_\mathcal{V}}. \tag{3.4}$$

3.1.2 Petrov–Galerkin-Ansatz

Bei einem Petrov–Galerkin-Verfahren zur Diskretisierung des betrachteten Randwertproblems verwendet man im allgemeinen einen vom Ansatzraum $\mathcal{V}_0 \subset \mathcal{H}_0^1(\Omega)$ verschiedenen Testraum $\mathcal{S}_0$ zum Aufstellen der diskreten schwachen Form und des linearen Gleichungssystems. Es werde $\mathcal{S}_0 \subset \mathcal{L}_2(\Omega)$ ebenfalls von nodalen Basisfunktionen $\{\varphi^0_{i,\mathcal{S}} : i = 1, ..., N_0\}$ aufgespannt. Die Steifigkeitsmatrix wird in diesem Fall definiert durch

$$T_0^{KB_\mathcal{V}\to MOM_\mathcal{S}} := (a(\varphi^0_{j,\mathcal{V}}, \varphi^0_{i,\mathcal{S}}))_{i,j}.$$

Man gewinnt $f_0 \in \mathcal{S}_0$ als die $\mathcal{L}_2(\Omega)$-Projektion der rechten Seite f auf den Testraum. Durch Einsetzen von u_0 in eine entsprechende diskrete schwache Form, die diesmal mittels $\mathcal{V}_0$ und $\mathcal{S}_0$ gebildet wird, erhalten wir dann wegen

$$a(\sum_{j=1}^{N_0} u_0(x_j^0)\varphi^0_{j,\mathcal{V}}, \varphi^0_{i,\mathcal{S}}) = (f, \varphi^0_{i,\mathcal{S}}), \quad (i = 1, ..., N_0),$$

das lineare Gleichungssystem

$$T_0^{KB_\mathcal{V} \to MOM_\mathcal{S}} u_0^{KB_\mathcal{V}} = f_0^{MOM_\mathcal{S}}. \tag{3.5}$$

Wir verzichten im weiteren Verlauf meist auf die mühevolle Darstellungskennzeichnung bei Vektoren sowie die Abbildungskennzeichnung bei Matrizen. Bisweilen identifizieren wir auch Funktionen und Abbildungen endlichdimensionaler Funktionenräume mit ihren Darstellungen durch Vektoren und Matrizen.

3.2 Zweiskalen-Darstellungen des diskretisierten Problems

3.2.1 Galerkin-artige Zweiskalen-Zerlegung

Wir untersuchen in diesem Abschnitt, wie sich das lineare Gleichungssystem (3.4) im Fall der einfachen Zerlegung des Ansatz- und Testraums

$$\mathcal{V}_0 = \mathcal{W}_1 \oplus \mathcal{V}_1, \tag{3.6}$$

transformiert. Man spricht auch von einer Zweiskalen-Zerlegung, da insgesamt zwei Skalen daran beteiligt sind. Hierbei sei $\mathcal{V}_1$ ein Unterraum, in dem grobskaligere Funktionen von $\mathcal{V}_0$ liegen. Wir nehmen an, daß diese mit Hilfe einer weiteren nodalen Basis dargestellt werden können, die bezüglich einem gröberen Gitter $\Omega_1 \subset \Omega_0$ definiert ist. Der Komplementraum $\mathcal{W}_1$ umfaßt dann Detail-Informationen, die beim Übergang von der feineren zur nächstgröberen Skala vernachlässigt werden. Wir setzen im folgenden voraus, daß sowohl für die Räume grobskaliger Anteile als auch für die Komplementräume Basen existieren, die in den Gitterpunkten des gröberen sowie in den Fein-ohne-Grobgitterpunkten $\Omega_0^f := \Omega_0 \backslash \Omega_1$ des feineren Gitters lokalisiert sind:

$$\mathcal{W}_1 = \operatorname{span}\{\psi_{i,\mathcal{V}}^1 : x_i^0 \in \Omega_0^f\}, \quad \mathcal{V}_1 = \operatorname{span}\{\varphi_{i,\mathcal{V}}^1 : x_i^0 \in \Omega_1\}. \tag{3.7}$$

Zwei Beispiele für solche Zerlegungen sind die einfachen Zerlegungen des Raums der stückweise linearen Funktionen auf $\Omega =]0,1[$ mit homogenen Randwerten mittels der hierarchischen Basis [129] und mittels linearer Prewavelets [91]. In beiden Fällen werden die Räume $\mathcal{V}_0$ und $\mathcal{V}_1$ aus skalierten und dilatierten linearen Splines aufgespannt, den bekannten Hut-Funktionen, die man sich in den Fein- beziehungsweise Grobgitterpunkten eines uniform verfeinerten Gitters verheftet denken kann. Die Komplementräume $\mathcal{W}_1$ werden im Fall der hierarchischen Zerlegung von Feingitter-Splines und im Prewavelet-Fall von linearen Prewavelets aufgespannt, die beide Male in den Fein-ohne-Grobgitterpunkten des feinen Gitters verankert sind. Abbildung 3.1 zeigt auf der linken Seite eine einfache hierarchische und auf der rechten Seite eine Prewavelet-Zerlegung auf $\Omega =]0,1[$. Die oberen beiden Bilder zeigen sämtliche Basisfunktionen des feineren Raums $\mathcal{V}_0$ zu einem uniformen Gitter mit Maschenweite $h_0 = \frac{1}{8}$, die unteren beiden Bilder die Basisfunktionen der nächstgrobskaligeren Räume, die für beide Zerlegungen übereinstimmen. In den mittleren Bildern werden jeweils die Basisfunktionen der unterschiedlichen Komplementräume dargestellt. Zur besseren Unterscheidung werden benachbarte Basisfunktionen im Wechsel mit durchgezogenen und unterbrochenen Linien gezeichnet.

Aufgrund des Zweiskalen-Ansatzes und der Inklusionen $\mathcal{V}_1 \subset \mathcal{V}_0$ und $\mathcal{W}_1 \subset \mathcal{V}_0$ existieren nun rechteckige Matrizen $P_{1,\mathcal{V}}^0$ und $Q_{1,\mathcal{V}}^0$, deren Spalten angeben, wie sich die Basisfunktionen $\varphi_{i,\mathcal{V}}^1$ von $\mathcal{V}_1$ und $\psi_{i,\mathcal{V}}^1$ von $\mathcal{W}_1$ gemäß den Multiskalen-Gesetzen, das heißt mittels *Skalierungs-*

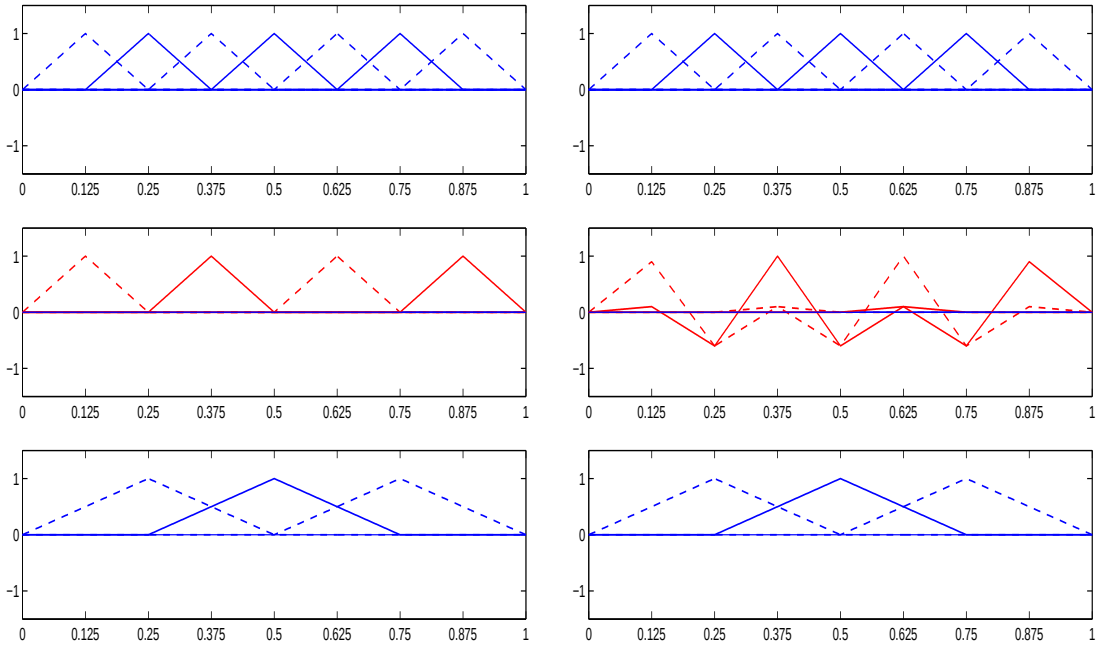


ABBILDUNG 3.1: *Einfache klassische hierarchische Basis Zerlegung (links) sowie Prewavelet-Zerlegung (rechts) auf* $\Omega =]0,1[$*, obere Bilder* $\mathcal{V}_0$*, mittlere Bilder* $\mathcal{W}_1$*, untere Bilder* $\mathcal{V}_1$*,* $h_0 = \frac{1}{8}$.

und *Waveletgleichungen,* aus solchen von $\mathcal{V}_0$ berechnen lassen [19]. Die zusammengesetzte schwachbesetzte quadratische Matrix $W^0_{1,\mathcal{V}} := [Q^0_{1,\mathcal{V}}, P^0_{1,\mathcal{V}}]$ beschreibt dann den Basiswechsel von einer Zweiskalen-Darstellung einer Funktion $u_0 \in \mathcal{V}_0$ bezüglich der direkten Zerlegung $\mathcal{W}_1 \oplus \mathcal{V}_1$ mit der Basis

$$\{\theta^1_{i,\mathcal{V}} : i = 1, ..., N_0\} := \{\psi^1_{i,\mathcal{V}} : x^0_i \in \Omega_0 \backslash \Omega_1\} \cup \{\varphi^1_{i,\mathcal{V}} : x^0_i \in \Omega_1\}$$

in die nodale Feingitterdarstellung bezüglich $\mathcal{V}_0$. Die Matrix $P^0_{1,\mathcal{V}}$ wird im Kontext von Mehrgitter-Verfahren auch als *Prolongation* bezeichnet und beschreibt eine Grobgitterfunktion des Levels 1 im nächstfeineren Raum der Stufe 0 [62]. Die Matrix $Q^0_{1,\mathcal{V}}$ tritt bei klassischen Mehrgitter-Verfahren nicht explizit in Erscheinung. Sie kann ebenfalls als Prolongation angesehen werden, die Basisfunktionen des Komplementraums $\mathcal{W}_1$ auf dem feineren Gitter darstellt. Der inverse Basiswechsel, der die nodale Feingitterdarstellung einer Funktion $u_0 \in \mathcal{V}_0$ in ihre Darstellung bezüglich der direkten Zerlegung $\mathcal{W}_1 \oplus \mathcal{V}_1$ überführt, ist als $W^1_{0,\mathcal{V}} := [Q^1_{0,\mathcal{V}}; P^1_{0,\mathcal{V}}] := (W^0_{1,\mathcal{V}})^{-1}$ gegeben durch rechteckige Matrizen $P^1_{0,\mathcal{V}}$ und $Q^1_{0,\mathcal{V}}$ mit im Vergleich zu $P^0_{1,\mathcal{V}}$ und $Q^0_{1,\mathcal{V}}$ transponierter Gestalt. (Man beachte die unterschiedliche Indizierung!) Im Gegensatz zu $P^0_{1,\mathcal{V}}$ und $Q^0_{1,\mathcal{V}}$ sind $P^1_{0,\mathcal{V}}$ und $Q^1_{0,\mathcal{V}}$ im allgemeinen jedoch nicht mehr schwachbesetzt. Eine fundamentale Erkenntnis ist, daß die Matrizen $P^1_{0,\mathcal{V}}$ und $Q^1_{0,\mathcal{V}}$ ähnlich wie im Fall der klassischen hierarchischen Basis Verfahren [16] nicht explizit zur Implementierung unserer Multiskalen-Verfahren gebraucht werden. In den Abbildungen 3.2 und 3.3 sind die Zweiskalen-Transformationen $W^0_{1,\mathcal{V}}$ und ihre Inversen $W^1_{0,\mathcal{V}}$ für den Fall einer eindimensionalen sowie zweidimensionalen Prewavelet-Zerlegung zu sehen. Die zweidimensionale Zerlegung geschieht hier mittels eines Tensorprodukt-Ansatzes. Wie wir später sehen, erhält $Q^0_{1,\mathcal{V}}$ dadurch eine dreibandige Struktur. Anhand der Bilder auf den rechten Seiten erkennt man deutlich, daß die Inversen $W^1_{0,\mathcal{V}}$ sich aus Matrizen $P^1_{0,\mathcal{V}}$ und $Q^1_{0,\mathcal{V}}$ mit im Vergleich zu $P^0_{1,\mathcal{V}}$ und $Q^0_{1,\mathcal{V}}$ transponierter Gestalt zusammensetzen. Aus Gründen der Übersichtlichkeit

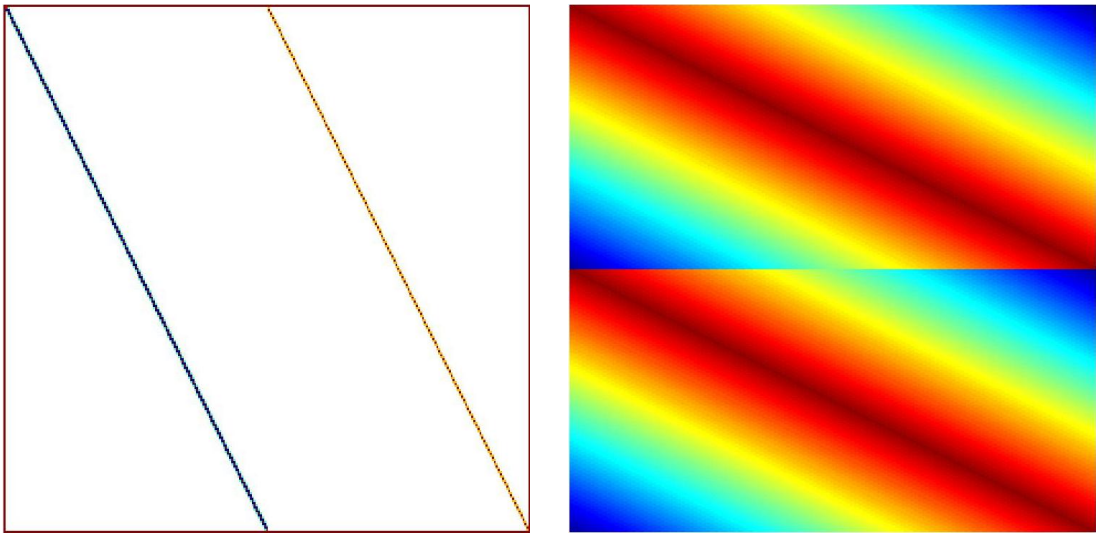


ABBILDUNG 3.2: *Zweiskalen-Transformation $W^0_{1,\mathcal{V}}$ (links) und ihre Inverse $W^1_{0,\mathcal{V}}$ (rechts) für eine eindimensionale Zerlegung mittels linearer Prewavelets auf $\Omega =]0,1[$, $h_0 = \frac{1}{256}$.*

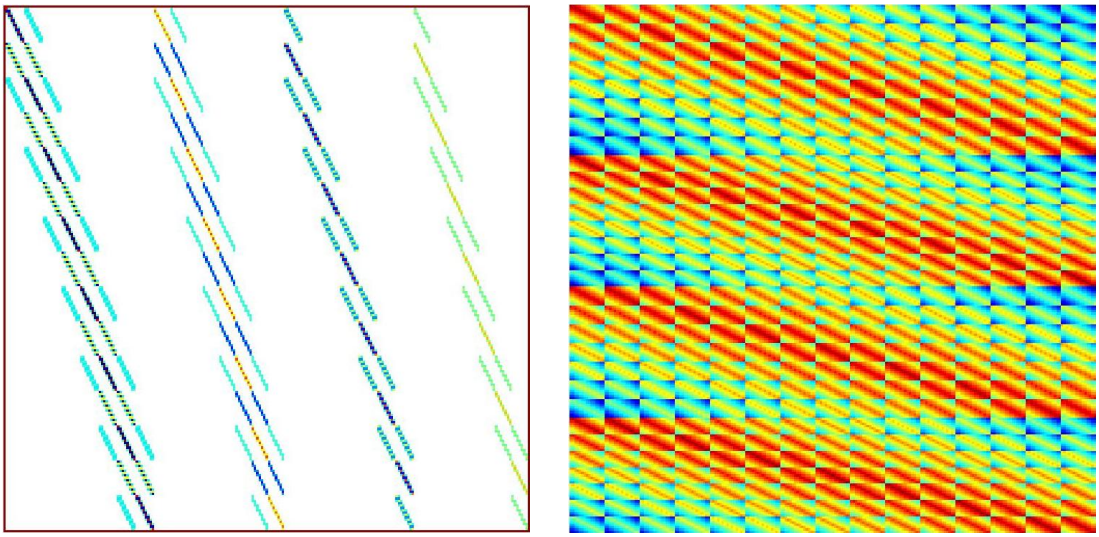


ABBILDUNG 3.3: *Zweiskalen-Transformation $W^0_{1,\mathcal{V}}$ (links) und ihre Inverse $W^1_{0,\mathcal{V}}$ (rechts) für eine zweidimensionale Zerlegung mittels linearer Prewavelets auf $\Omega =]0,1[^2$, $h_0 = \frac{1}{16}$, Tensorprodukt-Ansatz.*

wird im folgenden auf die Index-Kennzeichnung $\mathcal{V}$ bei Transformationsmatrizen verzichtet. Wir erhalten für die i-te Zweiskalen-Basisfunktion $\theta^1_{i,\mathcal{V}} = \sum_{m=1}^{N_0} (W^0_1)_{m,i} \varphi^0_{m,\mathcal{V}}$ und damit für die Einträge der Zweiskalen-Darstellung $\tilde{T}_0$ der Steifigkeitsmatrix T_0

$$(\tilde{T}_0)_{j,i} = a(\theta^1_{i,\mathcal{V}}, \theta^1_{j,\mathcal{V}}) = [(W^0_1)^t T_0 W^0_1]_{j,i}, \tag{3.8}$$

wie man leicht durch Einsetzen der Definitionen erkennt. Das lineare Gleichungssystem (3.4) transformiert sich dann bezüglich der einfachen Zerlegung zu

$$\begin{aligned} & (W^0_1)^t T_0 W^0_1 \tilde{u}_0 &&= (W^0_1)^t f_0 \\ \Longleftrightarrow \quad & \tilde{T}_0 \tilde{u}_0 &&= \tilde{f}_0, \end{aligned} \tag{3.9}$$

wobei $\tilde{u}_0 := (W_1^0)^{-1} u_0$ die nodale Zweiskalen-Darstellung der Funktion u_0 und $\tilde{f}_0 := (W_1^0)^t f_0$ die Momentendarstellung von f_0 unter der Zerlegung (3.6) bezeichnen.

Satz 6 (Äquivalenz zwischen dem ursprünglichen und dem einfach transformierten linearen Gleichungssystem)
Die linearen Gleichungssysteme (3.4) und (3.9) sind äquivalent.

Beweis:
Wir verzichten auf die Levelkennzeichung. Es sei $\tilde{u}$ eine Lösung von (3.9) und $u := W\tilde{u}$. Dann erfüllt u die Gleichung $W^t T u = W^t f$ und wegen der Invertierbarkeit von W^t auch (3.4). Umgekehrt sei u eine Lösung von (3.4). Dann gilt $W^{-t}\tilde{T}W^{-1}u = f$ und $\tilde{u} := W^{-1}u$ ist Lösung von (3.9).

□

Aufgrund der speziellen Gestalt von W_1^0 kann (3.9) auch in der expliziten Block-Form

$$\begin{pmatrix} A_1 & B_1 \\ C_1 & T_1 \end{pmatrix} \begin{pmatrix} d_1^u \\ s_1^u \end{pmatrix} = \begin{pmatrix} d_1^f \\ s_1^f \end{pmatrix} \tag{3.10}$$

mit den Matrizen

$$\begin{array}{rclcrcl} A_1 & := & (Q_1^0)^t T_0 Q_1^0, & & B_1 & := & (Q_1^0)^t T_0 P_1^0, \\ C_1 & := & (P_1^0)^t T_0 Q_1^0, & & T_1 & := & (P_1^0)^t T_0 P_1^0 \end{array}$$

sowie den Block-Vektoren $[d_1^u; s_1^u] := (W_1^0)^{-1} u_0$ und $[d_1^f; s_1^f] := (W_1^0)^t f_0$ geschrieben werden.

3.2.2 Petrov–Galerkin-artige Zweiskalen-Zerlegung

Wir betrachten nun das lineare Gleichungssystem (3.5), das man über einen Petrov–Galerkin-Ansatz gewinnt, der im allgemeinen voneinander verschiedene Ansatz- und Testräume $\mathcal{V}_0$ und $\mathcal{S}_0$ zum Aufstellen der diskreten schwachen Form verwendet. Man erhält zusätzliche Zweiskalen-Transformationen aufgrund der einfachen Zerlegung von $\mathcal{S}_0 = \mathrm{span}\{\varphi_{i,\mathcal{S}}^0 : x_i^0 \in \Omega_0\}$ als

$$\mathcal{S}_0 = \mathcal{F}_1 \oplus \mathcal{S}_1 \tag{3.11}$$

mit den entsprechenden Basisdarstellungen der Unterräume Detail- und grobskaliger Anteile

$$\mathcal{F}_1 = \mathrm{span}\{\psi_{i,\mathcal{S}}^1 : x_i^0 \in \Omega_0^f\}, \quad \mathcal{S}_1 = \mathrm{span}\{\varphi_{i,\mathcal{S}}^1 : x_i^0 \in \Omega_1\}. \tag{3.12}$$

Das Ausgangssystem (3.5) wird damit transformiert zum dazu äquivalenten System

$$\begin{array}{rcrcl} & & (W_{1,\mathcal{S}}^0)^t T_0 W_{1,\mathcal{V}}^0 \tilde{u}_0 & = & (W_{1,\mathcal{S}}^0)^t f_0 \\ \Longleftrightarrow & & \tilde{T}_0 \tilde{u}_0 & = & \tilde{f}_0, \end{array} \tag{3.13}$$

wobei $\tilde{u}_0 := (W_{1,\mathcal{V}}^0)^{-1} u_0$ die Zweiskalen-Darstellung der Funktion $u_0 \in \mathcal{V}_0$ gemäß der Zerlegung (3.6), $\tilde{f}_0 := (W_{1,\mathcal{S}}^0)^t f_0$ die Momentendarstellung von $f_0 \in \mathcal{S}_0$ relativ zur Zerlegung (3.11) und $\tilde{T}_0 := (W_{1,\mathcal{S}}^0)^t T_0 W_{1,\mathcal{V}}^0$ die Zweiskalen-transformierte Steifigkeitsmatrix im Petrov–Galerkin-Fall bezeichnen. Mit den Definitionen

$$\begin{array}{rclcrcl} A_1 & := & (Q_{1,\mathcal{S}}^0)^t T_0 Q_{1,\mathcal{V}}^0, & & B_1 & := & (Q_{1,\mathcal{S}}^0)^t T_0 P_{1,\mathcal{V}}^0, \\ C_1 & := & (P_{1,\mathcal{S}}^0)^t T_0 Q_{1,\mathcal{V}}^0, & & T_1 & := & (P_{1,\mathcal{S}}^0)^t T_0 P_{1,\mathcal{V}}^0 \end{array}$$

sowie den Block-Vektoren $[d_1^u; s_1^u] := (W_{1,\mathcal{V}}^0)^{-1} u_0$ und $[d_1^f; s_1^f] := (W_{1,\mathcal{S}}^0)^t f_0$ läßt sich (3.13) wieder in einer expliziten Block-Form analog zu (3.10) schreiben.

Wir fassen nun die Abbildungseigenschaften der unterschiedlichen Transformationsmatrizen und der ursprünglichen sowie Zweiskalen-transformierten Steifigkeitsmatrix zusammen. Der Deutlichkeit halber indizieren wir dabei die Darstellungen vollständig mit den Räumen, auf die sie sich beziehen.

$$\begin{array}{llllllll}
W_{1,\mathcal{V}}^0 & : KB_{\mathcal{W}_1\oplus\mathcal{V}_1} & \longrightarrow & KB_{\mathcal{V}_0}, & (W_{1,\mathcal{V}}^0)^{-1} & : KB_{\mathcal{V}_0} & \longrightarrow & KB_{\mathcal{W}_1\oplus\mathcal{V}_1}, \\
(W_{1,\mathcal{V}}^0)^t & : MOM_{\mathcal{V}_0} & \longrightarrow & MOM_{\mathcal{W}_1\oplus\mathcal{V}_1}, & (W_{1,\mathcal{V}}^0)^{-t} & : MOM_{\mathcal{W}_1\oplus\mathcal{V}_1} & \longrightarrow & MOM_{\mathcal{V}_0}, \\
W_{1,\mathcal{S}}^0 & : KB_{\mathcal{F}_1\oplus\mathcal{S}_1} & \longrightarrow & KB_{\mathcal{S}_0}, & (W_{1,\mathcal{S}}^0)^{-1} & : KB_{\mathcal{S}_0} & \longrightarrow & KB_{\mathcal{F}_1\oplus\mathcal{S}_1}, \\
(W_{1,\mathcal{S}}^0)^t & : MOM_{\mathcal{S}_0} & \longrightarrow & MOM_{\mathcal{F}_1\oplus\mathcal{S}_1}, & (W_{1,\mathcal{S}}^0)^{-t} & : MOM_{\mathcal{F}_1\oplus\mathcal{S}_1} & \longrightarrow & MOM_{\mathcal{S}_0},
\end{array}$$

$$\begin{array}{lll}
T_0 & : KB_{\mathcal{V}_0} & \longrightarrow MOM_{\mathcal{S}_0}, \\
\tilde{T}_0 & : KB_{\mathcal{W}_1\oplus\mathcal{V}_1} & \longrightarrow MOM_{\mathcal{F}_1\oplus\mathcal{S}_1}.
\end{array}$$

Es transformiert also beispielsweise $W_{1,\mathcal{V}}^0$ die Zweiskalen-Darstellung $\tilde{u}_0$ einer Funktion $u_0 \in \mathcal{V}_0$ in die zugehörige nodale Feingitterdarstellung. Die transformierte Steifigkeitsmatrix $\tilde{T}_0$ ordnet dem Vektor $\tilde{u}_0$ den Vektor $\tilde{f}_0$ zu, der die Momentendarstellung von $f_0 = T_0 u_0$ bezüglich der einfachen Zerlegung des Testraums $\mathcal{S}_0$ beschreibt.

3.3 Standard- und Nichtstandardform

Wir erweitern nun die Betrachtungen des letzten Abschnitts auf Zerlegungen, an denen mehr als zwei Skalen beteiligt sind. Es wird hier nur der reine Galerkin-Fall untersucht, dem der Feingitterraum $\mathcal{V}_0$ als Ansatz- und Testraum zugrunde liegt. Die Übertragung der Konstruktionen auf den Petrov–Galerkin-Fall ist ohne Probleme möglich.

3.3.1 Standardform

Es sei $W_{lt,s}^0$ die Matrix, die die Darstellung einer Funktion $u_0 \in \mathcal{V}_0$ mittels der rekursiv entstandenen Multiskalen-Basis zur Zerlegung (3.2)

$$E_s := \bigcup_{k=1}^{lt} \{\psi_{i,\mathcal{V}}^k : x_i^{k-1} \in \Omega_{k-1}\backslash\Omega_k\} \cup \{\varphi_{i,\mathcal{V}}^{lt} : x_i^{lt} \in \Omega_{lt}\} \tag{3.14}$$

in ihre Darstellung durch die nodale Feingitterbasis von $\mathcal{V}_0$ transformiert. Größen, die sich auf die Multiskalen-Basis (3.14) beziehen, werden durch einen unteren Index $_s$ gekennzeichnet.

Definition 2 (Standardform)
Wir definieren als Standardform *des diskreten Feingitteroperators T_0 zur Multiskalen-Basis (3.14) die Matrix*

$$\tilde{T}_{0,s} := (W_{lt,s}^0)^t T_0 W_{lt,s}^0. \tag{3.15}$$

Die Standardform ist also die Darstellung des Feingitteroperators unter dem zugehörigen Basiswechsel [19]. Das entsprechend transformierte lineare Gleichungssystem lautet

$$\begin{array}{rrcl}
 & (W_{lt,s}^0)^t T_0 W_{lt,s}^0 \tilde{u}_{0,s} & = & (W_{lt,s}^0)^t f_0 \\
\Longleftrightarrow & \tilde{T}_{0,s} \tilde{u}_{0,s} & = & \tilde{f}_{0,s},
\end{array} \tag{3.16}$$

wobei $\tilde{u}_{0,s} := (W^0_{lt,s})^{-1} u_0$ die Multiskalen-Darstellung der Funktion u_0 und $\tilde{f}_{0,s} := (W^0_{lt,s})^t f_0$ die Momentendarstellung von f_0 unter der Zerlegung (3.2) bezeichnen. Wir untersuchen nun genauer die Struktur der hierbei auftretenden Transformationen, die für die konkrete Gestalt des transformierten linearen Gleichungssystems (3.16) von Bedeutung sind. Definiert man für $k = 1, ..., lt$ die *Synthese-Operatoren*

$$\begin{aligned} P^0_k : \mathcal{V}_k \longrightarrow \mathcal{V}_0; \quad P^0_k &:= P^0_1 \cdot ... \cdot P^{k-2}_{k-1} \cdot P^{k-1}_k, && (3.17)\\ Q^0_k : \mathcal{W}_k \longrightarrow \mathcal{V}_0; \quad Q^0_k &:= P^0_1 \cdot ... \cdot P^{k-2}_{k-1} \cdot Q^{k-1}_k && (3.18)\\ &= P^0_{k-1} \cdot Q^{k-1}_k \end{aligned}$$

ausgehend vom rekursiven Aufbau der Multiskalen-Zerlegung (3.2) über einfache Zerlegungen $\mathcal{V}_{k-1} = \mathcal{W}_k \oplus \mathcal{V}_k$ und den zugehörigen Transformationsmatrizen P^{k-1}_k und Q^{k-1}_k, so erhält man $W^0_{lt,s}$ in expliziter Form als

$$W^0_{lt,s} = [Q^0_1, ..., Q^0_{lt}, P^0_{lt}]. \tag{3.19}$$

Die inverse Matrix $W^{lt}_{0,s} := (W^0_{lt,s})^{-1}$, welche die nodale Feingitterdarstellung einer Funktion $u_0 \in \mathcal{V}_0$ in ihre Multiskalen-Darstellung transformiert, kann entsprechend mittels der für $k = 1, ..., lt$ erklärten *Analyse-Operatoren*

$$\begin{aligned} P^k_0 : \mathcal{V}_0 \longrightarrow \mathcal{V}_k; \quad P^k_0 &:= P^k_{k-1} \cdot P^{k-1}_{k-2} \cdot ... \cdot P^1_0, && (3.20)\\ Q^k_0 : \mathcal{V}_0 \longrightarrow \mathcal{W}_k; \quad Q^k_0 &:= Q^k_{k-1} \cdot P^{k-1}_{k-2} \cdot ... \cdot P^1_0 && (3.21)\\ &= Q^k_{k-1} \cdot P^{k-1}_0 \end{aligned}$$

gebildet werden. Die in den obigen Produkten auftretenden Faktoren P^k_{k-1} und Q^k_{k-1} sind dabei gerade diejenigen Teilmatrizen, aus denen sich die inversen Zweiskalen-Transformationen $W^k_{k-1} := (W^{k-1}_k)^{-1} := [Q^{k-1}_k, P^{k-1}_k]^{-1}$ als $W^k_{k-1} = [Q^k_{k-1}; P^k_{k-1}]$ zusammensetzen. Wir erhalten

$$W^{lt}_{0,s} = [Q^1_0; ...; Q^{lt}_0; P^{lt}_0]. \tag{3.22}$$

Definieren wir nun die grobskaligen Anteile s^u_k, s^f_k und Detail-Anteile d^u_k, d^f_k der Vektoren u_0 und f_0 für $k = (0), ..., lt$ bezüglich der Multiskalen-Zerlegung (3.2) durch

$$\begin{aligned} s^u_k &:= P^k_0 u_0, \quad & d^u_k &:= Q^k_0 u_0,\\ s^f_k &:= (P^0_k)^t f_0, \quad & d^f_k &:= (Q^0_k)^t f_0, \end{aligned} \tag{3.23}$$

wobei $P^0_0 := \mathbf{1}_{\mathcal{V}_0}$ als Einheitsmatrix der Dimesion $\dim(\mathcal{V}_0)$ zu setzen ist und d^u_0 sowie d^f_0 nicht definiert werden, so erhalten wir aufgrund von (3.19) und (3.22) die folgenden Block-Vektor Darstellungen

$$\begin{aligned} \tilde{u}_{0,s} &= (W^0_{lt,s})^{-1} u_0 &&= [d^u_1; ...; d^u_{lt}; s^u_{lt}],\\ \tilde{f}_{0,s} &= (W^0_{lt,s})^t f_0 &&= [d^f_1; ...; d^f_{lt}; s^f_{lt}]. \end{aligned}$$

Wir möchten an dieser Stelle nicht den vollständigen Aufbau von $\tilde{T}_{0,s}$ mit sämtlichen daran beteiligten Block-Matrizen für beliebige Zerlegungstiefen beschreiben, was anhand der expliziten Darstellung (3.19) von $W^0_{lt,s}$ prinzipiell möglich ist. Wir beschränken uns auf die

Angabe der nachfolgenden Skizze, die den allgemeinen Aufbau einer Standardform im ersten nichttrivialen Fall $lt = 2$ zeigt.

$$\tilde{T}_{0,s} = \begin{array}{|c|c|c|} \hline A_1 & B_1^2 & B_1^3 \\ \hline C_1^2 & A_2 & B_2^3 \\ \hline C_1^3 & C_2^3 & T_2 \\ \hline \end{array}$$

Die darin auftretenden Blöcke sind dann aufgrund von (3.15) und (3.19) wie folgt definiert:

$$\begin{array}{lll} A_1 := (Q_1^0)^t T_0 Q_1^0, & B_1^2 := (Q_1^0)^t T_0 P_1^0 Q_2^1, & B_1^3 := (Q_1^0)^t T_0 P_1^0 P_2^1, \\ C_1^2 := (Q_2^1)^t (Q_1^0)^t T_0 Q_1^0, & A_2 := (Q_2^1)^t (P_1^0)^t T_0 P_1^0 Q_2^1, & B_2^3 := (Q_2^1)^t (P_1^0)^t T_0 P_1^0 P_2^1, \\ C_1^3 := (P_2^1)^t (Q_1^0)^t T_0 Q_1^0, & C_2^3 := (P_2^1)^t (P_1^0)^t T_0 P_1^0 Q_2^1, & T_2 := (P_2^1)^t (P_1^0)^t T_0 P_1^0 P_2^1. \end{array}$$

Auch im Fall von Multiskalen-Zerlegungen werden wir die im allgemeinen vollbesetzte inverse Transformation $W_{0,s}^{lt}$ nicht explizit im Verlauf von Multiskalen-Verfahren verwenden.

Wir zeigen jetzt, daß auch die Transformation $W_{lt,s}^0$ nicht unbedingt explizit aufgestellt zu werden braucht, um eine Funktion $u_0 \in \mathcal{V}_0$ aus ihrer Multiskalen-Darstellung in die nodale Feingitterdarstellung bezüglich $\mathcal{V}_0$ zu überführen. Man definiert dazu $S_{1,s}^0 := W_1^0$ und für $k = 2, ..., lt$ die quadratischen Matrizen

$$S_{k,s}^{k-1} := \begin{pmatrix} \mathbf{1}_{\mathcal{W}_1} & & & \\ & \ddots & & \\ & & \mathbf{1}_{\mathcal{W}_{k-1}} & \\ & & & W_k^{k-1} \end{pmatrix},$$

wobei $\mathbf{1}_{\mathcal{W}_i}$ für $i = 1, ..., k-1$ die Einheitsmatrix mit Dimension $\dim(\mathcal{W}_i)$ bezeichnet. Damit läßt sich $W_{lt,s}^0$ auch als das Produkt

$$W_{lt,s}^0 = S_{1,s}^0 \cdot ... \cdot S_{lt,s}^{lt-1} \tag{3.24}$$

schreiben. Durch Ausnutzen dieser Produktstruktur ist die Berechnung des Matrix-Vektor-Produkts $u_0 = W_{lt,s}^0 \tilde{u}_{0,s} = W_{lt,s}^0 [d_1^u; ...; d_{lt}^u; s_{lt}^u]$ in effizienter Weise möglich, das heißt mit einem Aufwand, der proportional zur Anzahl der Feingitterunbekannten ist, und kann wie folgt in Diagrammform dargestellt werden:

$$\begin{array}{ccccccccccccc} d_{lt}^u & & d_{lt-1}^u & & d_{lt-2}^u & & \cdots & & d_2^u & & d_1^u & & \\ & \overset{Q_{lt}^{lt-1}}{\searrow} & & \overset{Q_{lt-1}^{lt-2}}{\searrow} & & \searrow & & \searrow & & \overset{Q_2^1}{\searrow} & & \overset{Q_1^0}{\searrow} & \\ s_{lt}^u & \xrightarrow{P_{lt}^{lt-1}} & s_{lt-1}^u & \xrightarrow{P_{lt-1}^{lt-2}} & s_{lt-2}^u & \longrightarrow & \cdots & \longrightarrow & s_2^u & \xrightarrow{P_2^1} & s_1^u & \xrightarrow{P_1^0} & s_0^u \end{array} \tag{3.25}$$

Es ist klar, daß die transponierte Matrix $(W_{lt,s}^0)^t$, die die Transformation der Momentendarstellung einer Funktion $f_0 \in \mathcal{V}_0$ bezüglich der Feingitterbasis hin in die Momentendarstellung

relativ zur Multiskalen-Basis von $\mathcal{V}_0$ beschreibt, ebenfalls in faktorisierter Form geschrieben werden kann. Wir erhalten für die Berechnung von $\tilde{f}_{0,s} = [d_1^f; ...; d_{lt}^f; s_{lt}^f] = (W_{lt,s}^0)^t f_0$ das Diagramm:

$$\begin{array}{ccccccccccccccc} d_{lt}^f & & & d_{lt-1}^f & & & d_{lt-2}^f & & \cdots & & d_2^f & & d_1^f & & \\ & \overset{(Q_{lt}^{lt-1})^t}{\nwarrow} & & & \overset{(Q_{lt-1}^{lt-2})^t}{\nwarrow} & & & \nwarrow & & \nwarrow & & \overset{(Q_2^1)^t}{\nwarrow} & & \overset{(Q_1^0)^t}{\nwarrow} & \\ s_{lt}^f & \overset{(P_{lt}^{lt-1})^t}{\longleftarrow} & s_{lt-1}^f & & \overset{(P_{lt-1}^{lt-2})^t}{\longleftarrow} & s_{lt-2}^f & \longleftarrow & \cdots & \longleftarrow & s_2^f & \overset{(P_2^1)^t}{\longleftarrow} & s_1^f & \overset{(P_1^0)^t}{\longleftarrow} & s_0^f \end{array} \tag{3.26}$$

Der Vollständigkeit halber geben wir noch die Diagramme für Transformationen mittels faktorisierter Darstellungen der inversen Transformationen $W_{0,s}^{lt} = (W_{lt,s}^0)^{-1}$ und $(W_{0,s}^{lt})^t = (W_{lt,s}^0)^{-t}$ an. Man erhält für die Berechnung von $\tilde{u}_{0,s} = [d_1^u; ...; d_{lt}^u; s_{lt}^u] = W_{0,s}^{lt} u_0$:

$$\begin{array}{ccccccccccccccc} d_{lt}^u & & & d_{lt-1}^u & & & d_{lt-2}^u & & \cdots & & d_2^u & & d_1^u & & \\ & \overset{Q_{lt-1}^{lt}}{\nwarrow} & & & \overset{Q_{lt-2}^{lt-1}}{\nwarrow} & & & \nwarrow & & \nwarrow & & \overset{Q_1^2}{\nwarrow} & & \overset{Q_0^1}{\nwarrow} & \\ s_{lt}^u & \overset{P_{lt-1}^{lt}}{\longleftarrow} & s_{lt-1}^u & & \overset{P_{lt-2}^{lt-1}}{\longleftarrow} & s_{lt-2}^u & \longleftarrow & \cdots & \longleftarrow & s_2^u & \overset{P_1^2}{\longleftarrow} & s_1^u & \overset{P_0^1}{\longleftarrow} & s_0^u \end{array} \tag{3.27}$$

Das Diagramm für die Transformation $f_0 = (W_{0,s}^{lt})^t \tilde{f}_{0,s} = (W_{0,s}^{lt})^t [d_1^f; ...; d_{lt}^f; s_{lt}^f]$ lautet:

$$\begin{array}{ccccccccccccccc} d_{lt}^f & & & d_{lt-1}^f & & & d_{lt-2}^f & & \cdots & & d_2^f & & d_1^f & & \\ & \overset{(Q_{lt-1}^{lt})^t}{\searrow} & & & \overset{(Q_{lt-2}^{lt-1})^t}{\searrow} & & & \searrow & & \searrow & & \overset{(Q_1^2)^t}{\searrow} & & \overset{(Q_0^1)^t}{\searrow} & \\ s_{lt}^f & \overset{(P_{lt-1}^{lt})^t}{\longrightarrow} & s_{lt-1}^f & & \overset{(P_{lt-2}^{lt-1})^t}{\longrightarrow} & s_{lt-2}^f & \longrightarrow & \cdots & \longrightarrow & s_2^f & \overset{(P_1^2)^t}{\longrightarrow} & s_1^f & \overset{(P_0^1)^t}{\longrightarrow} & s_0^f \end{array} \tag{3.28}$$

Die Schemata (3.25) und (3.26) weisen die gleiche Struktur auf wie die schnelle diskrete Wavelet-Rück- und -Hintransformation, die auch als Pyramidenalgorithmen bekannt sind [81]. Sie bilden die Grundlage für die effiziente Implementierung additiver, BPX-artiger Multiskalen-Vorkonditionierer für diskrete Operatoren, die von elliptischen partiellen Differentialgleichungen herrühren [24, 33, 130, 137]. Wir gehen hierauf an späterer Stelle noch genauer ein.

Will man für solche Probleme hingegen Multiskalen-Verfahren konstruieren, die auf wesentlich mehr Information über den transformierten Operator zugreifen als nur seine (Block-) Diagonale, so ist die Standardform $\tilde{T}_{0,s}$ schon wegen der Anzahl ihrer nichtverschwindenden Einträge keine adäquate Darstellung mehr. Sie besitzt aufgrund von (3.15) und (3.19) eine $(lt+1) \times (lt+1)$ Blockstruktur, deren einzelne Blöcke die gegenseitigen Kopplungen der an der Zerlegung (3.2) beteiligten Unterräume durch den Operator beschreiben. Aufgrund der sich so ergebenden (Multi-)„Fingerstruktur“ von $\tilde{T}_{0,s}$ ist klar, daß die Standardform im Falle uniformer Gitterverfeinerung im allgemeinen wenigstens $\mathcal{O}(N_0(lt+1))$ signifikante Einträge besitzt. Abbildung 3.4 zeigt die Besetzungsmuster der Standardform für den mittels Finite-Differenzen auf $\Omega =]0,1[^d$ $(d = 1, 2)$ diskretisierten Laplace-Operator in einer (links) und zwei (rechts) Raumdimensionen. Es wurde dazu jeweils eine Prewavelet-Zerlegung des einbeziehungsweise zweidimensionalen Feingitterraums durchgeführt. Die Zerlegung in zwei Dimensionen erfolgte über einen klassischen Tensorprodukt-Ansatz [58]. Alleine das Aufstellen der Standardform bewirkt, daß ein damit gebildetes Multiskalen-Verfahren nicht mehr mit

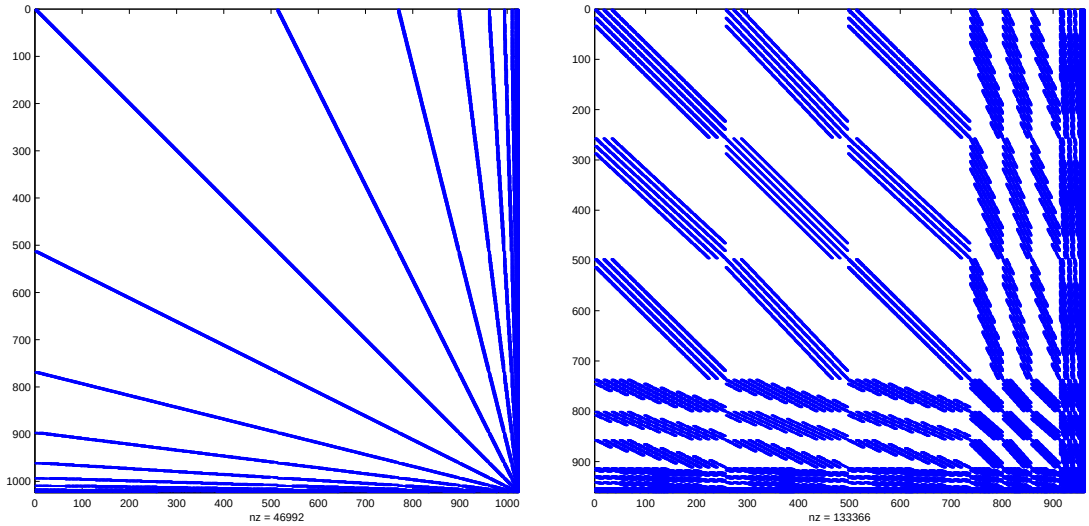


ABBILDUNG 3.4: *Standardform $\tilde{T}_{0,s}$ für Finite-Differenzen Diskretisierungen des* 1D *(links) bzw.* 2D *(rechts) Laplace-Operators auf* $\Omega =]0,1[^d$ $(d = 1, 2)$, $h_0 = \frac{1}{1024}$ (1D) *bzw.* $h_0 = \frac{1}{32}$ (2D), *vollständige Zerlegung des Ansatz- und Testraums mittels linearer Prewavelets, Tensorprodukt-Ansatz in* 2D.

linearer Komplexität, also mit einem Aufwand proportional zur Anzahl der Feingitterunbekannten, durchgeführt werden kann. Eine interessante Frage ist daher, ob es für diskrete Operatoren T_0 mit einem Besetzungsmuster mit $\mathcal{O}(N_0)$ Einträgen möglich ist, eine Multiskalen-Darstellung mit ebenfalls nur $\mathcal{O}(N_0)$ Einträgen zu finden. In [19] wurde im Zusammenhang mit der Entwicklung schneller Algorithmen für allgemeine Caldéron–Zygmund Operatoren die sogenannte *Nichtstandardform* eingeführt, die beispielsweise für Integraloperatoren nach Kompression $\mathcal{O}(N_0)$-Algorithmen zur schnellen Lösung der zugehörigen Integralgleichungen ermöglicht. Mit ihrer Hilfe läßt sich eine positive Antwort auf die oben gestellte Frage finden.

3.3.2 Nichtstandardform

Die Kernidee zur Konstruktion der Nichtstandardform $\tilde{T}_{0,ns}$ des diskreten Feingitteroperators T_0 bezüglich der Multiskalen-Zerlegung (3.2) besteht darin, rekursiv nur die Grobgitteroperator-Blöcke T_k weiter zu zerlegen, die bei den Transformationen hinsichtlich einfacher Zerlegungen entstehen.

Definition 3 (NICHTSTANDARDFORM)
Wir definieren mit Hilfe der Blöcke

$$\begin{aligned} A_k &:= (Q_k^{k-1})^t T_{k-1} Q_k^{k-1}, \quad B_k := (Q_k^{k-1})^t T_{k-1} P_k^{k-1}, \\ C_k &:= (P_k^{k-1})^t T_{k-1} Q_k^{k-1}, \quad T_k := (P_k^{k-1})^t T_{k-1} P_k^{k-1} \end{aligned}$$

für $k = 1, ..., lt-1$ *als* Nichtstandardform

$$\tilde{T}_{0,ns} := \{\{A_k, B_k, C_k\}_{k=1,...,lt}, T_{lt}\}. \tag{3.29}$$

Die in $\tilde{T}_{0,ns}$ enthaltenen Blöcke können formal als Untermatrizen in eine erweiterte Systemmatrix eingebettet werden, die wir ebenfalls als Nichtstandardform und mit $\tilde{T}_{0,ns}$ bezeichen

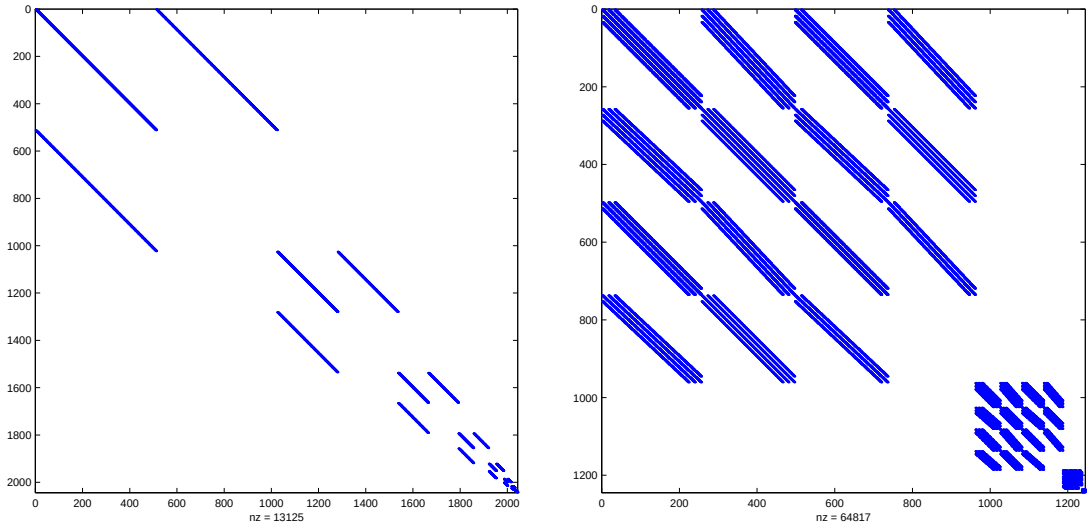


ABBILDUNG 3.5: *Nichtstandardform $\tilde{T}_{0,ns}$ für Finite-Differenzen Diskretisierungen des* 1D *(links) bzw.* 2D *(rechts) Laplace-Operators auf* $\Omega =]0,1[^d$ $(d = 1, 2)$, $h_0 = \frac{1}{1024}$ (1D) *bzw.* $h_0 = \frac{1}{32}$ (2D), *vollständige Zerlegung des Ansatz- und Testraums mittels linearer Prewavelets, Tensorprodukt-Ansatz in* 2D.

wollen. Die nachfolgende Skizze stellt den allgemeinen Aufbau einer Nichtstandardform im ersten nichttrivialen Fall $lt = 2$ dar.

$$\tilde{T}_{0,ns} = \begin{array}{|c|c|c|} \hline A_1 & B_1 & \multirow{2}{*}{} \\ \cline{1-2} C_1 & & \\ \hline \multirow{2}{*}{} & A_2 & B_2 \\ \cline{2-3} & C_2 & T_2 \\ \hline \end{array}$$

Abbildung 3.5 zeigt die Besetzungsmuster der Nichtstandardform für den mittels Finite-Differenzen auf $\Omega =]0,1[^d$ $(d = 1, 2)$ diskretisierten Laplace-Operator in einer (links) und zwei (rechts) Raumdimensionen. Es wurde dazu jeweils eine Prewavelet-Zerlegung des ein- beziehungsweise zweidimensionalen Feingitterraums durchgeführt. Hat der Feingitteroperator T_0 ein Besetzungsmuster mit $\mathcal{O}(N_0)$ Einträgen, so ist durch ein Argument mittels der Geometrischen Reihe klar, daß $\tilde{T}_{0,ns}$ ebenfalls nur $\mathcal{O}(N_0)$ Einträge ungleich Null besitzt.

Wir geben nun eine Interpretation der Nichtstandardform anhand eines speziellen Multiskalen-Erzeugendensystems. Da die Nebendiagonalblöcke C_k und B_k jeweils Abbildungen der Unterräume $\mathcal{W}_k$ und $\mathcal{V}_k$ in die Dualräume $\mathcal{V}_k'$ und $\mathcal{W}_k'$ bedeuten, ist die Nichtstandardform eine Multiskalen-Darstellung des Operators, die bezüglich der folgenden Erzeugendensystem-

Darstellung von $\mathcal{V}_0$

$$\mathcal{V}_0 = (\mathcal{W}_1 + \mathcal{V}_1) + (\mathcal{W}_2 + \mathcal{V}_2) + ... + (\mathcal{W}_{lt} + \mathcal{V}_{lt}) \tag{3.30}$$

mit der überbestimmten „Basis“

$$E_{ns} := \bigcup_{k=1}^{lt} \Big(\{\psi_{i,\mathcal{V}}^k \, : x_i^{k-1} \in \Omega_{k-1} \backslash \Omega_k\} \cup \{\varphi_{i,\mathcal{V}}^k \, : x_i^{k-1} \in \Omega_k\} \Big) \tag{3.31}$$

definiert ist. Jede nodale Darstellung einer Funktion $u_0 \in \mathcal{V}_0$ mit Hilfe des erweiterten Systems (3.31) kann mittels einer Abbildung $S_{lt,ns}^0$ in ihre nodale Feingitterdarstellung überführt werden, wobei $S_{lt,ns}^0$ auf natürliche Weise in Faktoren zerfällt:

Es sei $S_{1,ns}^0 := W_1^0$;

Für $2 \leq k \leq lt$ definieren wir Abbildungen

$$S_{k,ns}^{k-1} : (\mathcal{W}_1 \times \mathcal{V}_1) \times ... \times (\mathcal{W}_k \times \mathcal{V}_k) \longrightarrow (\mathcal{W}_1 \times \mathcal{V}_1) \times ... \times (\mathcal{W}_{k-1} \times \mathcal{V}_{k-1})$$

über ihre Darstellungen durch die rechteckigen Matrizen

$$S_{k,ns}^{k-1} := \begin{pmatrix} \begin{pmatrix} \mathbf{1}_{\mathcal{W}_1} & \\ & \mathbf{1}_{\mathcal{V}_1} \end{pmatrix} & & \\ & \ddots & \\ & & \left[\begin{pmatrix} \mathbf{1}_{\mathcal{W}_{k-1}} & \\ & \mathbf{1}_{\mathcal{V}_{k-1}} \end{pmatrix}, W_k^{k-1} \right] \end{pmatrix}.$$

Dann ist $S_{lt,ns}^0 := S_{1,ns}^0 \cdot S_{2,ns}^1 \cdot ... \cdot S_{lt,ns}^{lt-1}$ die Produktdarstellung der gesuchten Abbildung. Die Transformation einer erweiterten nodalen Darstellung $[d_1^u; s_1^u; ...; d_{lt}^u; s_{lt}^u]$ einer Funktion $u_0 \in \mathcal{V}_0$ in ihre nodale Feingitterdarstellung kann in Diagrammform wie folgt beschrieben werden:

$$\begin{array}{ccccccccccccc}
 & & s_{lt-1}^u & & s_{lt-2}^u & & \cdots & & s_2^u & & s_1^u & & \\
 & & \downarrow & & \downarrow & & & & \downarrow & & \downarrow & & \\
s_{lt}^u & \xrightarrow[Q_{lt}^{lt-1}]{P_{lt}^{lt-1}} & s_{lt-1}^u & \xrightarrow[Q_{lt-1}^{lt-2}]{P_{lt-1}^{lt-2}} & s_{lt-2}^u & \longrightarrow & \cdots & \longrightarrow & s_2^u & \xrightarrow[Q_2^1]{P_2^1} & s_1^u & \xrightarrow[Q_1^0]{P_1^0} & s_0^u \\
 & \nearrow & & \nearrow & & \nearrow & & \nearrow & & \nearrow & & \nearrow & \\
d_{lt}^u & & d_{lt-1}^u & & d_{lt-2}^u & & \cdots & & d_2^u & & d_1^u & &
\end{array} \tag{3.32}$$

Die nach unten führenden Pfeile auf jeder Stufe bedeuten lediglich ein Hinzuaddieren des auf diesem Level dargestellten grobskaligen Anteils, auf dessen weitere Zerlegung verzichtet wurde, zu den gerade interpolierten Werten. Die Bestandteile $d_1^u, s_1^u, ..., d_{lt}^u, s_{lt}^u$ der so durchlaufenen Darstellungen bezüglich des erweiterten Systems sind nicht identisch mit den durch (3.23) erklärten Anteilen, da letztere eine eindeutige Darstellung von u_0 ergeben.
Die Sequenz für die transponierte Abbildung $(S_{lt,ns}^0)^t$, die die Momentendarstellung einer Funktion $f_0 \in \mathcal{V}_0$ in eine erweiterte Momentendarstellung bezüglich (3.31) transformiert,

erhält man dann als

$$\begin{array}{ccccccccccccccc}
 & & s^f_{lt-1} & & s^f_{lt-2} & & \cdots & & s^f_2 & & s^f_1 & & \\
 & & \uparrow & & \uparrow & & & & \uparrow & & \uparrow & & \\
s^f_{lt} & \overset{(P^{lt-1}_{lt})^t}{\longleftarrow} & s^f_{lt-1} & \overset{(P^{lt-2}_{lt-1})^t}{\longleftarrow} & s^f_{lt-2} & \longleftarrow & \cdots & \longleftarrow & s^f_2 & \overset{(P^1_2)^t}{\longleftarrow} & s^f_1 & \overset{(P^0_1)^t}{\longleftarrow} & s^f_0 \\
 & \underset{(Q^{lt-1}_{lt})^t}{\swarrow} & & \underset{(Q^{lt-2}_{lt-1})^t}{\swarrow} & & \swarrow & & \swarrow & & \underset{(Q^1_2)^t}{\swarrow} & & \underset{(Q^0_1)^t}{\swarrow} & \\
d^f_{lt} & & d^f_{lt-1} & & d^f_{lt-2} & & \cdots & & d^f_2 & & d^f_1 & &
\end{array} \tag{3.33}$$

Die nach oben führenden Pfeile bedeuten hierbei, daß jeder grobskalige Anteil s^f_k von f_0 explizit in die erweiterte Darstellung aufgenommen wird. Man gelangt so zu einer *vollständig redundanten* Nichtstandard-Momentendarstellung von $f_0 \in \mathcal{V}_0$

$$\tilde{f}_{0,ns} = [d^f_1; s^f_1; ...; d^f_{lt}; s^f_{lt}] = (S^0_{lt,ns})^t f_0. \tag{3.34}$$

Neben den Detail-Anteilen treten nämlich zusätzlich auch sämtliche grobskaligen Anteile auf. Die Matrix $\tilde{T}_{0,ns}$ bildet lediglich eine Teilmatrix der zu (3.31) gehörenden Nichtstandard-Erzeugendensystem-Darstellung von T_0

$$\tilde{T}_{0,nsezs} := (S^0_{lt,ns})^t T_0 S^0_{lt,ns}. \tag{3.35}$$

Im Unterschied zu $\tilde{T}_{0,ns}$ ist $\tilde{T}_{0,nsezs}$ eine Multiskalen-Darstellung der Steifigkeitsmatrix T_0, die redundante Information enthält. Wir gehen im nächsten Unterkapitel noch genauer auf solche Erzeugendensystem-Darstellungen von Steifigkeitsmatrizen ein.

Die Nichtstandardform $\tilde{T}_{0,ns}$ kann nun gerade als diejenige Teilmatrix von $\tilde{T}_{0,nsezs}$ verstanden werden, deren Produkt mit der vollständig redundanten Nichtstandarddarstellung

$$\tilde{u}_{0,ns} = [d^u_1; s^u_1; ...; d^u_{lt}; s^u_{lt}] = S^{lt}_{0,ns} u_0 \tag{3.36}$$

einer Funktion $u_0 \in \mathcal{V}_0$ eine nichtredundante, erweiterte Momentendarstellung des Produkts $T_0 u_0$ bezüglich (3.31) ergibt. Die Abbildung $S^{lt}_{0,ns}$ ist hierbei gerade durch die Umkehrung von Schema (3.32)

$$\begin{array}{ccccccccccccccc}
 & & s^u_{lt-1} & & s^u_{lt-2} & & \cdots & & s^u_2 & & s^u_1 & & \\
 & & \uparrow & & \uparrow & & & & \uparrow & & \uparrow & & \\
s^u_{lt} & \overset{P^{lt}_{lt-1}}{\longleftarrow} & s^u_{lt-1} & \overset{P^{lt-1}_{lt-2}}{\longleftarrow} & s^u_{lt-2} & \longleftarrow & \cdots & \longleftarrow & s^u_2 & \overset{P^2_1}{\longleftarrow} & s^u_1 & \overset{P^1_0}{\longleftarrow} & s^u_0 \\
 & \underset{Q^{lt}_{lt-1}}{\swarrow} & & \underset{Q^{lt-1}_{lt-2}}{\swarrow} & & \swarrow & & \swarrow & & \underset{Q^2_1}{\swarrow} & & \underset{Q^1_0}{\swarrow} & \\
d^u_{lt} & & d^u_{lt-1} & & d^u_{lt-2} & & \cdots & & d^u_2 & & d^u_1 & &
\end{array} \tag{3.37}$$

definiert. Algorithmisch gesehen ist das Matrix-Vektor-Produkt $\tilde{T}_{0,ns}\tilde{u}_{0,ns}$ äquivalent zu einer rekursiven Berechnung von $T_0 u_0$ über die Zweiskalen-Zerlegungen $\mathcal{V}_{k-1} = \mathcal{W}_k \oplus \mathcal{V}_k$ für $k = 1, ..., lt$. Betrachten wir das Produkt $\tilde{T}_0\tilde{u}_0$ ausgehend von der einfachen Zerlegung (3.6) des feinsten Levels (siehe dazu auch die Block-Darstellung (3.10)), so bildet die Summe $(A_1 d^u_1 + B_1 s^u_1)$ gerade den ersten Detail-Anteil des Ergebnisses und wird als solcher in dessen erweiterter Momentendarstellung bezüglich (3.31) gespeichert. Der erste Summand von $(C_1 d^u_1 + T_1 s^u_1)$ wird als Bestandteil des grobskaligen Anteils ebenfalls entsprechend abgelegt. Der zweite Summand $T_1 s^u_1$ wird nun im Verlauf der rekursiven Zerlegung erst auf den gröberen

Skalen berechnet und über seine Detail- und grobskaligen Anteile im erweiterten System dargestellt. Man erhält so eine nichtredundante Nichtstandard-Momentendarstellung für das Produkt $T_0 u_0$, die mit dem Ergebnis des Matrix-Vektor-Produkts $\tilde{T}_{0,ns}\tilde{u}_{0,ns}$ übereinstimmt. Nichtstandardformen operieren also auf vollständig redundanten Darstellungen und liefern stets einen Ergebnisvektor in nichtredundanter erweiterter Darstellung zurück.
Ist $(\widetilde{T_0^{-1}})_{ns}$ die Nichtstandardform der Inversen T_0^{-1}, so wirkt diese auf Vektoren der Gestalt (3.34) und ergibt eine nichtredundante Nichtstandarddarstellung von $T_0^{-1} f_0 = u_0$. Diese kann nun mittels der Abbildung $S^0_{lt,ns}$ in die nodale Feingitterdarstellung bezüglich $\mathcal{V}_0$ zurücktransformiert werden, und wir erhalten

$$u_0 = S^0_{lt,ns}(\widetilde{T_0^{-1}})_{ns}(S^0_{lt,ns})^t f_0. \tag{3.38}$$

Im Zusammenhang mit der Matrix-Vektor-Multiplikation konnten wir die Nichtstandardform $\tilde{T}_{0,ns}$ trotz ihrer speziellen Multiskalen-Struktur als gewöhnliche Matrix auffassen. Die Berechnung von $T_0 u_0$ über die rekursive Zerlegung und das Abspeichern von grobskaligen Bestandteilen des Ergebnisses produziert die gleiche nichtredundante Darstellung wie die gewöhnliche Matrix-Vektor-Multiplikation im Nichtstandardsystem. Wir entwickeln im nachfolgenden Kapitel 4 multiplikative und additive Multiskalen-Verfahren, die auf unterschiedlichen Approximationen $\tilde{M}^{-1}_{0,ns}$ an $(\widetilde{T_0^{-1}})_{ns}$ beruhen. Zu diesem Zwecke reicht es nicht mehr aus, $\tilde{T}_{0,ns}$ als gewöhnliche Matrix anzusehen, da diese unter Umständen nicht invertierbar ist. Dies führt dazu, daß die Anwendung von $\tilde{M}^{-1}_{0,ns}$ zumeist über Prozeduren realisiert wird, die der zugrundeliegenden Multiskalen-Struktur des Nichtstandardsystems Rechnung tragen. Über die Anwendung von

$$M^{-1}_{0,ns} := S^0_{lt,ns}\tilde{M}^{-1}_{0,ns}(S^0_{lt,ns})^t \tag{3.39}$$

auf Vektoren können wir schließlich $M^{-1}_{0,ns}$ als approximative Multiskalen-Inverse des Feingitteroperators für eine lineare Iteration oder als Multiskalen-Vorkonditionierer für Krylovraum-Verfahren einsetzen.

3.3.3 Konversion zwischen dem Standard- und Nichtstandardsystem

Wir beschäftigen uns nun mit der Frage, wie die Konvertierung von Vektoren und Matrizen aus ihrer Standard- in die Nichtstandarddarstellung und umgekehrt geschieht. Damit weisen wir im nächsten Kapitel nach, daß die effizient über die Nichtstandardform erklärten Verfahren äquivalent zu entsprechenden Verfahren sind, die mit Hilfe der Standardform formuliert werden.

Konvertierung von Vektoren

1. $\tilde{u}_{0,s} \longrightarrow \tilde{u}_{0,ns} =: (\tilde{u}_{0,s})_{ns}$

Es sei $u_0 \in \mathcal{V}_0$ und $\tilde{u}_{0,s} := W^{lt}_{0,s} u_0$ die zugehörige Standarddarstellung. Die vollständig redundante Nichtstandarddarstellung $\tilde{u}_{0,ns}$ erhalten wir, indem wir ausgehend von $k = lt$ aus den Anteilen d^u_k und s^u_k von $\tilde{u}_{0,s}$ den Anteil s^u_{k-1} mittels W^{k-1}_k berechnen, diesen abspeichern und die Prozedur sukzessive für $k = lt - 1, ..., 2$ fortsetzen.

Man erreicht dies ebenfalls durch folgenden Umweg über die nodale Feingitterdarstellung:

$$\tilde{u}_{0,ns} = S^{lt}_{0,ns} W^0_{lt,s} \tilde{u}_{0,s}. \tag{3.40}$$

2. $\tilde{u}_{0,ns} \longrightarrow \tilde{u}_{0,s} =: (\tilde{u}_{0,ns})_s$

Es sei $u_0 \in \mathcal{V}_0$ und $\tilde{u}_{0,ns} := S^{lt}_{0,ns} u_0$ die zugehörige vollständig redundante Nichtstandarddarstellung. Dann erhält man $\tilde{u}_{0,s}$ durch Weglassen aller darin enthaltenen grobskaligen Anteile bis auf s^u_{lt}.

Ist $\tilde{u}_{0,ns}$ hingegen eine nichtredundante Nichtstandarddarstellung, so wird beginnend mit $k = 1$ der Anteil s^u_k zunächst mittels W^{k+1}_k in seinen Detail- und grobskaligen Bestandteil zerlegt. Diese werden zu den entsprechenden Anteilen auf der nächstgröberen Skala hinzuaddiert, bevor man diese Prozedur für $k = 2, ..., lt - 1$ fortsetzt. Die so erhaltenen Detail-Anteile bilden zusammen mit dem zuletzt bestimmten Anteil s^u_{lt} bezüglich der gröbsten Skala die Block-Vektoren, aus denen $\tilde{u}_{0,s}$ besteht.

Man erreicht letzteres ebenfalls durch einen Umweg über die nodale Feingitterdarstellung:

$$\tilde{u}_{0,s} = W^{lt}_{0,s} S^0_{lt,ns} \tilde{u}_{0,ns}. \tag{3.41}$$

3. $\tilde{f}_{0,s} \longrightarrow \tilde{f}_{0,ns} =: (\tilde{f}_{0,s})_{ns}$

Es sei $f_0 \in \mathcal{V}_0$ und $\tilde{f}_{0,s} := (W^0_{lt,s})^t f_0$ die zugehörige Standarddarstellung aus Momentensicht. Die vollständig redundante Nichtstandard-Momentendarstellung $\tilde{f}_{0,ns}$ erhalten wir, indem wir ausgehend von $k = lt$ aus den Anteilen d^f_k und s^f_k von $\tilde{f}_{0,s}$ den Anteil s^f_{k-1} mittels $(W^k_{k-1})^t$ berechnen, diesen abspeichern und die Prozedur sukzessive für $k = lt - 1, ..., 2$ fortführen.

Man erreicht dies nunmehr durch einen Umweg über die Feingitter-Momentendarstellung:

$$\tilde{f}_{0,ns} = (S^0_{lt,ns})^t (W^{lt}_{0,s})^t \tilde{f}_{0,s} = (W^{lt}_{0,s} S^0_{lt,ns})^t \tilde{f}_{0,s}. \tag{3.42}$$

4. $\tilde{f}_{0,ns} \longrightarrow \tilde{f}_{0,s} =: (\tilde{f}_{0,ns})_s$

Es sei $f_0 \in \mathcal{V}_0$ und $\tilde{f}_{0,ns} := (S^0_{lt,ns})^t f_0$ die zugehörige vollständig redundante Nichtstandard-Momentendarstellung. Dann erhält man $\tilde{f}_{0,s}$ durch Weglassen aller darin enthaltenen grobskaligen Anteile bis auf s^f_{lt}.

Ist $\tilde{f}_{0,ns}$ hingegen eine nichtredundante Nichtstandard-Momentendarstellung, so wird beginnend mit $k = 1$ der Anteil s^f_k zunächst mittels $(W^k_{k+1})^t$ in seinen Detail- und grobskaligen Bestandteil zerlegt. Diese werden zu den entsprechenden Anteilen auf der nächstgröberen Skala hinzuaddiert, bevor man diese Prozedur für $k = 2, ..., lt - 1$ fortsetzt. Die so erhaltenen Detail-Anteile bilden zusammen mit dem zuletzt erhaltenen Anteil s^f_{lt} bezüglich der gröbsten Skala die Block-Vektoren, aus denen $\tilde{f}_{0,s}$ besteht.

Man erreicht letzteres ebenfalls durch einen Umweg über die Feingitter-Momentendarstellung:

$$\tilde{f}_{0,s} = (W^0_{lt,s})^t (S^{lt}_{0,ns})^t \tilde{f}_{0,ns} = (S^{lt}_{0,ns} W^0_{lt,s})^t \tilde{f}_{0,ns}. \tag{3.43}$$

Konvertierung von Matrizen

1. $\tilde{T}_{0,ns} \longrightarrow \tilde{T}_{0,s} =: (\tilde{T}_{0,ns})_s$

Ausgehend von der Nichtstandardform $\tilde{T}_{0,ns}$ eines Operators T_0 erhält man seine Standardform $\tilde{T}_{0,s}$, indem man die Blöcke B_k und C_k mit Hilfe von Transformationen $W^k_{lt,s}$ und $(W^k_{lt,s})^t$ für $k = 1, ..., lt$ jeweils weiter zerlegt und zu einer Matrix der Größe von T_0 „zusammenschiebt“. Die Transformationen $W^k_{lt,s}$ werden dabei analog zu (3.19) gebildet.

2. $\tilde{T}_{0,s} \longrightarrow \tilde{T}_{0,ns} =: (\tilde{T}_{0,s})_{ns}$

Die Konvertierung der Standardform $\tilde{T}_{0,s}$ einer Matrix T_0 in die entsprechende Nichtstandardform $\tilde{T}_{0,ns}$ geschieht durch den umgekehrten Prozeß. Hierbei werden sukzessive die inversen Zweiskalen-Transformationen W_k^{k+1} und $(W_k^{k+1})^t$ auf die Nebendiagonalblöcke von $\tilde{T}_{0,s}$ angewandt.

Wie man leicht sieht, sind die Berechnungen des Matrix-Vektor-Produkts im Standard- und Nichtstandardsystem mit den Konvertierungsroutinen kompatibel, das heißt

$$\tilde{T}_{0,s}\tilde{u}_{0,s} = ((\tilde{T}_{0,s})_{ns}\tilde{u}_{0,ns})_s, \tag{3.44}$$

$$(\tilde{T}_{0,ns}\tilde{u}_{0,ns})_s = (\tilde{T}_{0,ns})_s\tilde{u}_{0,s}. \tag{3.45}$$

3.4 Einbettung in ein Multiskalen-Erzeugendensystem

Wir führen in diesem Abschnitt zunächst eine weitere nichtklassische Multilevel-Darstellung von diskreten Operatoren ein, die auf der gleichzeitigen Betrachtung verschiedener Diskretisierungslevel beruht [55, 56]. Die dazu notwendigen Funktionensysteme werden mit Hilfe sämtlicher nodaler Basisfunktionen unterschiedlicher Level gebildet. Man betrachtet damit wie im Fall der Nichtstandardform keine eindeutigen Basen der Ausgangsräume mehr, sondern nur noch Erzeugendensysteme. In [55, 56] wurde bewiesen, daß moderne Multilevel-Löser für lineare Gleichungssysteme, wie beispielsweise Mehrgitter-Verfahren und Multilevel-Vorkonditionierer, sich als klassische iterative Methoden (Gauß–Seidel, Jacobi-Vorkonditionierer) über den nach Transformation entstehenden semidefiniten Gleichungssystemen interpretieren lassen. Wir zeigen hier, daß die zur Standardform gehörenden Multiskalen-Transformationen des vorherigen Abschnitts jeweils auch als Produkte der entsprechenden Erzeugendensystem-Transformationen zusammen mit geeigneten Blockdiagonalmatrizen darstellbar sind. Damit können die von uns betrachteten Multiskalen-Methoden auch auf der Basis der Erzeugendensysteme in [55, 56] interpretiert werden, was im folgenden Kapitel 4 geschieht. Sogar noch größere Multiskalen-Systeme sind denkbar. Wir zeigen, daß es möglich ist, sowohl die Standard- und Nichtstandardform eines Operators als auch seine Multilevel-Erzeugendensystem-Darstellung in eine allgemeinere Multiskalen-Erzeugendensystem-Darstellung einzubetten. Wir beschränken uns in diesem Abschnitt wiederum auf den Galerkin-Fall, da die Übertragung der Konstruktionen auf den Petrov–Galerkin-Fall unproblematisch ist.

3.4.1 Multilevel-Erzeugendensystem

Betrachten wir die folgende Multilevel-Erzeugendensystem-Darstellung des Ausgangsraums

$$\mathcal{V}_0 = \mathcal{V}_0 + \mathcal{V}_1 + \mathcal{V}_2 + ... + \mathcal{V}_{lt}, \tag{3.46}$$

mit der überbestimmten „Basis“

$$E_{ezs} := \bigcup_{k=0}^{lt} \{\varphi_{i,\mathcal{V}}^k : x_i^k \in \Omega_k\}, \tag{3.47}$$

so kann jede Darstellung $\tilde{u}_{0,ezs}$ einer Funktion $u_0 \in \mathcal{V}_0$ über das erweiterte System (3.47) mittels einer Abbildung $S_{lt,ezs}^0$ in ihre nodale Feingitterdarstellung überführt werden. Ähnlich wie die Transformation $S_{lt,ns}^0$ zerfällt $S_{lt,ezs}^0$ auf natürliche Weise in Faktoren:

Es sei $S^0_{1,ezs} : \mathcal{V}_0 \times \mathcal{V}_1 \longrightarrow \mathcal{V}_0$ definiert durch die Darstellung über die rechteckige Matrix $S^0_{1,ezs} := [\mathbf{1}_{\mathcal{V}_0}, P^0_1]$;

Für $2 \leq k \leq lt$ definieren wir Abbildungen

$$S^{k-1}_{k,ezs} : \mathcal{V}_0 \times \mathcal{V}_1 \times ... \times \mathcal{V}_k \longrightarrow \mathcal{V}_0 \times \mathcal{V}_1 \times ... \times \mathcal{V}_{k-1}$$

über ihre Darstellungen durch die rechteckigen Matrizen

$$S^{k-1}_{k,ezs} := \begin{pmatrix} \mathbf{1}_{\mathcal{V}_0} & & & \\ & \mathbf{1}_{\mathcal{V}_1} & & \\ & & \ddots & \\ & & & \left[\mathbf{1}_{\mathcal{V}_{k-1}}, P^{k-1}_k\right] \end{pmatrix}.$$

Dann ist $S^0_{lt,ezs} := S^0_{1,ezs} \cdot S^1_{2,ezs} \cdot ... \cdot S^{lt-1}_{lt,ezs}$ die Produktdarstellung der gesuchten Abbildung. Die Transformation einer Multilevel-nodalen Darstellung $[s^u_0; s^u_1; ...; s^u_{lt}]$ einer Funktion $u_0 \in \mathcal{V}_0$ in ihre nodale Feingitterdarstellung kann in Diagrammform wie folgt beschrieben werden:

$$\begin{array}{ccccccccccccc} & & s^u_{lt-1} & & s^u_{lt-2} & & \cdots & & s^u_2 & & s^u_1 & & s^u_0 \\ & & \downarrow & & \downarrow & & & & \downarrow & & \downarrow & & \downarrow \\ s^u_{lt} & \xrightarrow{P^{lt-1}_{lt}} & s^u_{lt-1} & \xrightarrow{P^{lt-2}_{lt-1}} & s^u_{lt-2} & \longrightarrow & \cdots & \longrightarrow & s^u_2 & \xrightarrow{P^1_2} & s^u_1 & \xrightarrow{P^0_1} & s^u_0 \end{array} \tag{3.48}$$

Die nach unten führenden Pfeile auf jeder Stufe bedeuten lediglich ein Hinzuaddieren des auf diesem Level dargestellten nodalen Anteils zu den gerade interpolierten Werten. Die Sequenz für die umgekehrte Transformation $S^{lt}_{0,ezs}$ einer Funktion $u_0 \in \mathcal{V}_0$ aus ihrer nodalen Feingitterdarstellung in eine Multilevel-nodale Darstellung erhält man als:

$$\begin{array}{ccccccccccccc} & & s^u_{lt-1} & & s^u_{lt-2} & & \cdots & & s^u_2 & & s^u_1 & & s^u_0 \\ & & \uparrow & & \uparrow & & & & \uparrow & & \uparrow & & \uparrow \\ s^u_{lt} & \xleftarrow{P^{lt}_{lt-1}} & s^u_{lt-1} & \xleftarrow{P^{lt-1}_{lt-2}} & s^u_{lt-2} & \longleftarrow & \cdots & \longleftarrow & s^u_2 & \xleftarrow{P^2_1} & s^u_1 & \xleftarrow{P^1_0} & s^u_0 \end{array} \tag{3.49}$$

Die nach oben führenden Pfeile bedeuten hierbei, daß jeder grobskalige Anteil s^u_k von u_0 explizit in die Multilevel-nodale Darstellung aufgenommen wird. Man gelangt so zu einer vollständig redundanten nodalen Darstellung

$$\tilde{u}_{0,ezs} = [s^u_0; s^u_1; ...; s^u_{lt}] = S^{lt}_{0,ezs} u_0 \tag{3.50}$$

von u_0 innerhalb des erweiterten Systems (3.47), da sämtliche grobskaligen Bestandteile zusätzlich zur Feingitterdarstellung s^u_0 auftreten. Genauso wie im Falle von $W^{lt}_{0,s}$ und $S^{lt}_{0,ns}$ sind die Teilmatrizen P^{k+1}_k der bei der Faktorisierung von $S^{lt}_{0,ezs}$ auftretenden Faktoren im allgemeinen vollbesetzt. Aufgrund der schwachen Formulierung wird $S^{lt}_{0,ezs}$ jedoch nicht zur Implementierung von Multilevel-Verfahren gebraucht, sondern lediglich die transponierte Abbildung $(S^0_{lt,ezs})^t$. Sie beschreibt die Transformation einer Funktion f_0 aus ihrer Momentendarstellung bezüglich der Feingitterbasis von $\mathcal{V}_0$ in die vollständig redundante Multilevel-Momentendarstellung

$$\tilde{f}_{0,ezs} = [s^f_0; s^f_1; ...; s^f_{lt}] = (S^0_{lt,ezs})^t f_0 \tag{3.51}$$

bezüglich des erweiterten Systems (3.47). Wir erhalten dafür das nachfolgende Schema:

$$\begin{array}{ccccccccccc} & & s^f_{lt-1} & & s^f_{lt-2} & & \cdots & & s^f_2 & & s^f_1 & & s^f_0 \\ & & \uparrow & & \uparrow & & & & \uparrow & & \uparrow & & \uparrow \\ s^f_{lt} & \xleftarrow{(P^{lt-1}_{lt})^t} & s^f_{lt-1} & \xleftarrow{(P^{lt-2}_{lt-1})^t} & s^f_{lt-2} & \longleftarrow & \cdots & \longleftarrow & s^f_2 & \xleftarrow{(P^1_2)^t} & s^f_1 & \xleftarrow{(P^0_1)^t} & s^f_0 \end{array} \tag{3.52}$$

Der Vollständigkeit halber geben wir noch das zu $(S^{lt}_{0,ezs})^t$ gehörende Diagramm an:

$$\begin{array}{ccccccccccc} & & s^f_{lt-1} & & s^f_{lt-2} & & \cdots & & s^f_2 & & s^f_1 & & s^f_0 \\ & & \downarrow & & \downarrow & & & & \downarrow & & \downarrow & & \downarrow \\ s^f_{lt} & \xrightarrow{(P^{lt}_{lt-1})^t} & s^f_{lt-1} & \xrightarrow{(P^{lt-1}_{lt-2})^t} & s^f_{lt-2} & \longrightarrow & \cdots & \longrightarrow & s^f_2 & \xrightarrow{(P^2_1)^t} & s^f_1 & \xrightarrow{(P^1_0)^t} & s^f_0 \end{array} \tag{3.53}$$

Wir definieren nun mit Hilfe der Erzeugendensystem-Transformation $S^0_{lt,ezs}$ und ihrer Transponierten $(S^0_{lt,ezs})^t$

$$\tilde{T}_{0,ezs} := (S^0_{lt,ezs})^t T_0 S^0_{lt,ezs}, \tag{3.54}$$

die sogenannte *Multilevel-Erzeugendensystem-Darstellung* von T_0 [55, 56]. Die Matrix $\tilde{T}_{0,ezs}$ ist aufgrund von Nulleigenwerten nur noch semidefinit und besitzt gleichen Rang wie T_0. Das damit analog zum Standardsystem (3.16) entstehende lineare Gleichungssystem

$$\begin{array}{rrcl} & (S^0_{lt,ezs})^t T_0 S^0_{lt,ezs} \tilde{u}_{0,ezs} & = & (S^0_{lt,ezs})^t f_0 \\ \Longleftrightarrow & \tilde{T}_{0,ezs} \tilde{u}_{0,ezs} & = & \tilde{f}_{0,ns} \end{array} \tag{3.55}$$

ist lösbar, da die rechte Seite $\tilde{f}_{0,ns} := (S^0_{lt,ezs})^t f_0$ in konsistenter Weise gebildet wird. Die vollständig redundante Multilevel-nodale Darstellung $S^{lt}_{0,ezs} u_0$ der Lösung des Ausgangssystems (3.50) ist jedoch keine Lösung. Aufgrund der Semidefinitheit der erweiterten Systemmatrix ist die Lösung zu (3.55) nicht mehr eindeutig bestimmt. Unterschiedliche Lösungen $\tilde{u}^1_{0,ezs}$ und $\tilde{u}^2_{0,ezs}$ sind aber in dem Sinne äquivalent, daß sie nach Anwendung von $S^0_{lt,ezs}$ den gleichen Vektor

$$u_0 = S^0_{lt,ezs} \tilde{u}^1_{0,ezs} = S^0_{lt,ezs} \tilde{u}^2_{0,ezs},$$

die eindeutig bestimmte Lösung des Ausgangssystems, ergeben. Da $\tilde{T}_{0,ezs}$ nicht invertierbar ist, ist es ungeachtet der Kosten nicht möglich, einen direkter Löser wie etwa die Gauß-Elimination zum Lösen von (3.55) im Erzeugendensystem einzusetzen. Ähnlich wie im Fall der Nichtstandardform erhält man jedoch durch die Erzeugendensystem-Darstellung der Feingitterinversen $(\widetilde{T_0^{-1}})_{ezs}$ ein direktes Verfahren. Es gilt

$$u_0 = S^0_{lt,ezs} (\widetilde{T_0^{-1}})_{ezs} (S^0_{lt,ezs})^t f_0. \tag{3.57}$$

Man vergleiche dies auch mit der Darstellung (3.38). Stattdessen kann man zur Lösung auch ein lineares Iterationsverfahren

$$\begin{array}{rcl} \tilde{u}^{i+1}_{0,ezs} & = & \tilde{u}^i_{0,ezs} - \tilde{M}^{-1}_{0,ezs} (\tilde{f}_{0,ezs} - \tilde{T}_{0,ezs} \tilde{u}^i_{0,ezs}) \\ & = & \tilde{u}^i_{0,ezs} - \tilde{M}^{-1}_{0,ezs} (S^0_{lt,ezs})^t (f_0 - T_0 u^i_0) \end{array} \tag{3.58}$$

mit invertierbarer Matrix $\tilde{M}_{0,ezs}$ heranziehen, wobei $u^i_0 = S^0_{lt,ezs} \tilde{u}^i_{0,ezs}$ ist. Durch Linkstransformation mittels $S^0_{lt,ezs}$ ist (3.58) auch als lineare Iteration mit Matrix der zweiten Normalform [64]

$$M^{-1}_{0,ezs} := S^0_{lt,ezs} \tilde{M}^{-1}_{0,ezs} (S^0_{lt,ezs})^t \tag{3.59}$$

im ursprünglichen Feingittersystem interpretierbar. Man vergleiche dies wiederum mit der Darstellung (3.39). Es kann $\tilde{M}^{-1}_{0,ezs}$ auch als Vorkonditionierer für Krylovraum-Verfahren bezüglich des erweiterten Systems eingesetzt werden. Über die Transformation $S^0_{lt,ezs}$ ist eine Interpretation als Multilevel-vorkonditioniertes Krylovraum-Verfahren für das Ausgangssystem möglich. Der Multilevel-Vorkonditionierer ist dann gerade $M^{-1}_{0,ezs}$, wie in (3.59) definiert.

3.4.2 Standardsystem und Multilevel-Erzeugendensystem

Wir zeigen nun, wie die zur Standardform $\tilde{T}_{0,s}$ gehörenden Multiskalen-Transformationen jeweils auch als Produkte der entsprechenden Erzeugendensystem-Transformationen mit geeigneten Blockdiagonalmatrizen dargestellt werden können. Dazu wird die Transformation $W^0_{lt,s}$ in einen zweistufigen Prozeß zerlegt, bei dem jeder Detail-Anteil einer in die nodale Feingitterdarstellung zurück zu transformierenden Standarddarstellung $\tilde{u}_{0,s}$ einer Funktion $u_0 \in \mathcal{V}_0$ zunächst einmal in einen nächstgrobskaligeren Anteil transformiert wird. Dieser Schritt ist simultan durchführbar und erfolgt beispielsweise mit Hilfe der Blockdiagonalmatrix

$$Q^0_{lt,s} := blockdiag(Q^0_1, ..., Q^{lt-1}_{lt}, \mathbf{1}_{\mathcal{V}_{lt}}). \tag{3.60}$$

Man gelangt so zu einer Multilevel-nodalen Darstellung von u_0. Anschließend werden in einem zweiten Schritt alle Multilevel-nodalen Anteile mittels $S^0_{lt,s}$ „aufsummiert“, so daß wir für die gesamte Transformation das Produkt

$$W^0_{lt,s} = S^0_{lt,ezs} Q^0_{lt,s} \tag{3.61}$$

erhalten. Das entsprechende Diagramm setzt sich additiv aus dem ersten Anteil des Diagramms (3.25) und dem zu $S^0_{lt,ezs}$ gehörenden Diagramm (3.48) zusammen:

$$\begin{array}{ccccccccccccccc}
d^u_{lt} & & d^u_{lt-1} & & d^u_{lt-2} & & \cdots & & d^u_2 & & d^u_1 & & \\
& \overset{Q^{lt-1}_{lt}}{\searrow} & & \overset{Q^{lt-2}_{lt-1}}{\searrow} & & \searrow & & \searrow & & \overset{Q^1_2}{\searrow} & & \overset{Q^0_1}{\searrow} & \\
& & s^{d^u_{lt}}_{lt-1} & & s^{d^u_{lt-1}}_{lt-2} & & \cdots & & s^{d^u_3}_2 & & s^{d^u_2}_1 & & s^{d^u_1}_0 \\
& & \downarrow & & \downarrow & & & & \downarrow & & \downarrow & & \downarrow & \\
s^u_{lt} & \overset{P^{lt-1}_{lt}}{\longrightarrow} & s^u_{lt-1} & \overset{P^{lt-2}_{lt-1}}{\longrightarrow} & s^u_{lt-2} & \longrightarrow & \cdots & \longrightarrow & s^u_2 & \overset{P^1_2}{\longrightarrow} & s^u_1 & \overset{P^0_1}{\longrightarrow} & s^u_0
\end{array} \tag{3.62}$$

Die grobskaligen Anteile der explizit erscheinenden Erzeugendensystem-Darstellung werden hierbei durch diejenigen Detail-Anteile oben indiziert, von denen sie unmittelbar abstammen. Die transponierte Abbildung $(W^0_{lt,s})^t$, die den Wechsel von der Feingitter-Momentendarstellung einer Funktion $f_0 \in \mathcal{V}_0$ hin in die Momentendarstellung bezüglich der Multiskalen-Basis beschreibt, kann damit ebenfalls als das Produkt

$$(W^0_{lt,s})^t = (Q^0_{lt,s})^t (S^0_{lt,ezs})^t \tag{3.63}$$

geschrieben werden. Eine zu transformierende Momentendarstellung f_0 bezüglich der Feingitterbasis wird also zunächst mittels $(S^0_{lt,s})^t$ in eine Multilevel-Momentendarstellung gebracht. Anschließend wird von jedem Anteil s^f_k über die Transformationen $(Q^k_{k+1})^t$ der Detail-Anteil relativ zur nächstgröberen Skala $k+1$ extrahiert. Das zugehörige Diagramm setzt sich additiv aus dem ersten Anteil des Diagramms (3.26) und dem zu $(S^0_{lt,s})^t$ gehörenden Diagramm

(3.52) zusammen:

$$\begin{array}{ccccccccccccc}
d_{lt}^f & & d_{lt-1}^f & & d_{lt-2}^f & & \cdots & & d_2^f & & d_1^f & & \\
& {\scriptstyle (Q_{lt}^{lt-1})^t}\nwarrow & & {\scriptstyle (Q_{lt-1}^{lt-2})^t}\nwarrow & & \nwarrow & & \nwarrow & & {\scriptstyle (Q_2^1)^t}\nwarrow & & {\scriptstyle (Q_1^0)^t}\nwarrow & \\
& & s_{lt-1}^f & & s_{lt-2}^f & & \cdots & & s_2^f & & s_1^f & & s_0^f \\
& & \uparrow & & \uparrow & & & & \uparrow & & \uparrow & & \uparrow \\
s_{lt}^f & {\scriptstyle (P_{lt}^{lt-1})^t}\longleftarrow & s_{lt-1}^f & {\scriptstyle (P_{lt-1}^{lt-2})^t}\longleftarrow & s_{lt-2}^f & \longleftarrow & \cdots & \longleftarrow & s_2^f & {\scriptstyle (P_2^1)^t}\longleftarrow & s_1^f & {\scriptstyle (P_1^0)^t}\longleftarrow & s_0^f
\end{array} \tag{3.64}$$

Man erhält entsprechende Diagramme für Anwendungen der Produkte

$$W_{0,s}^{lt} = Q_{0,s}^{lt} S_{0,ezs}^{lt}, \tag{3.65}$$

$$(W_{0,s}^{lt})^t = (S_{0,ezs}^{lt})^t (Q_{0,s}^{lt})^t, \tag{3.66}$$

wobei $Q_{0,s}^{lt} := blockdiag(Q_0^1, ..., Q_{lt-1}^{lt}, \mathbf{1}_{\mathcal{V}_{lt}})$ ist. Sie werden analog zu (3.64) und (3.62) gebildet.

3.4.3 Multiskalen-Erzeugendensystem

Es sind selbstverständlich noch größere Multiskalen-Systeme vorstellbar. Wir zeigen zum Abschluß diese Kapitels eine Möglichkeit, die bislang betrachteten Multiskalen-Darstellungen eines Operators T_0 in eine allgemeinere Multiskalen-Erzeugendensystem-Darstellung einzubetten. Wir betrachten dazu die Zerlegung

$$\mathcal{V}_0 = \mathcal{V}_0 + (\mathcal{W}_1 + \mathcal{V}_1) + (\mathcal{W}_2 + \mathcal{V}_2) + ... + (\mathcal{W}_{lt} + \mathcal{V}_{lt}) \tag{3.67}$$

zusammen mit der überbestimmten „Basis“

$$E_{msezs} := \{\varphi_{i,\mathcal{V}}^0 : x_i^0 \in \Omega_0\} \cup \bigcup_{k=1}^{lt} \left\{ \{\psi_{i,\mathcal{V}}^k : x_i^{k-1} \in \Omega_{k-1} \backslash \Omega_k\} \cup \{\varphi_{i,\mathcal{V}}^k : x_i^{k-1} \in \Omega_k\} \right\}. \tag{3.68}$$

Jede Darstellung $\tilde{u}_{0,msezs}$ einer Funktion $u_0 \in \mathcal{V}$ innerhalb des erweiterten Systems (3.68) kann mittels einer Abbildung $S_{lt,msezs}^0$ in ihre nodale Feingitterdarstellung bezüglich $\mathcal{V}_0$ überführt werden, wobei $S_{lt,msezs}^0$ auf natürliche Weise in Faktoren zerfällt:

Es sei $S_{1,msezs}^0 : \mathcal{V}_0 \times (\mathcal{W}_1 \times \mathcal{V}_1) \longrightarrow \mathcal{V}_0$ definiert durch die Darstellung über die rechteckige Matrix $S_{1,msezs}^0 := [\mathbf{1}_{\mathcal{V}_0}, W_1^0]$;

Für $2 \leq k \leq lt$ definieren wir Abbildungen

$$S_{k,msezs}^{k-1} : \mathcal{V}_0 \times (\mathcal{W}_1 \times \mathcal{V}_1) \times ... \times (\mathcal{W}_k \times \mathcal{V}_k) \longrightarrow \mathcal{V}_0 \times (\mathcal{W}_1 \times \mathcal{V}_1) \times ... \times (\mathcal{W}_{k-1} \times \mathcal{V}_{k-1})$$

über ihre Darstellungen durch die rechteckigen Matrizen

$$S_{k,msezs}^{k-1} := \begin{pmatrix} \mathbf{1}_{\mathcal{V}_0} & & & \\ & \begin{pmatrix} \mathbf{1}_{\mathcal{W}_1} & \\ & \mathbf{1}_{\mathcal{V}_1} \end{pmatrix} & & \\ & & \ddots & \\ & & & \left[\begin{pmatrix} \mathbf{1}_{\mathcal{W}_{k-1}} & \\ & \mathbf{1}_{\mathcal{V}_{k-1}} \end{pmatrix}, W_k^{k-1} \right] \end{pmatrix}.$$

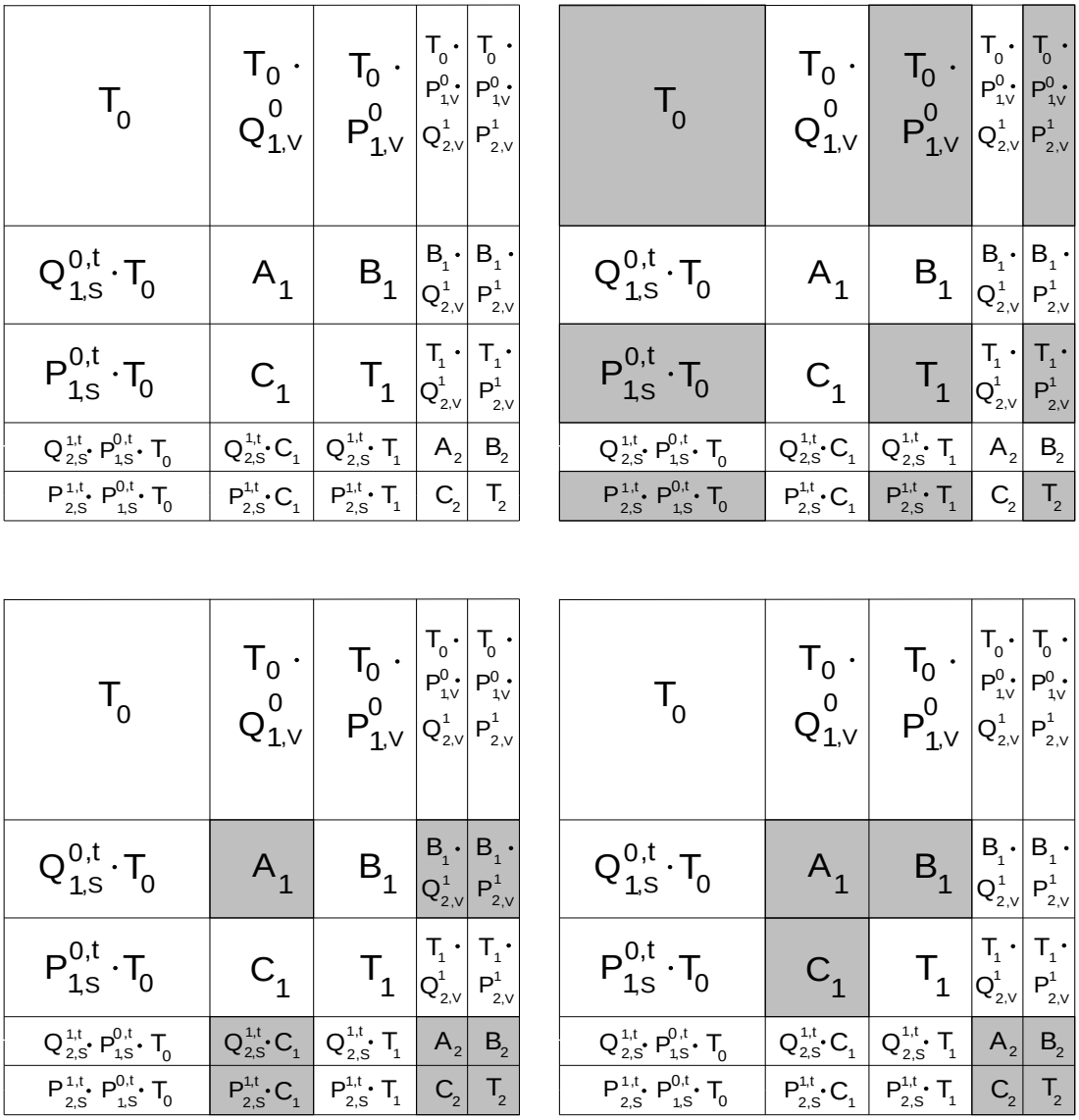


ABBILDUNG 3.6: *Schematische Darstellung der Petrov–Galerkin Multiskalen-Erzeugendensystem-Darstellung (oben links), der darin enthaltenen Multilevel-Erzeugendensystem-Darstellung (oben rechts, grau unterlegt) der Standardform (unten links, grau unterlegt) und der Nichtstandardform (unten rechts, grau unterlegt) im Fall $lt = 2$.*

Dann ist $S^0_{lt,msezs} := S^0_{1,msezs} \cdot S^1_{2,msezs} \cdot ... \cdot S^{lt-1}_{lt,msezs}$ die Produktdarstellung der gesuchten Abbildung. Wir erhalten als erweiterte Systemmatrix

$$\tilde{T}_{0,msezs} := (S^0_{lt,msezs})^t T_0 S^0_{lt,msezs}, \tag{3.69}$$

die sowohl die Standard- und Nichtstandardform von T_0 als auch die Multilevel-Erzeugendensystem-Darstellung gemäß [55, 56] als Untermatrizen enthält.

Abbildung 3.6 zeigt für $lt = 2$ im allgemeinen Petrov–Galerkin-Fall den schematischen Aufbau von $\tilde{T}_{0,msezs}$, als grau unterlegte Teilmatrizen die gewöhnliche Erzeugendensystem-Darstellung sowie auch Standard- und Nichtstandardform von T_0.

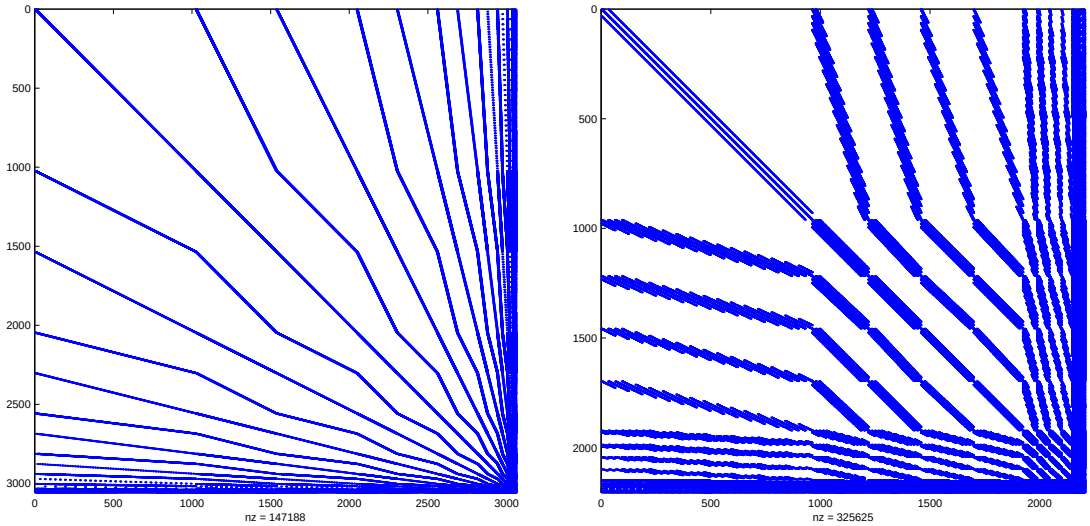


ABBILDUNG 3.7: *Multiskalen-Erzeugendensystem-Darstellung $\tilde{T}_{0,msezs}$ für Finite-Differenzen Diskretisierungen des* 1D *(links) bzw.* 2D *(rechts) Laplace-Operators auf* $\Omega =]0,1[^d$ $(d = 1, 2)$, $h_0 = \frac{1}{1024}$ (1D) *bzw.* $h_0 = \frac{1}{32}$ (2D), *vollständige Zerlegung des Ansatz- und Testraums mittels linearer Prewavelets, Tensorprodukt-Ansatz in* 2D.

In Abbildung 3.7 sehen wir die Besetzungsmuster der Multiskalen-Erzeugendensystem-Darstellung für den mittels Finite-Differenzen auf $\Omega =]0,1[^d$ $(d = 1, 2)$ diskretisierten Laplace-Operator in einer (links) und zwei (rechts) Raumdimensionen. Es wurde dazu jeweils eine Prewavelet-Zerlegung der ein- beziehungsweise zweidimensionalen Ansatz- und Testräume durchgeführt.

Kapitel 4

Iterative Verfahren für Multiskalen-transformierte lineare Gleichungssysteme

In diesem Kapitel stellen wir *multiplikative* und *additive Multiskalen-Methoden* vor, die der Lösung des Feingittersystems (3.5) ausgehend von der Standard- und der Nichtstandardform des diskreten Operators T_0 bezüglich der Multiskalen-Zerlegungen (3.2) und (3.3) dienen.

Wir betrachten zuerst in recht allgemeiner Weise *Matrixsplittings* der Standard- und Nichtstandardform eines diskreten Operators T_0 und untersuchen, unter welchen Bedingungen darüber definierte Iterationsverfahren im Standard- und Nichtstandardsystem einander äquivalent sind. Darüberhinaus wird gezeigt, wie sich diese Iterationen dann auch aus Sicht von Multilevel-Erzeugendensystemen, das heißt auf der Basis klassischer Multilevel-Verfahren darstellen lassen.

In dem danach folgenden zentralen Unterkapitel 4.2 werden zwei zunächst unterschiedlich erscheinende multiplikative Multiskalen-Verfahren definiert. Das erste Verfahren basiert auf einer approximativen Faktorisierung der Nichtstandardform $\tilde{T}_{0,ns}$ von T_0, die äquivalent zu einer entsprechenden approximativen Faktorisierung der Standardform $\tilde{T}_{0,s}$ ist. Man gelangt so zu einer approximativen Multiskalen-Inversen $M_{0,ns}^{-1}$ von T_0, die sich effizienter berechnen läßt als die dazu äquivalente über die Standardform definierte approximative Multiskalen-Inverse $M_{0,s}^{-1}$. Es kann $M_{0,ns}^{-1}$ beispielsweise innerhalb einer linearen Iteration der Form

$$x_0^{i+1} = x_0^i + M_{0,ns}^{-1}(f_0 - T_0 x_0^i), \quad (i = 0, 1, 2, ...), \tag{4.1}$$

oder auch als Multiskalen-Vorkonditionierer für Krylovraum-Verfahren (siehe Anhang) eingesetzt werden. Iteration (4.1) läßt sich zudem als Block-Gauß–Seidel Verfahren im Nichtstandardsystem interpretieren, das äquivalent zu einem entsprechenden Block-Gauß–Seidel Verfahren für das Standardsystem ist. Die zweite Technik beruht auf dem sukzessiven approximativen Lösen der Teilgleichungen für die Detail- und grobskaligen Komponenten des auf Nichtstandardform transformierten Systems, wobei die vorhergehenden Korrekturen jeweils in die darauffolgenden einfließen. Das Verfahren gehört damit zur Klasse der *multiplikativen Unterraumkorrektur-Methoden* [127]. Es kann auch als eine Verallgemeinerung der klassischen Hierarchischen Basis Mehrgitter-Methode (HBMG) angesehen [12, 16] und mittels speziell definierter Glättungsiterationen sogar als gewöhnliches Mehrgitter-Verfahren interpretiert werden. Beide Verfahren werden zunächst im Fall der einfachen Zerlegungen (3.6) und (3.11)

vorgestellt. Danach werden die entsprechenden Algorithmen anhand der Nichtstandardform $\tilde{T}_{0,ns}$ im Multiskalen-Fall besprochen. Es zeigt sich, daß sie bei jeweils exakter Invertierung der Blöcke A_k $(k = 1, ..., lt)$ äquivalent sind. Bei lediglich approximativer Invertierung dieser Blöcke ist die zweite Technik der ersten überlegen, da hierbei in den Nachkorrekturschritten modifizierte Residuen gebildet werden, die genau zu den zu lösenden Teilproblemen passen. Wir zeigen, daß eine effiziente Implementierung beider Verfahren auch ohne das vollständige Aufstellen der Nichtstandardform möglich ist. Neben den Zweiskalen-Transformationen für die Ansatz- und Testseite werden dazu im Fall einer lediglich approximativen Invertierung der Blöcke A_k nur deren approximative Inversen $\breve{A}_k^{-1}$ und die jeweiligen Grobgitteroperatoren T_k bezüglich der verschiedenen Skalen sowie der Feingitteroperator T_0 benötigt.
Nach dem Vorstellen der Grundversionen der beiden multiplikativen Multiskalen-Algorithmen diskutieren wir dann Möglichkeiten, diese weiter zu verbessern. Wir schlagen *verbesserte Varianten* vor, die genauere Schur-Komplement-Approximationen zur Bestimmung der im Verlauf der Verfahren zu berechnenden verallgemeinerten hierarchischen Residuen verwenden. Diese verbesserten Varianten sind unseres Wissens in dieser Form bislang nicht in der Literatur aufgetreten. Darüberhinaus vergleichen wir die Verfahren noch mit drei verwandten Ansätze aus der Literatur. Es sind dies eine in [48, 49] vorgeschlagene schnelle Wavelet-basierte LU-Faktorisierung, die Multilevel-Vorkonditionierung mittels AMLI (AMLI = Algebraic Multi-Level Iteration) gemäß [5, 6] sowie die Methode der Approximativen Zyklischen Reduktion [96, 97].

Es wird schließlich noch auf der Basis multiplikativer Verfahren und ihrer Implementierung mittels der Nichtstandardform eine Herleitung der entsprechenden additiven Variante als *additive Unterraumkorrektur-Methode* gegeben [127]. Eine effiziente Realisierung ist hier aber bereits mit Hilfe des Standardsystems auf einfache Weise möglich.

4.1 Matrixsplittings der Standard- und Nichtstandardform

Wir betrachten in diesem Unterkapitel additive *Matrixsplittings* der Standard- und der Nichtstandardform eines diskreten Operators T_0. Wir zeigen zunächst, daß es zu jedem Matrixsplitting der Nichtstandardform $\tilde{T}_{0,ns}$ ein äquivalentes Splitting der zugehörigen Standardform $\tilde{T}_{0,s}$ gibt und umgekehrt. Äquivalent bedeutet hierbei, daß die Anwendung der an einem Splitting beteiligten Nichtstandardform-Matrizen auf jeden Vektor $\tilde{x}_{0,ns}$ in vollständig redundanter Darstellung nach Konversion ins Standardsystem das gleiche Ergebnis liefert, wie die Anwendung der entsprechenden Standardform-Matrizen auf den zugehörigen Vektor $\tilde{x}_{0,s}$.

Satz 7 (Äquivalente Splittings von Standard- und Nichtstandardform)
Jedes Matrixsplitting $\tilde{T}_{0,s} = \tilde{M}_{0,s} + \tilde{R}_{0,s}$ der Standardform eines Operators T_0 induziert ein eindeutig bestimmtes Splitting $\tilde{T}_{0,ns} = \tilde{M}_{0,ns} + \tilde{R}_{0,ns}$ der Nichtstandardform, so daß für alle Vektoren $x_0 \in \mathbb{R}^{N_0}$ die Beziehungen

$$\tilde{M}_{0,s}\tilde{x}_{0,s} = (\tilde{M}_{0,ns}\tilde{x}_{0,ns})_s, \tag{4.2}$$

$$\tilde{R}_{0,s}\tilde{x}_{0,s} = (\tilde{R}_{0,ns}\tilde{x}_{0,ns})_s \tag{4.3}$$

gelten. Umgekehrt wird durch ein beliebiges Matrixsplitting $\tilde{T}_{0,ns} = \tilde{M}_{0,ns} + \tilde{R}_{0,ns}$ der Nichtstandardform von T_0 ein eindeutig bestimmtes Splitting $\tilde{T}_{0,s} = \tilde{M}_{0,s} + \tilde{R}_{0,s}$ der Standardform induziert, so daß für jeden Vektor $x_0 \in \mathbb{R}^{N_0}$ die Beziehungen (4.2) und (4.3) ebenfalls erfüllt sind.

Beweis:

Die Algorithmen zur gegenseitigen Konvertierung von Standard- und Nichtstandardform-Matrizen sind (algebraisch) linear (siehe Kapitel 3.3.3), so daß für Matrizen $\tilde{Z}_{0,s} = \tilde{X}_{0,s} + \tilde{Y}_{0,s}$ und $\tilde{Z}_{0,ns} = \tilde{X}_{0,ns} + \tilde{Y}_{0,ns}$, die die Form solcher Matrizen haben, gilt

$$\begin{aligned}(\tilde{Z}_{0,s})_{ns} &= (\tilde{X}_{0,s} + \tilde{Y}_{0,s})_{ns} = (\tilde{X}_{0,s})_{ns} + (\tilde{Y}_{0,s})_{ns},\\ (\tilde{Z}_{0,ns})_s &= (\tilde{X}_{0,ns} + \tilde{Y}_{0,ns})_s = (\tilde{X}_{0,ns})_s + (\tilde{Y}_{0,ns})_s.\end{aligned}$$

Ausgehend von einem Splitting in einem der beiden Darstellungsformate (Standard- oder Nichtstandardform) beweist dies die Existenz eines entsprechenden Splittings im anderen Format.

Die Berechnungen des Matrix-Vektor-Produkts in der Standard- und Nichtstandarddarstellung sind kompatibel mit den Konvertierungsroutinen, das heißt für jeden Vektor $x_0 \in \mathbb{R}^{N_0}$ haben wir gemäß (3.44) und (3.45)

$$\begin{aligned}\tilde{M}_{0,s}\tilde{x}_{0,s} &= ((\tilde{M}_{0,s})_{ns}\tilde{x}_{0,ns})_s,\\ (\tilde{M}_{0,ns}\tilde{x}_{0,ns})_s &= (\tilde{M}_{0,ns})_s\tilde{x}_{0,s}.\end{aligned}$$

Damit folgt die Beziehung (4.2) und ebenso (4.3).

Zum Nachweis der Eindeutigkeit der induzierten Splittings betrachten wir nun zwei Zerlegungen $\tilde{T}_{0,ns} = \tilde{M}_{0,ns} + \tilde{R}_{0,ns} = \tilde{M}^1_{0,ns} + \tilde{R}^1_{0,ns}$, so daß für jeden Vektor $x_0 \in \mathbb{R}^{N_0}$

$$\begin{aligned}\tilde{M}_{0,s}\tilde{x}_{0,s} &= (\tilde{M}_{0,ns}\tilde{x}_{0,ns})_s = (\tilde{M}^1_{0,ns}\tilde{x}_{0,ns})_s,\\ \tilde{R}_{0,s}\tilde{x}_{0,s} &= (\tilde{R}_{0,ns}\tilde{x}_{0,ns})_s = (\tilde{R}^1_{0,ns}\tilde{x}_{0,ns})_s\end{aligned}$$

gelten. Die Algorithmen zur gegenseitigen Konvertierung von Standard- und Nichtstandardvektoren sind ebenfalls linear, so daß also gilt

$$\begin{aligned}\mathbf{0} &= ((\tilde{M}_{0,ns} - \tilde{M}^1_{0,ns})\tilde{x}_{0,ns})_s,\\ \mathbf{0} &= ((\tilde{R}_{0,ns} - \tilde{R}^1_{0,ns})\tilde{x}_{0,ns})_s.\end{aligned}$$

Hierbei bezeichnet $\mathbf{0}$ einen Nullvektor passender Größe. Da das Matrix-Vektor-Produkt im Nichtstandardsystem eine nichtredundante Darstellung im erweiterten System liefert, folgen $\tilde{M}_{0,ns} = \tilde{M}^1_{0,ns}$ und $\tilde{R}_{0,ns} = \tilde{R}^1_{0,ns}$. Die Eindeutigkeit des induzierten Splittings bezüglich der Standardform wird mit ähnlichen Argumenten nachgewiesen.

□

Zum besseren Verständnis des obigen Satzes zeigen wir nun detailliert die Mechanismen zur Konvertierung gegebener Matrixsplittings im Fall $lt = 2$. Wir betrachten zunächst ein Splitting der Nichtstandardform $\tilde{T}_{0,ns} = \tilde{M}_{0,ns} + \tilde{R}_{0,ns}$ für den Ausgangsoperator T_0. Die hierdurch induzierten Matrixsplittings der Blöcke A_1, A_2, T_2, B_2 und C_2 von $\tilde{T}_{0,ns}$ liefern direkt solche der entsprechenden gleichen Blöcke innerhalb der Standardform. Ausgehend von den Zerlegungen

$$\begin{aligned}B_1 &= {}^{M}B_1 + {}^{R}B_1,\\ C_1 &= {}^{M}C_1 + {}^{R}C_1\end{aligned}$$

ergibt sich das gewünschte Splitting der Standardform $\tilde{T}_{0,s}$ mit den Definitionen

$$\begin{aligned} B_1^2 &= B_1 Q_{2,\mathcal{V}}^1 = ({}^M B_1 + {}^R B_1) Q_{2,\mathcal{V}}^1 =: {}^M B_1^2 + {}^R B_1^2, \\ B_1^3 &= B_1 P_{2,\mathcal{V}}^1 = ({}^M B_1 + {}^R B_1) P_{2,\mathcal{V}}^1 =: {}^M B_1^3 + {}^R B_1^3, \\ C_1^2 &= (Q_{2,\mathcal{S}}^1)^t C_1 = (Q_{2,\mathcal{S}}^1)^t ({}^M C_1 + {}^R C_1) =: {}^M C_1^2 + {}^R C_1^2, \\ C_1^3 &= (P_{2,\mathcal{S}}^1)^t C_1 = (P_{2,\mathcal{S}}^1)^t ({}^M C_1 + {}^R C_1) =: {}^M C_1^3 + {}^R C_1^3. \end{aligned}$$

Wir betrachten nun ein Splitting der Standardform $\tilde{T}_{0,s} = \tilde{M}_{0,s} + \tilde{R}_{0,s}$. Die Matrixsplittings der Blöcke A_1, A_2, T_2, B_2^3 und C_2^3 von $\tilde{T}_{0,s}$ liefern wiederum solche der entsprechenden gleichen Blöcke innerhalb der Nichtstandardform. Sind

$$\begin{aligned} B_1^2 &= {}^M B_1^2 + {}^R B_1^2, \\ B_1^3 &= {}^M B_1^3 + {}^R B_1^3, \\ C_1^2 &= {}^M C_1^2 + {}^R C_1^2, \\ C_1^3 &= {}^M C_1^3 + {}^R C_1^3, \end{aligned}$$

dann definieren wir

$$\begin{aligned} {}^M B_1 &:= [{}^M B_1^2, {}^M B_1^3] \cdot (W_{2,\mathcal{V}}^1)^{-1}, \\ {}^R B_1 &:= [{}^R B_1^2, {}^R B_1^3] \cdot (W_{2,\mathcal{V}}^1)^{-1}, \\ {}^M C_1 &:= (W_{2,\mathcal{S}}^1)^{-t} \cdot [{}^M C_1^2; {}^M C_1^3], \\ {}^R C_1 &:= (W_{2,\mathcal{S}}^1)^{-t} \cdot [{}^R C_1^2; {}^R C_1^3]. \end{aligned}$$

Man erhält so

$$\begin{aligned} B_1 &= [B_1^2, B_1^3] \cdot (W_{2,\mathcal{V}}^1)^{-1} = [{}^M B_1^2 + {}^R B_1^2, {}^M B_1^3 + {}^R B_1^3] \cdot (W_{2,\mathcal{V}}^1)^{-1} \\ &= [{}^M B_1^2, {}^M B_1^3] \cdot (W_{2,\mathcal{V}}^1)^{-1} + [{}^R B_1^2, {}^R B_1^3] \cdot (W_{2,\mathcal{V}}^1)^{-1} \\ &= {}^M B_1 + {}^R B_1, \\ C_1 &= (W_{2,\mathcal{S}}^1)^{-t} \cdot [C_1^2; C_1^3] = (W_{2,\mathcal{S}}^1)^{-t} \cdot [{}^M C_1^2 + {}^R C_1^2; {}^M C_1^3 + {}^R C_1^3] \\ &= (W_{2,\mathcal{S}}^1)^{-t} \cdot [{}^M C_1^2; {}^M C_1^3] + (W_{2,\mathcal{S}}^1)^{-t} \cdot [{}^R C_1^2; {}^R C_1^3] \\ &= {}^M C_1 + {}^R C_1 \end{aligned}$$

und damit das gewünschte Splitting der Nichtstandardform $\tilde{T}_{0,ns}$. Um zu überprüfen, ob die so definierten Zerlegungen miteinander konsistent sind, berechnen wir

$$\begin{aligned} {}^M B_1 Q_{2,\mathcal{V}}^1 &= [{}^M B_1^2, {}^M B_1^3] \cdot (W_{2,\mathcal{V}}^1)^{-1} Q_{2,\mathcal{V}}^1 \\ &= [{}^M B_1^2, {}^M B_1^3] \cdot [\mathbf{1}; \mathbf{0}] = {}^M B_1^2, \\ (Q_{2,\mathcal{S}}^1)^t {}^M C_1 &= (Q_{2,\mathcal{S}}^1)^t (W_{2,\mathcal{S}}^1)^{-t} \cdot [{}^M C_1^2; {}^M C_1^3] \\ &= [\mathbf{1}, \mathbf{0}] \cdot [{}^M C_1^2; {}^M C_1^3] = {}^M C_1^2 \end{aligned}$$

sowie ${}^M B_1 P_{2,\mathcal{V}}^1 = {}^M B_1^3$ und $(P_{2,\mathcal{S}}^1)^t {}^M C_1 = {}^M C_1^3$. Hierbei bezeichnen wir mit $\mathbf{1}$ Einheitsmatrizen sowie mit $\mathbf{0}$ Matrizen, die nur aus Nullen bestehen, und jeweils passende Größe besitzen. Für die ^{R}-Ausdrücke ergeben sich die analogen Beziehungen.

Wir weisen schließlich für dieses Beispiel die Beziehung (4.2) explizit nach und berechnen dazu den Ausdruck $(\tilde{M}_{0,ns}\tilde{x}_{0,ns})_s$, wobei $\tilde{x}_{0,ns} = [d_1^x; s_1^x; d_2^x; s_2^x]$ ist. Wir erhalten

$$(\tilde{M}_{0,ns}\tilde{x}_{0,ns})_s = \begin{pmatrix} {}^MA_1 d_1^x + {}^MB_1 s_1^x \\ {}^MC_1 d_1^x \\ {}^MA_2 d_2^x + {}^MB_2 s_2^x \\ {}^MC_2 d_2^x + {}^MT_2 s_2^x \end{pmatrix}_s = \begin{pmatrix} {}^MA_1 d_1^x + [{}^MB_1^2, {}^MB_1^3] \cdot (W_{2,\mathcal{V}}^1)^{-1} s_1^x \\ (W_{2,\mathcal{S}}^1)^{-t} \cdot [{}^MC_1^2, {}^MC_1^3] d_1^x \\ {}^MA_2 d_2^x + {}^MB_2 s_2^x \\ {}^MC_2 d_2^x + {}^MT_2 s_2^x \end{pmatrix}_s$$

$$= \begin{pmatrix} {}^MA_1 d_1^x + {}^MB_1^2 d_2^x + {}^MB_1^3 s_2^x \\ {}^MC_1^2 d_1^x + {}^MA_2 d_2^x + {}^MB_2^3 s_2^x \\ {}^MC_1^3 d_1^x + {}^MC_2^3 d_2^x + {}^MT_2 s_2^x \end{pmatrix} = \tilde{M}_{0,s}\tilde{x}_{0,s}.$$

Beziehung (4.3) kann in der gleichen Weise gezeigt werden.

Sind die Zweiskalen-Transformationen $W_{2,\mathcal{V}}^1$ und $W_{2,\mathcal{S}}^1$ schwachbesetzt, so besteht das durch ein Splitting einer Nichtstandardform $\tilde{T}_{0,ns}$ mittels schwachbesetzter Matrizen induzierte Splitting von $\tilde{T}_{0,s}$ wieder aus schwachbesetzten Matrizen. Der umgekehrte Sachverhalt ist nicht unmittelbar einzusehen, da ausgehend von einem Splitting der Standardform $\tilde{T}_{0,s}$ mit schwachbesetzten Matrizen das entsprechende Splitting im Nichtstandardsystem durch die inversen Basiswechsel $W_{1,\mathcal{V}}^2 = (W_{2,\mathcal{V}}^1)^{-1}$ und $W_{1,\mathcal{S}}^2 = (W_{2,\mathcal{S}}^1)^{-1}$ induziert wird. Diese sind aber im allgemeinen vollbesetzt, wie wir in Kapitel 3 gesehen haben. Nehmen wir zum Beispiel an, daß ${}^MB_1^2$ und ${}^MB_1^3$ ein schwaches Besetzungsmuster aufweisen, so denken wir uns den Block $[{}^MB_1^2, {}^MB_1^3]$ entstanden mittels $W_{2,\mathcal{V}}^1$ aus einer notwendig schwachbesetzten Matrix ${}^M\bar{B}_1$, nämlich $[{}^MB_1^2, {}^MB_1^3] = {}^M\bar{B}_1 \cdot W_{2,\mathcal{V}}^1$. Damit ist dann aber

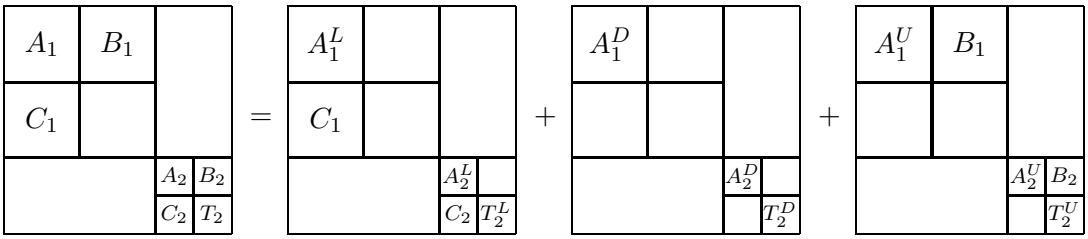
$$\begin{aligned} {}^MB_1 &= [{}^MB_1^2, {}^MB_1^3] \cdot (W_{2,\mathcal{V}}^1)^{-1} \\ &= {}^M\bar{B}_1 \cdot W_{2,\mathcal{V}}^1 \cdot (W_{2,\mathcal{V}}^1)^{-1} = {}^M\bar{B}_1 \end{aligned}$$

schwachbesetzt. Die schwachbesetzten Blöcke der am Splitting der Standardform beteiligten Matrizen lassen sich allgemein als solche interpretieren, die im Verlauf der Transformation auf Standardform aus zwangsläufig schwachbesetzten Blöcken vorheriger Zerlegungen entstanden sind. Die Rücktransformationen mittels der vollbesetzten inversen Basiswechsel führen gerade zu jenen schwachbesetzten Blöcken, die vorher zerlegt wurden.

Selbstverständlich gelten die bislang gemachten Aussagen auch für Fälle, bei denen mehr als zwei Summanden am Matrixsplitting beteiligt sind. Als konkretes Beispiel kann man das *fundamentale Splitting* der Standardform $\tilde{T}_{0,s}$ in ihren Block-unteren, Block-diagonalen und Block-oberen Anteil betrachten, mit dessen Hilfe sich beispielsweise multiplikative Multiskalen-Verfahren zur Lösung des Feingittersystems formulieren lassen. Es lautet im Fall $lt = 2$:

$$\begin{array}{|c|c|c|} \hline A_1 & B_1^2 & B_1^3 \\ \hline C_1^2 & A_2 & B_2^3 \\ \hline C_1^3 & C_2^3 & T_2 \\ \hline \end{array} = \begin{array}{|c|c|c|} \hline A_1^L & & \\ \hline C_1^2 & A_2^L & \\ \hline C_1^3 & C_2^3 & T_2^L \\ \hline \end{array} + \begin{array}{|c|c|c|} \hline A_1^D & & \\ \hline & A_2^D & \\ \hline & & T_2^D \\ \hline \end{array} + \begin{array}{|c|c|c|} \hline A_1^U & B_1^2 & B_1^3 \\ \hline & A_2^U & B_2^3 \\ \hline & & T_2^U \\ \hline \end{array}$$

Die oberen Indizes ^{L}, ^{U} und ^{D} kennzeichnen hierbei den strikten unteren und oberen sowie diagonalen Anteil der jeweiligen Blöcke. Das so erzeugte Splitting ist äquivalent zum folgenden Splitting der Nichtstandardform $\tilde{T}_{0,ns}$:



Wir zeigen jetzt, daß ein Iterationsverfahren, das auf einem Matrixsplitting der Standardform basiert, unter Umständen auch mittels des Nichtstandardsystems durchgeführt werden kann, und machen dazu die folgende Annahme. Es gebe zu einer approximativen Inversen $\tilde{M}_{0,s}^{-1}$ von $\tilde{T}_{0,s} = \tilde{M}_{0,s} + \tilde{R}_{0,s}$ eine Prozedur im Nichtstandardsystem, die mit $\tilde{M}_{0,ns}^{-1}$ bezeichnet werde und die durch das entsprechende Matrixsplitting der Nichtstandardform $\tilde{T}_{0,ns} = \tilde{M}_{0,ns} + \tilde{R}_{0,ns}$ induziert wird, so daß für jeden Vektor $x_0 \in \mathbb{R}^{N_0}$ die Beziehung

$$\tilde{M}_{0,s}^{-1}\tilde{x}_{0,s} = (\tilde{M}_{0,ns}^{-1}\tilde{x}_{0,ns})_s \tag{4.4}$$

erfüllt ist. Hierbei bedeutet der Ausdruck $\tilde{M}_{0,ns}^{-1}\tilde{x}_{0,ns}$ die Anwendung der Prozedur auf einen Vektor in völlig redundanter Nichtstandarddarstellung. Diese Voraussetzung trifft insbesondere auf das fundamentale Splitting zu, wie wir im nächsten Unterkapitel sehen werden. Sie gilt auch, wenn das Splitting der Standardform durch eine unvollständige Faktorisierung mittels unterer und oberer Dreiecksmatrizen $\tilde{L}_{0,s}$ und $\tilde{U}_{0,s}$ definiert wird:

$$\tilde{T}_{0,s} = \tilde{L}_{0,s}\tilde{U}_{0,s} + \tilde{R}_{0,s}. \tag{4.5}$$

Das hierzu äquivalente Matrixsplitting der Nichtstandardform kann genauso über die unteren und oberen Nichtstandardmatrizen $\tilde{L}_{0,ns} := (\tilde{L}_{0,s})_{ns}$ und $\tilde{U}_{0,ns} := (\tilde{U}_{0,s})_{ns}$, die durch Konversion ins Nichtstandardsystem entstehen, formuliert werden.

Satz 8 (Äquivalenz von Iterationsverfahren mittels des Standard- und Nichtstandardsystems)
Es seien die Matrixsplittings $\tilde{T}_{0,s} = \tilde{M}_{0,s} + \tilde{R}_{0,s}$ und $\tilde{T}_{0,ns} = \tilde{M}_{0,ns} + \tilde{R}_{0,ns}$ äquivalent im Sinne von Satz 7 und es existiere diesbezüglich eine Prozedur $\tilde{M}_{0,ns}^{-1}$, so daß Beziehung (4.4) erfüllt ist. Unter der Voraussetzung gleicher Startvektoren erhält man durch die damit gebildeten Iterationsverfahren

$$\begin{aligned}
&\text{i)} && x_0^{i+1} = x_0^i + W_{lt,s,\mathcal{V}}^0 \tilde{M}_{0,s}^{-1} (W_{lt,s,\mathcal{S}}^0)^t (f_0 - T_0 x_0^i), && (i = 0,1,2,...),\\
&\text{ii)} && x_0^{i+1} = x_0^i + S_{lt,ezs,\mathcal{V}}^0 Q_{lt,s}^0 \tilde{M}_{0,s}^{-1} (Q_{lt,s}^0)^t (S_{lt,ezs,\mathcal{S}}^0)^t (f_0 - T_0 x_0^i), && (i = 0,1,2,...),\\
&\text{iii)} && x_0^{i+1} = x_0^i + S_{lt,ns,\mathcal{V}}^0 \tilde{M}_{0,ns}^{-1} (S_{lt,ns,\mathcal{S}}^0)^t (f_0 - T_0 x_0^i), && (i = 0,1,2,...),
\end{aligned}$$

in jedem Schritt die gleichen Iterierten. Betrachtet man ausgehend von dem Startvektor $\tilde{x}_{0,s}^0 := W_{0,s,\mathcal{V}}^{lt} x_0^0$ das Iterationsverfahren im Standardsystem

$$\text{iv)} \quad \tilde{x}_{0,s}^{i+1} = \tilde{x}_{0,s}^i + \tilde{M}_{0,s}^{-1}(\tilde{f}_{0,s} - \tilde{T}_{0,s}\tilde{x}_{0,s}^i), \quad (i = 0,1,2,...),$$

so stimmen dessen Iterierten nach Rücktransformation in die nodale Feingitterdarstellung mit den Iterierten aus i), ii) und iii) überein.

Beweis:

Die Gleichheit der Iterierten aus i) und ii) folgt direkt mit den Produktdarstellungen (3.61) und (3.63). Der Beweis der restlichen Behauptungen erfolgt per Induktion über die Iterationsschritte. Der Induktionsanfang ($i = 0$) ist klar nach den Voraussetzungen. Wir nehmen nun an, daß die Iterierten aus i) und iii) im i-ten Schritt übereinstimmen und für die Iterierte $\tilde{x}^i_{0,s}$ aus iv) $x^i_0 = W^0_{lt,s,\mathcal{V}}\tilde{x}^i_{0,s}$ gilt. Es seien dann $r^i_0 = f_0 - T_0 x^i_0$, $\tilde{r}^i_{0,s} = (W^0_{lt,s,\mathcal{S}})^t r^i_0$ und $\tilde{r}^i_{0,ns} = (S^0_{lt,ns,\mathcal{S}})^t r^i_0$ die Residuen nach dem i-ten Iterationsschritt von i) und iii) in Feingitter-, Standard- sowie vollständig redundanter Nichtstandard-Momentendarstellung. Aufgrund von (4.4) erhalten wir mit der expliziten Darstellung (3.41) der zugehörigen Konversionsroutine, da $\tilde{M}^{-1}_{0,ns}\tilde{r}^i_{0,ns}$ eine nichtredundante Nichtstandarddarstellung ist,

$$\tilde{M}^{-1}_{0,s}\tilde{r}^i_{0,s} \quad = \quad W^{lt}_{0,s,\mathcal{V}} S^0_{lt,ns,\mathcal{V}} \tilde{M}^{-1}_{0,ns}\tilde{r}^i_{0,ns}.$$

Rücktransformation beider Seiten mittels $W^0_{lt,s,\mathcal{V}}$ zeigt dann, daß bei beiden Iterationen i) und iii) die gleichen Korrekturen der Iterierten x^i_0 berechnet werden. Hieraus ergibt sich die erste Behauptung. Wegen

$$\tilde{f}_{0,s} - \tilde{T}_{0,s}\tilde{x}^i_{0,s} = (W^0_{lt,s,\mathcal{S}})^t (f_0 - T_0 W^0_{lt,s,\mathcal{V}} W^{lt}_{0,s,\mathcal{V}} x^i_0)$$

stimmt $\tilde{r}^i_{0,s}$ nach Induktionsannahme ebenfalls mit dem Residuum nach dem i-ten Iterationsschritt von iv) überein. Es wird aus Sicht des Standardsystems also bei der Iteration iv) die gleiche Korrektur berechnet wie bei i). Damit folgt auch die zweite Behauptung.

□

4.2 Multiplikative Multiskalen-Verfahren

4.2.1 Multiplikative Verfahren für einfache Zerlegungen

Wir untersuchen zum besseren Verständnis der beiden im nächsten Abschnitt vorgestellten Multiskalen-Verfahren diese zunächst im Fall der Zerlegungstiefe $lt = 1$, das heißt für ein Gleichungssystem der Form (3.10), hervorgerufen durch die einfachen direkten Splittings von Ansatz- und Testraum $\mathcal{V}_0 = \mathcal{W}_1 \oplus \mathcal{V}_1$ und $\mathcal{S}_0 = \mathcal{F}_1 \oplus \mathcal{S}_1$. Unsere Darstellung erfolgt dabei direkt im betrachteten Zweiskalen-System, um beide Methoden besser vergleichen zu können.

Symmetrische Block-Gauß-Seidel Iteration

Wir untersuchen die Iteration

$$\tilde{x}^{i+1}_0 = \tilde{x}^i_0 + \tilde{M}^{-1}_0(\tilde{f}_0 - \tilde{T}_0\tilde{x}^i_0), \quad (i = 0, 1, 2, ...), \tag{4.6}$$

wobei $\tilde{M}_0$ über folgende approximative Faktorisierung von $\tilde{T}_0$ erklärt ist:

$$\tilde{M}_0 := \begin{pmatrix} A_1 & \mathbf{0} \\ C_1 & T_1 \end{pmatrix} \begin{pmatrix} \mathbf{1} & A_1^{-1}B_1 \\ \mathbf{0} & \mathbf{1} \end{pmatrix} =: \tilde{L}_0\tilde{U}_0. \tag{4.7}$$

Es entspricht damit (4.6) einer symmetrischen Block-Gauß-Seidel Iteration für das einfach transformierte System, da gilt:

$$\tilde{M}_0 = \left[\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ C_1 & \mathbf{0} \end{pmatrix} + \begin{pmatrix} A_1 & \mathbf{0} \\ \mathbf{0} & T_1 \end{pmatrix}\right] \begin{pmatrix} A_1 & \mathbf{0} \\ \mathbf{0} & T_1 \end{pmatrix}^{-1} \left[\begin{pmatrix} \mathbf{0} & B_1 \\ \mathbf{0} & \mathbf{0} \end{pmatrix} + \begin{pmatrix} A_1 & \mathbf{0} \\ \mathbf{0} & T_1 \end{pmatrix}\right].$$

Mit den Matrizen $\hat{A}_1\check{A}_1 := A_1$ (LU-Faktorisierung), $\hat{C}_1 := C_1\check{A}_1^{-1}$, $\check{B}_1 := \hat{A}_1^{-1}B_1$ und $\hat{T}_1\check{T}_1 := T_1$ (LU-Faktorisierung) kann die Faktorisierung von $\tilde{M}_0$ auch folgendermaßen geschrieben werden:

$$\tilde{M}_0 = \begin{pmatrix} \hat{A}_1 & \mathbf{0} \\ \hat{C}_1 & \hat{T}_1 \end{pmatrix} \begin{pmatrix} \check{A}_1 & \check{B}_1 \\ \mathbf{0} & \check{T}_1 \end{pmatrix}. \tag{4.8}$$

Die Approximation an $\tilde{T}_0$ besteht darin, den bei der Block-Gauß-Elimination der Fein-ohne-Grobgitterunbekannten aus den transformierten Grobgittergleichungen entstehenden Schur-Komplement Operator

$$S_1 := \tilde{T}_0/T_1 := T_1 - C_1A_1^{-1}B_1$$

alleine durch dessen ersten Anteil, den Grobgitteroperator T_1, zu nähern. Eine theoretische Rechtfertigung hierfür ist im Falle symmetrisch positiv definiter Operatoren die spektrale Äquivalenz der Grobgitteroperatoren und der zugehörigen Schur-Komplement Operatoren [31].

Die Anwendung von $\tilde{M}_0^{-1} = \tilde{U}_0^{-1}\tilde{L}_0^{-1}$ auf $\tilde{f}_0 - \tilde{T}_0\tilde{x}_0^i$ wird gemäß der Darstellung (4.7) mittels einer Vorwärts- und Rückwärtssubstitution realisiert und verlangt das zweimalige Anwenden von A_1^{-1} sowie die einmalige Anwendung von T_1^{-1} (zu bestimmende Teilvektoren werden jeweils innerhalb ihrer Bestimmungsgleichung **fett** dargestellt):

i) *Vorwärtssubstitution*

Löse

$$\begin{pmatrix} A_1 & \mathbf{0} \\ C_1 & T_1 \end{pmatrix} \begin{pmatrix} \mathbf{d_{1,v}} \\ \mathbf{s_{1,v}} \end{pmatrix} \stackrel{!}{=} \begin{pmatrix} d_1^f - A_1d_1^{x^i} - B_1s_1^{x^i} \\ s_1^f - C_1d_1^{x^i} - T_1s_1^{x^i} \end{pmatrix}.$$

Man erhält

$$\begin{aligned} d_{1,v} &= A_1^{-1}(d_1^f - A_1d_1^{x^i} - B_1s_1^{x^i}), \\ s_{1,v} &= T_1^{-1}(s_1^f - C_1d_1^{x^i} - T_1s_1^{x^i} - C_1d_{1,v}). \end{aligned}$$

ii) *Rückwärtssubstitution*

Löse

$$\begin{pmatrix} \mathbf{1} & A_1^{-1}B_1 \\ \mathbf{0} & \mathbf{1} \end{pmatrix} \begin{pmatrix} \mathbf{d_{1,r}} \\ \mathbf{s_{1,r}} \end{pmatrix} \stackrel{!}{=} \begin{pmatrix} d_{1,v} \\ s_{1,v} \end{pmatrix}.$$

Man erhält

$$\begin{aligned} s_{1,r} &= s_{1,v}, \\ d_{1,r} &= d_{1,v} - A_1^{-1}B_1s_{1,v}. \end{aligned}$$

Insgesamt ergibt sich damit als neue Iterierte

$$\tilde{x}_0^{i+1} := \begin{pmatrix} d_1^{x^i} + d_{1,r} \\ s_1^{x^i} + s_{1,r} \end{pmatrix} = \begin{pmatrix} d_1^{x^i} + d_{1,v} - A_1^{-1}B_1s_{1,v} \\ s_1^{x^i} + s_{1,v} \end{pmatrix}. \tag{4.9}$$

Sukzessive Unterraumkorrektur

Die zweite Technik basiert auf dem sukzessiven Lösen der Teilgleichungen für die Detail- und grobskaligen Komponenten des einfach transformierten Systems (3.10) und kann damit als Unterraumkorrektur-Methode [127] angesehen werden. Ausgehend von einer Anfangsapproximation $\tilde{x}_0^i = [d_1^{x^i}; s_1^{x^i}]$ kann man die erste Korrektur bezüglich der Detail-Komponente als eine Vorglättung betrachten, die die transformierten Gleichungen in den Fein-ohne-Grobgitter-Freiheitsgraden löst. Dies wird bei dem nachfolgenden Lösen bezüglich der grobskaligen Komponente berücksichtigt. Die approximative Grobgitterlösung hat ihrerseits Auswirkung auf einen darauffolgenden Nachglättungsschritt, der den grobskaligen Anteil der bisherigen Approximation nun aber nicht mehr beeinflußt.

i) *Vorkorrektur der $\mathcal{W}_1$-Komponente* (ausgehend von $\tilde{x}_0^i$):

Löse

$$[A_1, B_1] \begin{pmatrix} d_1^{x^i} + \mathbf{d_{1,v}} \\ s_1^{x^i} \end{pmatrix} \stackrel{!}{=} d_1^f.$$

Man erhält

$$d_{1,v} = A_1^{-1}(d_1^f - A_1 d_1^{x^i} - B_1 s_1^{x^i})$$

und korrigiert zu

$$\tilde{x}_0^{i+\frac{1}{3}} := \begin{pmatrix} d_1^{x^i} + d_{1,v} \\ s_1^{x^i} \end{pmatrix}.$$

ii) *Korrektur der $\mathcal{V}_1$-Komponente* (ausgehend von $\tilde{x}_0^{i+\frac{1}{3}}$):

Löse

$$[C_1, T_1] \begin{pmatrix} d_1^{x^i} + d_{1,v} \\ s_1^{x^i} + \mathbf{s_{1,v}} \end{pmatrix} \stackrel{!}{=} s_1^f.$$

Man erhält

$$s_{1,v} = T_1^{-1}(s_1^f - C_1 d_1^{x^i} - T_1 s_1^{x^i} - C_1 d_{1,v})$$

und korrigiert zu

$$\tilde{x}_0^{i+\frac{2}{3}} := \begin{pmatrix} d_1^{x^i} + d_{1,v} \\ s_1^{x^i} + s_{1,v} \end{pmatrix}.$$

iii) *Nachkorrektur der $\mathcal{W}_1$-Komponente* (ausgehend von $\tilde{x}_0^{i+\frac{2}{3}}$):

Löse

$$[A_1, B_1] \begin{pmatrix} d_1^{x^i} + d_{1,v} + \mathbf{d_{1,n}} \\ s_1^{x^i} + s_{1,v} \end{pmatrix} \stackrel{!}{=} d_1^f.$$

Man erhält

$$A_1 d_{1,n} = d_1^f - A_1 d_1^{x^i} - B_1 s_1^{x^i} - A_1 d_{1,v} - B_1 s_{1,v},$$

wegen $A_1 d_{1,v} = d_1^f - A_1 d_1^{x^i} - B_1 s_1^{x^i}$

$$d_{1,n} = -A_1^{-1} B_1 s_{1,v},$$

und korrigiert schließlich zu

$$\tilde{x}_0^{i+1} := \begin{pmatrix} d_1^{x^i} + d_{1,v} - A_1^{-1} B_1 s_{1,v} \\ s_1^{x^i} + s_{1,v} \end{pmatrix}. \tag{4.10}$$

Im Fall mehrerer Skalen führt das rekursive Anwenden dieses Vorgehens zu einem Verfahren, das von seiner Struktur her einem Mehrgitter-V-$(1,1)$-Zyklus entspricht [62], bei dem außer auf dem gröbsten Gitter die levelweisen Korrekturen nur in den jeweiligen Fein-ohne-Grobgitter-Freiheitsgraden erfolgen.

Der direkte Vergleich der neuen Iterierten aus den Formeln (4.9) und (4.10) liefert die Äquivalenz der beiden Verfahren. Das gilt nicht mehr, wenn man die Korrektur der Detail-Komponente nur noch approximativ durchführt. Man erhält dann für beide Verfahren zwar

$$\begin{aligned} d_{1,v} &= \breve{A}_1^{-1}(d_1^f - A_1 d_1^{x^i} - B_1 s_1^{x^i}), \\ s_{1,v} &= T_1^{-1}(s_1^f - C_1 d_1^{x^i} - T_1 s_1^{x^i} - C_1 d_{1,v}), \end{aligned}$$

wenn $\breve{A}_1^{-1}$ eine approximative Inverse von A_1 bezeichnet. Im Nachkorrekturschritt erhalten wir aber

$$d_{1,n} = \breve{A}_1^{-1}(d_1^f - A_1 d_1^{x^i} - B_1 s_1^{x^i} - A_1 d_{1,v} - B_1 s_{1,v})$$

mit $A_1 d_{1,v} = A_1 \breve{A}_1^{-1}(d_1^f - A_1 d_1^{x^i} - B_1 s_1^{x^i}) \neq d_1^f - A_1 d_1^{x^i} - B_1 s_1^{x^i}$, was das Nichtübereinstimmen der Verfahren erklärt. Es ist klar, daß das sukzessive Verfahren in diesem Fall bessere Ergebnisse liefert als die Iteration mittels der approximativen Block-Faktorisierung, da man dabei im Nachkorrektursschritt automatisch das passende Residuum bildet.

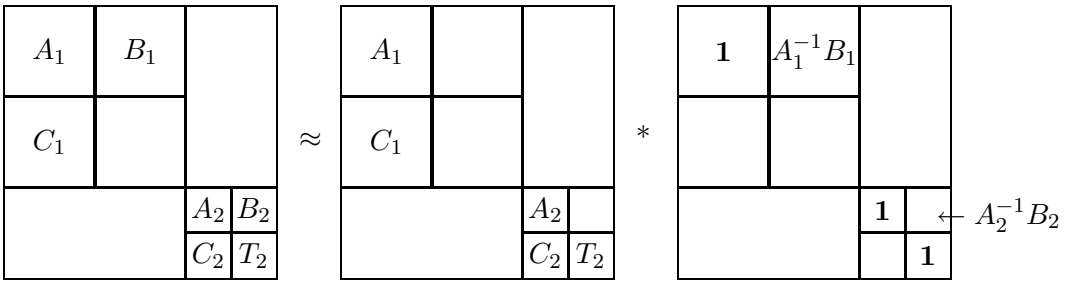
4.2.2 Multiplikative Multiskalen-Verfahren

Symmetrische Block-Gauss-Seidel Iteration mittels Nichtstandardform

Wir betrachten nunmehr die Iteration im Knotenbasissystem

$$x_0^{i+1} = x_0^i + M_{0,ns}^{-1}(f_0 - T_0 x_0^i), \quad (i = 0, 1, 2, ...), \tag{4.11}$$

wobei $M_{0,ns}^{-1}$ mit Hilfe einer über mehrere Skalen fortgesetzten approximativen Faktorisierung der in T_0 implizit enthaltenen Grobgitteroperatoren erklärt ist. Dies kann auch als approximative Faktorisierung der Nichtstandardform von T_0 interpretiert werden. Im Fall $lt = 2$ verdeutlicht die nachfolgende Skizze das Vorgehen.



Dem damit erklärten Matrixsplitting der Nichtstandardform

$$\tilde{T}_{0,ns} = \tilde{L}_{0,ns}\tilde{U}_{0,ns} + \tilde{R}_{0,ns} \tag{4.12}$$

entspricht über die konvertierten Matrizen $\tilde{L}_{0,s} = (\tilde{L}_{0,ns})_s$ und $\tilde{U}_{0,s} = (\tilde{U}_{0,ns})_s$ ein äquivalentes Splitting der Form (4.5) im Standardsystem, das auf einer approximativen Faktorisierung von $\tilde{T}_{0,s}$ mittels Block-unterer und Block-oberer Standardform-Matrizen beruht. Wir erinnern dazu daran, daß bei der Konvertierung einer Nichtstandardform in ihre Standardform lediglich

die Nebendiagonalblöcke B_k und C_k für $k = 1, ..., lt$ jeweils weiter zerlegt werden. Die Implementierung eines Lösers auf der Basis des Splittings (4.12) ist aber effizienter als mittels des dazu äquivalenten Splittings der Standardform, da die Nichtstandardform die Fingerstruktur der auf Standardform transformierten Operatoren vermeidet (siehe Kapitel 3.3).

Durch das geschickte Anwenden der levelweisen Zweiskalen-Transformationen $W^k_{k+1,\mathcal{V}}$ auf der Ansatz- und der transponierten Zweiskalen-Transformationen $(W^k_{k+1,\mathcal{S}})^t$ auf der Testseite kann man ausgehend von der Nichtstandardform $\tilde{T}_{0,ns}$ Algorithmen für eine Multiskalen-Vorwärts- und -Rückwärtssubstitution finden, die das explizite Arbeiten im Nichtstandardsystem vermeiden. Sie werden durch die nachfolgenden beiden Prozeduren beschrieben. Der Übersichtlichkeit halber verzichten wir auf die Übergabe von T_0, $\tilde{T}_{0,ns}$ sowie der Folgen der Zweiskalen-Transformationen $\{W^{k-1}_{k,\mathcal{V}}\}_{k=1,...lt}$ und $\{W^{k-1}_{k,\mathcal{S}}\}_{k=1,...lt}$ als Parameter.

Algorithmus 1a (Multiskalen-Vorwärtssubstitution)

`function` $[d_{1,v}, ..., d_{lt,v}, s_{lt,v}] = \text{ms\_vorsubst}(lt, x^i_0, f_0)$

1. $s^{res}_0 = f_0 - T_0 x^i_0$; (Berechnen des Residuums)
2. `for` $k = 1 : lt$
   (a) $[d^{res}_k; s^{res}_k] = (W^{k-1}_{k,\mathcal{S}})' s^{res}_{k-1}$; (Zerlegen des Residuums)
   (b) $d_{k,v} = A^{-1}_k d^{res}_k$; (Vorkorrektur)
   (c) $s^{res}_k = s^{res}_k - C_k d_{k,v}$; (Modifikation des grobskaligen Residuen-Anteils)

   `end`
3. $s_{lt,v} = T^{-1}_{lt} s^{res}_{lt}$; (Korrektur der gröbsten Komponente)

Algorithmus 1b (Multiskalen-Rückwärtssubstitution)

`function` $s_{0,r} = \text{ms\_rücksubst}(lt, d_{1,v}, ..., d_{lt,v}, s_{lt,v})$

1. $s_{lt,r} = s_{lt,v}$;
2. `for` $k = lt : -1 : 1$
   (a) $d_{k,r} = d_{k,v} - A^{-1}_k B_k s_{k,r}$; (Nachkorrektur)
   (b) $s_{k-1,r} = W^{k-1}_{k,\mathcal{V}}[d_{k,r}; s_{k,r}]$; (Rücktransformation)

   `end`

Wir erhalten nach Anwenden dieser beiden Prozeduren die neue Iterierte durch

$$x^{i+1}_0 := x^i_0 + s_{0,r}. \tag{4.13}$$

Würde man bei der Konstruktion der Faktoren $\tilde{L}_{0,ns}$ und $\tilde{U}_{0,ns}$ der Nichtstandardform jeweils anstelle der Grobgitteroperatoren T_k die entsprechenden Schur-Komplement Operatoren S_k als Ausgangspunkt für die nächste einfache Zerlegung verwenden, so erhielte man hierdurch einen direkten Multiskalen-Löser. Man rechnet leicht nach, daß Algorithmus 1a das gleiche Ergebnis liefert wie eine Vorwärtssubstitution mittels $\tilde{L}^{-1}_{0,s}$ angewandt auf $\tilde{r}_{0,s}$, den Residuenvektor $r_0 = f_0 - T_0 x^i_0$ in seiner Momentendarstellung bezüglich der Zerlegung (3.3). Wendet man im Anschluß daran auf das Ergebnis eine Rückwärtssubstitution mittels $\tilde{U}^{-1}_{0,s}$ an und transformiert den so erhaltenen Vektor in Multiskalen-nodaler Darstellung relativ zur Zerlegung (3.2) zurück in seine Feingitterdarstellung, so erhält man das gleiche Resultat wie durch Algorithmus 1b.

Sukzessive Unterraumkorrektur mittels Nichtstandardform

Exakt die gleichen Iterierten erhalten wir durch den Aufruf

$$x_0^{i+1} = \text{ms\_v\_zyklus}(lt, 1, x_0^i, f_0),$$

wobei die Prozedur "ms_v_zyklus" wie folgt erklärt ist:

Algorithmus 2 (MULTISKALEN-V$(1,1)$-ZYKLUS)
`function` $x = \text{ms\_v\_zyklus}(lt, k, x, f)$;

1. $[d^{res}; s^{res}] = (W_{k,\mathcal{S}}^{k-1})'(f - T_{k-1}x)$; (Zerlegen des Residuums)
2. $d_v = A_k^{-1} d^{res}$; (Vorkorrektur)
3. $s_m^{res} = s^{res} - C_k d_v$; (Modifikation des grobskaligen Residuen-Anteils)
4. `if` $(k == lt)$
   $s_v = T_{lt}^{-1} s_m^{res}$;
   `else`
   $s_v = \text{ms\_v\_zyklus}(lt, k+1, \mathbf{0}, s_m^{res})$;
5. $d_n = A_k^{-1}(d^{res} - A_k d_v - B_k s_v)$; (Nachkorrektur)
6. $x = x + W_{k,\mathcal{V}}^{k-1}[d_v + d_n; s_v]$; (Rücktransformation und Korrektur)

Wir nehmen hierbei an, daß die Vektoren x, f und $\mathbf{0}$ stets die passende Größe besitzen. Für $k = 1$ stimmen die Schritte 1-3 offensichtlich mit den Schritten 1 und 2.a - 2.c aus der Vorwärtssubstitutionsroutine überein. Letztere zerlegt dann im Rahmen einer for-Schleife sukzessive die modifizierten grobskaligen Anteile weiter, um Korrekturen bezüglich der Detail-Anteile und schließlich auf dem gröbsten Level eine Grobgitterkorrektur zu berechnen.
Genau die gleichen Zerlegungen werden bei den rekursiven Aufrufen von "ms_v_zyklus" gemacht, da hier jeweils als rechte Seite der gleiche modifizierte grobskalige Anteil zusammen mit einem entsprechend dimensionierten Nullvektor als Startapproximation übergeben wird. Man gelangt so innerhalb der Rekursion bis zum gröbsten Level lt, wo man die gleiche Grobgitterkorrektur berechnet wie in Algorithmus 1a.
Die Multiskalen-Rückwärtssubstitution durch Algorithmus 1b besteht aus einer Folge von Nachkorrekturen bezüglich der Detail-Anteile, die zusammen mit der jeweils aktuellen Grobskalenkorrektur durch Rücktransformation sukzessive auf die nächstfeinere Skala transportiert werden, um schließlich in einer Feingitterkorrektur der bisherigen Iterierten einzumünden.
Wegen $d_v + d_n = d_v + A_k^{-1}(d^{res} - A_k d_v - B_k s_v) = d_v - A_k^{-1} B_k s_v$ werden im Falle exakter Invertierung der A_k in Schritt 5 der rekursiven Prozedur genau die gleichen Nachkorrekturen bezüglich der Detail-Anteile bestimmt wie mit Algorithmus 1b. Sie werden nämlich jeweils zusammen mit der bisherigen Grobskalenkorrektur durch Rücktransformation auf die nächstfeinere Skala transportiert und direkt zu den Startapproximationen aufaddiert, die bis auf die Anfangsapproximation x_0^i alle Null sind. Die Feingitterkorrektur (4.13) der bisherigen Iterierten geschieht damit automatisch. Es gilt also der folgende Satz.

Satz 9 (Äquivalenz der approximativen Multiskalen-Faktorisierung und des Multiskalen-V(1,1)-Zyklus)
Im Falle exakter Korrekturen bezüglich der Detail-Komponenten, d.h. exakter Invertierung der Blöcke A_k für $k = 1, ..., lt$, sind die Iteration mittels der approximativen Multiskalen-Faktorisierung der Steifigkeitsmatrix und der Multiskalen-$V(1,1)$-Zyklus äquivalent.

Beweis:
Der Beweis folgt direkt aus der unmittelbar vorangehenden Argumentation.

□

Im Fall lediglich approximativer Nachkorrekturen sollte der Multiskalen-V(1, 1)-Zyklus aufgrund der gleichen Argumente wie bei der Diskussion der Verfahren für die einfache Zerlegung dann bessere Ergebnisse liefern als die Iteration mittels der approximativen Multiskalen-Faktorisierung der Nichtstandardform. Die während der Zerlegung des Ausgangsproblems über mehrere Skalen entstehenden Teilprobleme sind selbst jedoch nur Approximationen der Feingittergleichung. Dies kann dazu führen, daß in gewissen Fällen die erste Variante (Algorithmen 1a, 1b) trotzdem etwas besser abschneidet als die zweite (Algorithmus 2).

Wir zeigen nun, daß unser Multiskalen-V(1,1)-Zyklus auch als gewöhnlicher Mehrgitter-V(1,1)-Zyklus interpretiert werden kann, der eine spezielle Glättungsiteration auf den einzelnen Leveln benutzt. In diesem Zusammenhang sprechen wir dann auch von einem *Multiskalen-Glätter*, der zu dem Multiskalen-Verfahren gehört. Ein Glättungsschritt für ein gewöhnliches Mehrgitter-Verfahren ist auf dem feinsten Level zumeist durch eine klassische Iteration der Form

$$x_0^{i+\frac{1}{3}} = x_0^i + M_0^{-1}(f_0 - T_0 x_0^i) \tag{4.14}$$

gegeben. Ist M_0 gleich dem unteren Dreieck von T_0, so erhalten wir eine Gauß–Seidel-Glättung, ist M_0 gleich der Diagonalen von T_0, so definiert dies eine Jacobi-Glättung. Die Schritte 1 und 2 des Multiskalen-V(1,1)-Zyklus können nun zusammen wie folgt als spezieller Vorglättungsschritt in einem gewöhnlichen Mehrgitter-Verfahren aufgefaßt werden. Das Feingittersystem wird zunächst in zwei Komponenten für die jeweils rauhen und glatten Anteile der Iterationsfehler aufgespalten, was einfach durch eine Zweiskalen-Transformation des Gleichungssystems geschieht. Eine Korrektur wird dann in Schritt 2 nur bezüglich der rauhen Komponente durch $d_v = \breve{A}_1^{-1}(Q_{1,\mathcal{S}}^0)^t(f_0 - T_0 x_0^i)$ (approximativ) berechnet. Wir transformieren nun diese Korrektur mittels $Q_{1,\mathcal{V}}^0$ zurück in die nodale Darstellung bezüglich $\mathcal{V}_0$ und addieren sie zur bisherigen Iterierten, um zu sehen, was sie aus Sicht des Feingittersystems bedeutet. Man erhält

$$x_0^{i+\frac{1}{3}} = x_0^i + Q_{1,\mathcal{V}}^0 \breve{A}_1^{-1}(Q_{1,\mathcal{S}}^0)^t(f_0 - T_0 x_0^i) \tag{4.15}$$

und wir können damit (4.15) als einen speziellen Glätter ansehen. Im klassischen Mehrgitter-Verfahren wird nun das nach der Glättung entstehende neue und vorzugsweise glattere Residuum zur Grobgitter-Korrektur auf das nächstgröbere Level restringiert. Wir erhalten, indem wir die Abhängigkeit des Residuums von der Iterierten explizit kennzeichen,

$$\begin{aligned} s_1^{res,x_0^{i+\frac{1}{3}}} &= (P_{1,\mathcal{S}}^0)^t(f_0 - T_0 x_0^{i+\frac{1}{3}}) \\ &= s_1^{res,x_0^i} - (P_{1,\mathcal{S}}^0)^t T_0 Q_{1,\mathcal{V}}^0 \breve{A}_1^{-1}(Q_{1,\mathcal{S}}^0)^t(f_0 - T_0 x_0^i) \\ &= s_1^{res,x_0^i} - C_1 d_v. \end{aligned}$$

Genau das gleiche leistet auch Schritt 3 unseres Multiskalen-Zyklus. Der Grobgitterkorrekturschritt 4 unterscheidet sich offensichtlich nicht von demjenigen im gewöhnlichen Mehrgitter-Verfahren, und der Nachkorrekturschritt 5 kann wiederum über unseren speziellen Glätter (4.15) als Nachglättungsschritt verstanden werden. Da die gleichen Identifizierungen und Interpretationen ebenfalls auf den gröberen Leveln gelten, zeigt dies die Äquivalenz des Multiskalen-V(1,1)-Zyklus zu einem gewöhnlichen Mehrgitter-V(1,1)-Zyklus, der spezielle Glätter auf den einzelnen Leveln verwendet.

Satz 10 (Äquivalenz des Multiskalen-V(1,1)-Zyklus zu einem Mehrgitter-V(1,1)-Zyklus)
Der über Algorithmus 2 definierte Multiskalen-$V(1,1)$-Zyklus, bei dem die Korrekturen bezüglich der Detail-Anteile approximativ mittels $\breve{A}_k^{-1}$ für $k = 1, ..., lt$ erfolgen, ist äquivalent zu einem gewöhnlichen Mehrgitter-$V(1,1)$-Zyklus, der auf jedem Level $k-1$ für $k = 1, ..., lt$ die Matrizen $M_{k-1}^{-1} := Q_{k,\mathcal{V}}^{k-1} \breve{A}_k^{-1} (Q_{k,\mathcal{S}}^{k-1})^t$ zur Glättungsiteration benutzt.

Beweis:
Der Beweis folgt direkt aus der unmittelbar vorangehenden Argumentation.

□

Zur Veranschaulichung des Glättungsprozesses betrachten wir ein Poissonproblem mit homogenen Dirichletschen Randbedingungen auf dem Einheitsquadrat und der Lösung Null, diskretisiert durch einen 5-Punkte Stern und zentrale Differenzen auf einem Gitter mit Maschenweite $h_0 = \frac{1}{32}$. Wir berechnen die Residuenvektoren, die sich ausgehend von einem jeweils gleichen zufälligen Startvektor für unterschiedliche Multiskalen-Zerlegungen nach einem Multiskalen-Vorglättungsschritt (4.15) auf dem feinsten Level ergeben. Es wird $\breve{A}_1^{-1}$ mittels einer unvollständigen Zerlegung bezüglich des Besetzungsmusters von A_1 (ILU(0)) über eine Vorwärts- und Rückwärtssubstitution realisiert. Zum Vergleich berechnen wir noch die entsprechenden Residuenvektoren, die man durch den gewöhnlichen Glättungsschritt (4.14) auf dem feinsten Level mit Hilfe eines Gauß–Seidel Verfahrens und einer über ILU(0) definierten Iteration erhält. Die jeweiligen Residuen sind als Gitterfunktionen in Abbildung 4.1 dargestellt.

Implementierung ohne explizite Berechnung der Nichtstandardform

Die beiden vorgestellten multiplikativen Multiskalen-Verfahren benötigen lediglich die Anwendung der Nebendiagonalblöcke B_k und C_k auf entsprechend dimensionierte Vektoren. Wegen der Beziehungen

$$B_k = (Q_{k,\mathcal{S}}^{k-1})^t T_{k-1} P_{k,\mathcal{V}}^{k-1} \quad \text{und} \quad C_k = (P_{k,\mathcal{S}}^{k-1})^t T_{k-1} Q_{k,\mathcal{V}}^{k-1} \tag{4.16}$$

für $k = 1, ..., lt$ ist daher eine effiziente Implementierung der soeben besprochenen Algorithmen auch ohne das explizite Aufstellen der vollständigen Nichtstandardform möglich. Berechnet man (approximativ) die Aktionen der Inversen der Blöcke A_k mit Hilfe eines Iterationsverfahrens, bei dem nur die Anwendung der A_k auf entsprechend dimensionierte Vektoren verlangt wird, so kann man dazu auch die Darstellungen $A_k = (Q_{k,\mathcal{S}}^{k-1})^t T_{k-1} Q_{k,\mathcal{V}}^{k-1}$ für $k = 1, ..., lt$ benutzen.

Die bislang in der Literatur gemachten Vorschläge zur Verbesserung von Multilevel- und Multiskalen-Algorithmen zielen zumeist darauf hin, die Approximation der bei einer exakten Block-Faktorisierung eigentlich zu verwendenden Schur-Komplement Operatoren S_k durch

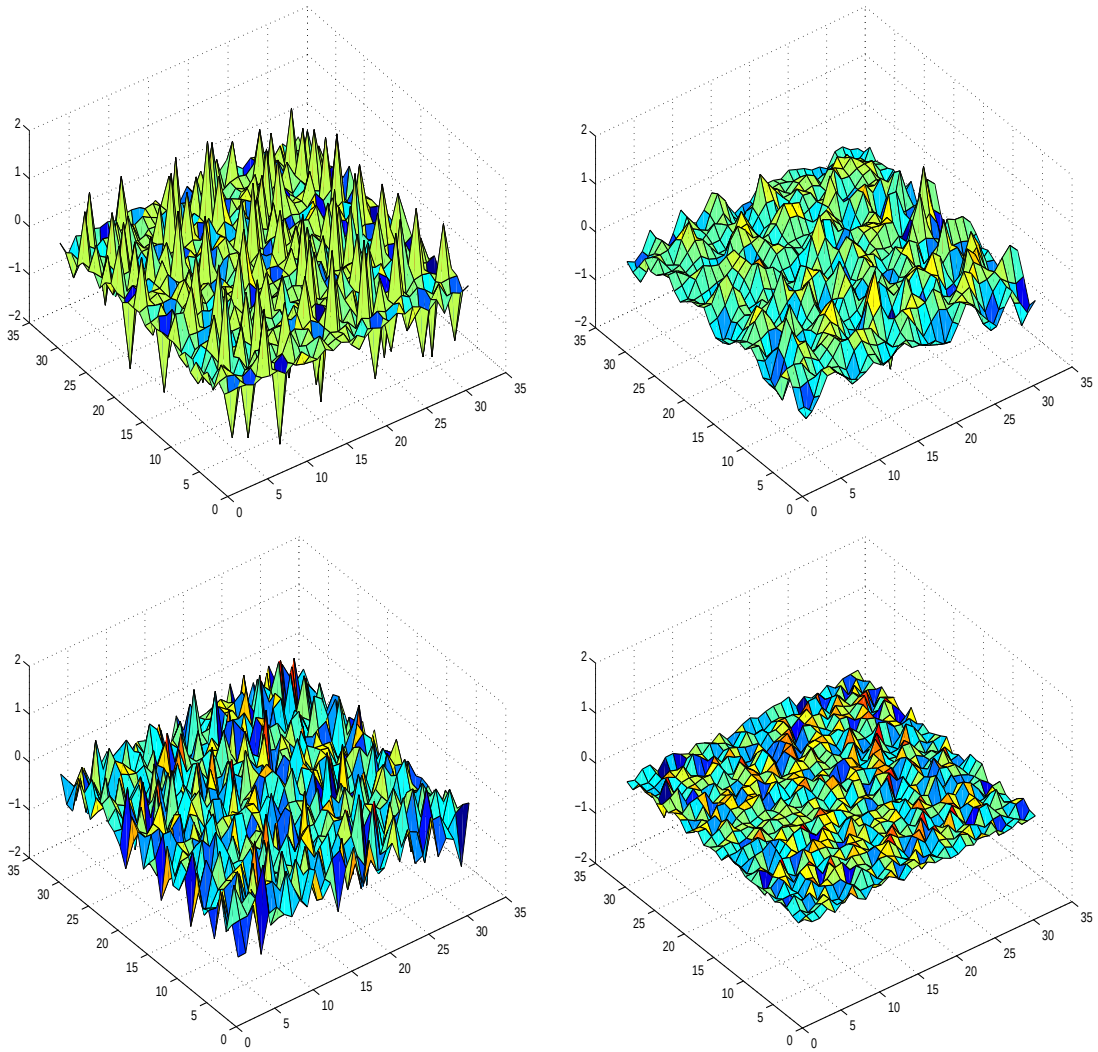


ABBILDUNG 4.1: *Obere Bilder: Feingitter-Residuum nach einem Multiskalen-Glättungsschritt; klassische hierarchische Basis Zerlegung (oben links) und Prewavelet-Zerlegung (oben rechts). Untere Bilder: Feingitter-Residuum nach einem Mehrgitter-Glättungsschritt; Gauß–Seidel (unten links) und ILU(0) (unten rechts) Laplace 5-Punkte Stern,* $h_0 = \frac{1}{32}$, $f_0 = \mathbf{0}$, *zufälliger Startvektor.*

die Grobgitteroperatoren T_k zu verbessern. Das bedeutet konkret, letztere so zu modifizieren, daß ihr Verhalten dem der exakten Schur-Komplement Operatoren genauer entspricht. Oft führt eine solche Anpassung zu Grobgitteroperatoren mit größer werdenden Sternbreiten, was das rekursive Fortsetzen solcher Prozeduren schwierig macht, wenn der Aufwand für die Anwendung des Algorithmus proportional zur Zahl der Feingitterunbekannten bleiben soll. Man erreicht eine Verbesserung aber auch implizit durch die geschickte Wahl der Multiskalen-Transformationen (siehe Kapitel 5 dieser Arbeit sowie [1, 37, 38, 54, 132, 133]) oder auch durch zusätzliche innere Rekursionen. In [5, 6] werden zum Beispiel rekursiv definierte polynomiale Schur-Komplement Approximationen zur Stabilisierung des ersten Verfahrens im Fall der klassischen hierarchischen Basis Zerlegung für zwei- und dreidimensionale symmetrisch positiv definite elliptische Probleme vorgeschlagen. Sie führen zu HBMG-artigen Verfahren mit einer allgemeineren Zyklus-Struktur als der des gewöhnlichen V-Zyklus und im

Fall des Laplace-Operators zu gitterweitenunabhängigen Konvergenzraten. Der gleiche Effekt wird auch mit Hilfe eines hierarchischen W-Zyklus erzielt, der für solche Probleme ebenfalls zu optimalen Verfahren führt [54].

Verbesserte Varianten der vorgestellten Verfahren ergeben sich möglicherweise aber auch dadurch, daß man für $k = 2, ..., lt$ in den Formeln (4.16) Schur-Komplement Approximationen $\breve{S}_{k-1} = T_{k-1} - C_{k-1}\breve{A}_{k-1}^{-1}B_{k-1}$ statt der Operatoren T_{k-1} verwendet. Im Fall des Multiskalen-V(1,1)-Zyklus werden zusätzlich noch die $\breve{S}_{k-1}$ zur Berechnung der Residuen in Schritt 1 benutzt. Man nutzt also Schur-Komplement Approximationen nur bei der Berechnung der verallgemeinerten hierarchischen Residuen und nicht darüberhinaus zur weiteren Zerlegung. Damit werden feinskalige Einflüsse unter Umständen besser auf die gröberen Skalen transportiert. Einen interessanten Ansatz, der darüberhinaus explizite Schur-Komplement Approximationen ebenfalls für die weiteren Zerlegungsschritte verwendet, findet man in der Arbeit [89]. Hierbei werden die Multiskalen-Zerlegungen aber lediglich durch eine Multilevel-artige Anordnung der Freiheitsgrade induziert.

Definiert man $\breve{S}_0 := T_0$, so erhalten wir für $k = 1, ..., lt$ anstelle von (4.16) die Darstellungen

$$B_k^S = (Q_{k,\mathcal{S}}^{k-1})^t \breve{S}_{k-1} P_{k,\mathcal{V}}^{k-1} \quad \text{und} \quad C_k^S = (P_{k,\mathcal{S}}^{k-1})^t \breve{S}_{k-1} Q_{k,\mathcal{V}}^{k-1}. \tag{4.17}$$

Die sich darüber ergebenden Varianten der Algorithmen sind unseres Wissens neu und in dieser Form bislang nicht in der Literatur bekannt.

Algorithmus 1a' (Verbesserte Multiskalen-Vorwärtssubstitution)
`function` $[d_{1,v}, ..., d_{lt,v}, s_{lt,v}] = \text{ms\_vorsubst}(lt, x_0^k, f_0)$

1. $s_0^{res} = f_0 - T_0 x_0^i$
2. `for` $k = 1 : lt$
   (a) $[d_k^{res}; s_k^{res}] = (W_{k,\mathcal{S}}^{k-1})' s_{k-1}^{res}$;
   (b) $d_{k,v} = \breve{A}_k^{-1} d_k^{res}$;
   (c) $s_k^{res} = s_k^{res} - (P_{k,\mathcal{S}}^{k-1})' \breve{S}_{k-1} Q_{k,\mathcal{V}}^{k-1} d_{k,v}$;

   `end`
3. $s_{lt,v} = T_{lt}^{-1} s_{lt}^{res}$;

Algorithmus 1b' (Verbesserte Multiskalen-Rückwärtssubstitution)
`function` $s_{0,r} = \text{ms\_rücksubst}(lt, d_{1,v}, ..., d_{lt,v}, s_{lt,v})$

1. $s_{lt,r} = s_{lt,v}$;
2. `for` $k = lt : -1 : 1$
   (a) $d_{k,r} = d_{k,v} - \breve{A}_k^{-1} (Q_{k,\mathcal{S}}^{k-1})' \breve{S}_{k-1} P_{k,\mathcal{V}}^{k-1} s_{k,r}$;
   (b) $s_{k-1,r} = W_{k,\mathcal{V}}^{k-1} [d_{k,r}; s_{k,r}]$;

   `end`

Im Fall des Multiskalen-V(1,1)-Zyklus werden außerdem die $\breve{S}_{k-1}$ zur Berechnung der Residuen in Schritt 1 verwendet. Für den rekursiven Algorithmus mit einer allgemeineren μ-fachen Zyklus-Struktur (μ = Rekursionstiefe) und approximativen Detail-Korrekturen erhalten wir dann mit den gleichen Berechnungen der Komponenten der verallgemeinerten hierarchischen Residuen:

Algorithmus 2' (Allgemeiner Multiskalen-Zyklus)

`function` $x = \text{ms\_zyklus}(lt, k, x, f, \mu)$;

1. $[d^{res}; s^{res}] = (W_{k,\mathcal{S}}^{k-1})'(f - \breve{S}_{k-1}x)$;
2. $d_v = \breve{A}_k^{-1} d^{res}$;
3. `for` $z = 1 : \mu$
   (a) $s_v = \mathbf{0}$;
   (b) $s_m^{res} = s^{res} - (P_{k,\mathcal{S}}^{k-1})'\breve{S}_{k-1} W_{k,\mathcal{V}}^{k-1}[d_v; s_v]$;
   `if` $(k == lt)$
   (c) $s_h = T_{lt}^{-1} s_m^{res}$;
   `else`
   (d) $s_h = \text{ms\_zyklus}(lt, k+1, \mathbf{0}, s_m^{res}, \mu)$;
   `end`
   (e) $s_v = s_v + s_h$;
   `end`
4. $d_n = \breve{A}_k^{-1}(d^{res} - (Q_{k,\mathcal{S}}^{k-1})'\breve{S}_{k-1} W_{k,\mathcal{V}}^{k-1}[d_v; s_v])$;
5. $x = x + W_{k,\mathcal{V}}^{k-1}[d_v + d_n; s_v]$;

In Kapitel 6 werden die Kosten dieser verbesserten Varianten diskutiert. Es zeigt sich, daß diese recht hoch sind. Die Kosten für einen verbesserten Multiskalen-V-Zyklus können beispielsweise höher liegen als für einen gewöhnlichen Multiskalen-W-Zyklus. Eine vollständig rekursive Berechnung der Schur-Komplement Approximationen $\breve{S}_{k-1} = T_{k-1} - C_{k-1}\breve{A}_{k-1}^{-1}B_{k-1}$, die statt B_{k-1} und C_{k-1} wiederum Darstellungen für B_{k-1}^S und C_{k-1}^S verwendet, ist ebenfalls möglich. Sie führt allerdings zu deutlich höheren Kosten. Man erhält hierdurch außerdem keine merkliche Verbesserung der Verfahren, da im Verlauf der approximativen Multiskalen-Faktorisierungen weiterhin die Grobgitteroperatoren als Schur-Komplement Approximationen zur weiteren Zerlegung verwendet werden.

Werden im Multiskalen-Zyklus Korrekturen der Detail-Komponenten bezüglich der Skala k berechnet, so finden diese aus Sicht des entsprechenden Mehrgitter-Verfahrens auf dem Level $k-1$ statt. Die Korrektur der Detail-Komponente sowie des grobskaligen Anteils bezüglich der gröbsten Skala lt geschieht daher in der Mehrgitter-Interpretation auf den Leveln $lt-1$ respektive lt. Mit diesen Bemerkungen geben wir abschließend als äquivalente Formulierung von Algorithmus 2' einen gewöhnlichen Mehrgitter-Zyklus an, der gemäß Satz 10 auf jedem Level $k-1$ für $k = 1, ..., lt$ die Matrizen $M_{k-1}^{-1} = Q_{k,\mathcal{V}}^{k-1}\breve{A}_k^{-1}(Q_{k,\mathcal{S}}^{k-1})^t$ zur Glättungsiteration benutzt.

Algorithmus 3 (Multiskalen-Zyklus als Mehrgitter-Verfahren)

`function` $x = \text{ms\_zyklus}(lt, k, x, f, \mu)$;

`if` $(k == (lt+1))$

$x = T_{lt}^{-1} f$;

`else`

$x = x + Q_{k,\mathcal{V}}^{k-1} \breve{A}_k^{-1} (Q_{k,\mathcal{S}}^{k-1})'(f - \breve{S}_{k-1} x)$;

$s^{res} = (P_{k,\mathcal{S}}^{k-1})'(f - \breve{S}_{k-1} x)$;

$s_v = \mathbf{0}$;

`for` $z = 1 : \mu$

$s_v = \text{ms\_zyklus}(lt, k+1, s_v, s_m^{res}, \mu)$;

`end`

$x = x + P_{k,\mathcal{V}}^{k-1} s_v$;

$x = x + Q_{k,\mathcal{V}}^{k-1} \breve{A}_k^{-1} (Q_{k,\mathcal{S}}^{k-1})'(f - \breve{S}_{k-1} x)$;

`end`

4.2.3 Vergleich mit ähnlichen Methoden

Wir vergleichen in diesem Abschnitt die vorgestellten Verfahren mit drei verwandten Ansätzen aus der Literatur, die die Entwicklung unserer Methoden beeinflußt haben. Es sind dies eine in [48, 49] vorgeschlagene schnelle Wavelet-basierte LU-Faktorisierung, die Multilevel-Vorkonditionierung mittels AMLI gemäß [5, 6] sowie die Methode der Approximativen Zyklischen Reduktion [96, 97].

Schnelle Wavelet-basierte LU-Faktorisierung

In [48, 49] wurde eine schnelle Wavelet-basierte LU-Faktorisierung als direkter Löser für lineare Gleichungssysteme vorgestellt, die durch Diskretisierung von Integralgleichungen mit N_0 Freiheitsgraden entstehen. Das Bestimmen der Lösung solcher Gleichungssysteme ist dabei bis auf eine Genauigkeit ε mit $\mathcal{O}(N_0)$ Aufwand möglich. Dies ist umso erstaunlicher, wenn man bedenkt, daß aufgrund der Nichtlokalität von Integraloperatoren der Aufwand zum Aufstellen der nodalen Steifigkeitsmatrix schon quadratisch mit der Anzahl der Freiheitsgrade wächst [63]. Das Verfahren eignet sich ebenfalls zur Lösung linearer Gleichungssysteme, die von Diskretisierungen partieller Differentialgleichungen herrühren. Die Frage, ob man mit der in [48, 49] vorgeschlagenen Methode zu robusten Lösern für Konvektions-Diffusions Probleme gelangt, wird in [107] genauer untersucht. Wir beschränken uns in der nachfolgenden Darstellung auf eindimensionale Probleme. Erweiterungen auf höherdimensionale Fälle sind aber möglich [31, 48, 107] und werden gegen Ende des Abschnitts diskutiert. Wir erörtern dann ebenfalls Schwierigkeiten, die mit dieser Methode bei der Lösung von Konvektions-Diffusions Gleichungen auftreten.

Ausgehend von einer orthogonalen Wavelet-Basis des $\mathcal{L}_2(\mathbb{R})$ nutzt die in [48, 49] vorgeschlagene Methode die Tatsache, daß Integraloperatoren T der Form

$$T(f)(x) = \int_{\mathbb{R}} K(x,y) f(y)\, dy \tag{4.18}$$

nach *Kompression* eine schwachbesetzte Darstellung über ihre Nichtstandardform besitzen [19]. Diese Aussage bleibt beim Übergang von den kontinuierlichen zu den diskretisierten Operatoren T_0 erhalten. Eine theoretische Grundlage hierfür bietet der nachfolgende Satz, der eine Aussage über das Abfallverhalten der Matrixeinträge der Nichtstandardform von Integraloperatoren im Ganzraumfall macht. Die Beschränkung auf den Ganzraumfall ergibt sich aus der folgenden Überlegung. Die Matrizen A_k, B_k, C_k und T_k haben dann ungeachtet einer Diskretisierung unendlich viele Einträge. Wir sagen auch, sie sind *bi-unendlich*. Gehört der kontinuierliche Operator T zu einem Randwertproblem bezüglich eines Gebiets Ω, so sind diese Blöcke nach Diskretisierung und Transformation Matrizen endlicher Größe. Für solche endlichen Matrizen ist die quantitative Beschreibung des Abfallverhaltens ihrer Einträge in Abhängigkeit von der Entfernung von der Diagonalen nicht ohne weiteres möglich, da es sich dabei um asymptotische Aussagen handelt.

Satz 11 (Abfall der Nichtstandardform-Einträge von Integraloperatoren)
Es sei T ein Integraloperator der Form (4.18) auf $\mathcal{L}_2(\mathbb{R})$ mit Kernfunktion $K(x,y)$, so daß die Abschätzungen

$$|K(x,y)| \leq |x-y|^{-1}, \tag{4.19}$$

$$|\partial_x^m K(x,y)| + |\partial_y^m K(x,y)| \leq C|x-y|^{-m-1} \tag{4.20}$$

für eine natürliche Zahl m erfüllt sind. Ist der durch $K(x,y)$ definierte Operator unbeschränkt, so gelte für beliebige dyadische Intervalle $I_j^k := [j2^{-k}, (j+1)2^{-k}[\subset \mathbb{R}$ mit $j,k \in \mathbb{Z}$ zudem die Abschätzung

$$|\int_{I_j^k \times I_j^k} K(x,y)dx\,dy| \leq C|I_j^k|. \tag{4.21}$$

Hat die verwendete Wavelet-Basis zur Zerlegung von $\mathcal{L}_2(\mathbb{R})$ m verschwindende Momente, das heißt für das zugehörige Wavelet ψ gilt $\int_{\mathbb{R}} x^p \psi(x)\,dx = 0$ für alle $p = 0, ..., m-1$, so erhält man für die Einträge $a_{i,j}^k$, $b_{i,j}^k$ und $c_{i,j}^k$ der Blöcke A_k, B_k und C_k der Nichtstandardform von T die Abschätzung

$$|a_{i,j}^k| + |b_{i,j}^k| + |c_{i,j}^k| \leq C(1+|i-j|)^{-m-1}. \tag{4.22}$$

Die Konstante $C = C(k,m)$ hängt ebenfalls von der jeweiligen Kernfunktion $K(x,y)$ ab.

Beweis: Siehe [19, 83].

□

Dieser Satz gilt ebenfalls für die allgemeinere Klasse von Calderón–Zygmund-Operatoren [83, 84]. Aufgrund des in Abschätzung (4.22) aufgezeigten Abfallverhaltens, ist es für numerische Zwecke oft ausreichend, mit *Bandapproximationen* von diskretisierten Integraloperatoren zu rechnen. Sie werden mit einem hochgestellten Index ^{B} gekennzeichnet. Bandapproximationen, die durch Vernachlässigen aller Einträge der Matrix außerhalb eines Bandes der Weite $w \in \mathbb{N}$ rechts und links um die Hauptdiagonale entstehen, werden auch mit dem hochgestelltem Index ^{2w} versehen. Wir approximieren nun die vollständig zerlegte Nichtstandardform von T_0 ($N_0 = 2^{lt}$) durch eine Bandapproximation $\tilde{T}_{0,ns}^B = \tilde{T}_{0,ns}^{2w}$, die nur Einträge der Blöcke A_k, B_k und C_k für $k = 1, ..., lt$ von $\tilde{T}_{0,ns}$ innerhalb Bänder der Weite $w \geq 2m$ rechts und links um deren Diagonalen berücksichtigt. Man erhält dann aufgrund von (4.22) die fast optimale Abschätzung des Kompressionsfehlers

$$|||\tilde{T}_{0,ns} - \tilde{T}_{0,ns}^B|||_\infty \leq \frac{C}{w^m} \log_2 N_0. \tag{4.23}$$

Hierbei ist $|||\cdot|||_\infty$ definiert als

$$|||\tilde{T}_{0,ns}|||_\infty := \sum_{k=1}^{lt}(||A_k||_\infty + ||B_k||_\infty + ||C_k||_\infty) + ||T_{lt}||_\infty. \tag{4.24}$$

In [19] wird gezeigt, daß Abschätzung (4.23) auch ohne den logarithmischen Faktor gilt und damit durch Wahl der Bandweite $w \geq (C/\varepsilon)^{\frac{1}{m}}$ eine Bandapproximation der diskreten Nichtstandardform mit Genauigkeit ε möglich ist. Hierzu verwenden die Autoren ein Splitting $T = L + M + R$ gemäß dem $T(1)$-Theorem von David und Journé [83], wobei L das Para-Produkt mit einer Funktion $\beta \in \mathcal{BMO}$, M die adjungierte Abbildung eines solchen Para-Produkts und R eine sogenannte Pseudo-Konvolution darstellt. Für eine diskretisierte Version des David–Journé-Splittings findet man, daß der entsprechende Operator R_0 optimal kompressibel ist und durch eine Bandapproximation mit $\mathcal{O}(N_0)$ Einträgen in der Nichtstandardform hinreichend gut approximiert werden kann. Die (adjungierten) diskreten Para-Produkte führen als bilineare Funktionale nach Kompression zu Matrizen mit ebenfalls nur $\mathcal{O}(N_0)$ Einträgen. Man erhält über den Umweg des David–Journé-Splittings von T, daß die oben betrachtete Bandapproximation $\tilde{T}^B_{0,ns}$ sogar die optimale Abschätzung

$$|||\tilde{T}_{0,ns} - \tilde{T}^B_{0,ns}|||_\infty \leq \frac{C}{w^m} \tag{4.25}$$

erfüllt.

Nach diesen theoretischen Aussagen über die Güte der Approximation von diskreten Operatoren mittels Bandapproximationen ihrer Nichtstandardform wollen wir uns nun mit der Frage beschäftigen, wie solche Bandapproximationen effizient, das heißt mit $\mathcal{O}(N_0)$ Aufwand, berechnet werden können. Für Differentialoperatoren ist die effiziente Transformation auf Nichtstandardform trivial möglich, wenn die Diskretisierung durch eine Finite-Elemente, Finite-Differenzen oder Finite-Volumen Methode erfolgt. Es ergeben sich so als Feinstskalendarstellungen schwachbesetzte Matrizen, und die Transformation dieser auf Nichtstandardform anhand der Transformationsmatrizen $P^k_{k+1,\mathcal{V}/\mathcal{S}}$ und $Q^k_{k+1,\mathcal{V}/\mathcal{S}}$ ist mit $\mathcal{O}(N_0)$ Aufwand möglich. Erfolgt die Diskretisierung eines Differentialoperators T über ein Petrov–Galerkin Finite-Elemente Verfahren mit Ansatz- und Testräumen $\mathcal{V}_0$ und $\mathcal{S}_0$ und zugehörigen Multiskalen-Analysen, die die Zweiskalen-Transformationen $W^k_{k+1,\mathcal{V}}$ und $W^k_{k+1,\mathcal{S}}$ definieren, so stimmen die Einträge der damit berechneten Blöcke A_k, B_k, C_k und T_k mit denjenigen überein, die man durch direkte Integration mittels der Funktionen des Nichtstandard-Erzeugendensystems der Ansatz- und Testseite erhält:

$$\begin{aligned}(A_k)_{i,j} &= a^k_{i,j} = (T\psi^k_{j,\mathcal{V}}, \psi^k_{i,\mathcal{S}}), \qquad (B_k)_{i,j} = b^k_{i,j} = (T\varphi^k_{j,\mathcal{V}}, \psi^k_{i,\mathcal{S}}),\\ (C_k)_{i,j} &= c^k_{i,j} = (T\psi^k_{j,\mathcal{V}}, \varphi^k_{i,\mathcal{S}}), \qquad (T_k)_{i,j} = t^k_{i,j} = (T\varphi^k_{j,\mathcal{V}}, \varphi^k_{i,\mathcal{S}}).\end{aligned}$$

Wir nennen diese Methode fortan die *direkte Transformation* auf Nichtstandardform im Unterschied zum bisher betrachteten *algebraischen* Verfahren, das mittels der Zweiskalen-Transformationen erfolgt. Wir diskutieren nun, inwieweit eine effiziente Transformation von Integraloperatoren auf Nichtstandardform mit diesen beiden Varianten möglich ist.

- Direkte Transformation von Integraloperatoren T auf Nichtstandardform

  Es werden hierbei die Einträge der Blöcke A_k, B_k und C_k für $k = 1, ..., lt$ sowie die Einträge von T_{lt} direkt durch Berechnen der Einträge der Steifigkeitsmatrix ausgehend von den Nichtstandard-Erzeugendensystemen für die Ansatz- und Testseite bestimmt.

Ist ε die Genauigkeit, mit der eine Approximation $\tilde{T}^B_{0,ns}$ von $\tilde{T}_{0,ns}$ berechnet werden soll, so existiert aufgrund von Abschätzung (4.25) eine Bandweite $w = w(\varepsilon, m, K)$, so daß dazu nur die Einträge innerhalb dieser Bandweite um die Diagonalen der jeweiligen Blöcke berechnet werden müssen. Der Aufwand hierfür beträgt $\mathcal{O}(wN_0)$.

- Algebraische Transformation von Integraloperatoren T auf Nichtstandardform

  Die algebraische Transformation eines Integraloperators T auf eine komprimierte Nichtstandarddarstellung ist in linearer Komplexität nur dann möglich, wenn die zugehörige Kernfunktion $K(x, y)$ neben den bereits geforderten Abfallgesetzen außerhalb der Diagonalen $x = y$ gewissen Glattheitsbedingungen genügt, durch die eine Interpolation gestattet wird. Wir skizzieren kurz das Vorgehen:

  Zunächst wird eine Approximation $T^B_0 = T^{2w}_0$ des Operators T_0 bezüglich der feinsten Skala aufgestellt, indem man nur Einträge von T_0 berücksichtigt, die innerhalb eines Bandes $[-w, w]$ der Gesamtweite $2w$ um die Hauptdiagonale liegen. Die Zweiskalen-Transformation von T^B_0 liefert nun die Blöcke A^B_1, B^B_1, C^B_1 sowie T^B_1, die ebenfalls Bandgestalt haben. In Abhängigkeit der Länge der verwendeten Skalierungsfilter verkleinert sich dabei die Ausgangsbandweite $2w$ für T^B_1. Wir nehmen der Einfachheit halber an, daß sich die Bandweite um den Faktor 2 reduziert, so daß also $T^B_1 = T^w_1$ nurmehr die Bandweite w besitzt. Um mit der Transformation fortzufahren, wird nun T^B_1 mit Hilfe einer Interpolationsformel symmetrisch zu einer Bandmatrix T^{2w}_1 der ursprünglichen Weite $2w$ erweitert. Wird diese Prozedur sukzessive fortgeführt, so erhält man schließlich eine Bandapproximation $\tilde{T}^B_{0,ns}$ an die im allgemeinen vollbesetzte Nichtstandardform $\tilde{T}_{0,ns}$ von T_0. Ist der Aufwand zum Berechnen der interpolierenden Einträge proportional zur Anzahl der Freiheitsgrade auf der betreffenden Skala, so ist die Konstruktion von $\tilde{T}^B_{0,ns}$ ebenfalls mit dem Aufwand $\mathcal{O}(wN_0)$ durchführbar.

  Bedenkt man, daß das soeben dargestellte Verfahren aufgrund des mangelnden Abfallverhaltens der Operatoreinträge bezüglich der nodalen Basis mit einer eigentlich unzulässigen Approximation $T^B_0 = T^{2w}_0$ an T_0 startet, so fragt man sich zu Recht, wie damit eine sinnvolle Approximation $\tilde{T}^B_{0,ns}$ an $\tilde{T}_{0,ns}$ möglich ist. Der Grund hierfür ist die Banderweiterung der resultierenden Grobgitteroperatoren $T^B_k = T^w_k \longrightarrow T^{2w}_k$ für $k = 1, ..., lt$ mittels besagter Interpolationsformel. Durch das Erweitern von $T^B_1 = T^w_1$ zu T^{2w}_1 wird implizit eine Interpolation von T_0 auf den Bändern $[-2w, -w]$ und $[w, 2w]$ berechnet und in den Transformationsprozeß miteinbezogen. Durch das Erweitern von $T^B_2 = T^w_2$ zu T^{2w}_2 geschieht selbiges hinsichtlich der Bänder $[-4w, -2w]$ und $[2w, 4w]$ von T_0, so daß durch das rekursive Fortsetzen der Prozedur schließlich die (vollbesetzte) Steifigkeitsmatrix T_0 über eine Interpolation in die algebraische Transformation miteinfließt.

Der in [48, 49] vorgestellte direkte Löser beruht auf einer komprimierten Darstellung der Operatoren über ihre Nichtstandardform und den Multiskalen-Vorwärts- und Rückwärtssubstitutionen mittels der Algorithmen 1a und 1b, wobei im Verlauf einer rekursiven Faktorisierung Bandapproximationen der entstehenden Schur-Komplement Operatoren verwendet werden. Eine solche Bandapproximation wird durch die nachfolgende Betrachtung — wiederum für den Ganzraumfall — motiviert.

Wir definieren ausgehend von dem in Satz 11 gezeigten Abfallverhalten der Matrixeinträge der Nichtstandardform von Integraloperatoren die folgende Klasse bi-unendlicher Matrizen

mit polynomialem Abfall:

$$\mathcal{M}_\alpha = \{(m_{i,j})_{i,j\in\mathbb{Z}} : \exists C = C(m) > 0 \quad |m_{i,j}| \leq C(1+|i-j|)^{-\alpha-1}\}.$$

Wir führen zudem noch analog die Klassen

$$\mathcal{M}_\alpha^k = \{(m_{i,j})_{i,j\in\mathbb{Z}} : \exists C = C(m,k) > 0 \quad |m_{i,j}| \leq C(1+|i-j|)^{-\alpha-1}\}$$

ein, die in einem späteren Beweis lediglich der besseren Unterscheidung verschiedener Skalen k dienen. Es gilt nun der grundlegende Satz:

Satz 12 (Eigenschaften bi-unendlicher Matrizen mit polynomialem Abfall)

i) Die Klasse $\mathcal{M}_\alpha^k$ ist abgeschlossen unter Addition und Multiplikation.

ii) Ist $M \in \mathcal{M}_\alpha^k$ invertierbar bezüglich des Folgenraums l_2, so gilt $M^{-1} \in \mathcal{M}_\alpha^k$.

Beweis:
Siehe [31, 114].

□

Mit Hilfe des Konzepts sogenannter H-Matrizen, die in dem Sinne schwach besetzt sind, daß nur wenige Daten zu ihrer Darstellung gebraucht werden, wird neuerdings auch in [66, 67, 70] der Aufbau einer Matrix-Klasse mit ähnlichen Eigenschaften wie den in Satz 12 beschriebenen vorangetrieben. Die Klasse der H-Matrizen ist darauf ausgerichtet, daß die Matrix-Vektor-Multiplikation in (fast) linearer Komplexität möglich ist sowie auch Summen und Produkte von H-Matrizen durch H-Matrizen in (fast) linearer Komplexität hinreichend genau approximiert werden können. Letzteres gilt ebenfalls für die Inverse einer H-Matrix, so daß die direkte Berechnung einer H-Matrix-Approximation eines Schur-Komplement Operators innerhalb dieser Klasse effizient durchgeführt werden kann. Die resultierende Matrizen-Arithmetik dient ebenfalls der Lösung von Integralgleichungen.

Zusammen mit Satz 11 ergibt sich aus Satz 12 eine Aussage über das Abfallverhalten der A-, B- und C-Block Einträge der im Verlauf einer Multiskalen-Faktorisierung von T_0 auftretenden Schur-Komplement Operatoren.

Satz 13 (Erhaltung des Abfallverhaltens für die Schur-Komplement Operatoren über endlich viele Skalen)
Es sei T ein Integraloperator der Form (4.18) auf $\mathcal{L}_2(\mathbb{R})$ und die zugrundeliegende Wavelet-Basis erfülle für diesen die Voraussetzungen von Satz 11. Es seien S_{k-1} der Schur-Komplement Operator bezüglich einer Skala $k-1 > 0$, der im Verlauf einer exakten Multiskalen-Faktorisierung ausgehend vom Feingitteroperator T_0 entsteht, und A_k^S, B_k^S sowie C_k^S die zugehörigen Blöcke nach einer entsprechenden Zweiskalen-Transformation. Dann gilt:

i) Es läßt sich S_{k-1} darstellen als Summe

$$S_{k-1} = T_{k-1} - F_{k-1}, \tag{4.26}$$

wobei T_{k-1} der Grobgitteroperator zur Skala $k-1$ ist und F_{k-1} zur Klasse $\mathcal{M}_m^{k-1}$ gehört.

ii) Es liegen A_k^S, B_k^S und C_k^S in der Klasse $\mathcal{M}_m^k$, das heißt die Einträge $a_{i,j}^{k,S}$, $b_{i,j}^{k,S}$ und $c_{i,j}^{k,S}$ dieser bi-unendlichen Matrizen erfüllen die Abschätzung

$$|a_{i,j}^{k,S}| + |b_{i,j}^{k,S}| + |c_{i,j}^{k,S}| \leq C(1+|i-j|)^{-m-1}. \tag{4.27}$$

Die Konstante $C = C(k, m)$ hängt ebenfalls von der jeweiligen Kernfunktion $K(x, y)$ ab.

Beweis:

Der Beweis erfolgt durch Induktion. Wir starten mit der Projektion T_0 von T auf die feinste Skala 0 und den Blöcken A_1, B_1, C_1 und T_1, die im Verlauf einer einfachen Zerlegung entstehen. Gemäß Satz 11 erfüllen die Einträge der Blöcke A_1, B_1 und C_1 die diesbezügliche Abschätzung (4.22), liegen also in $\mathcal{M}_m^1$. Ist A_1 invertierbar, so gilt nach Satz 12 das gleiche Abfallverhalten auch für die Einträge des Produkts $F_1 := C_1 A_1^{-1} B_1 \in \mathcal{M}_m^1$. Der Schur-Komplement Operator

$$S_1 = T_1 - F_1 = T_1 - C_1 A_1^{-1} B_1$$

hat damit die geforderte Darstellung (4.26) und zerfällt im Verlauf der nächsten einfachen Zerlegung in die Blöcke A_2^S, B_2^S, C_2^S und T_2^S. Diese lassen sich auch als die Differenzen der entsprechenden Blöcke von T_1 und F_1 schreiben. Die Einträge der Blöcke von T_1 haben wegen Satz 11 das richtige Abfallverhalten. Die Einträge der Blöcke von F_1 besitzen ebenfalls das richtige Abfallverhalten, da sie durch eine einfache Zerlegung von $F_1 \in \mathcal{M}_m^1$ entstehen. Damit haben die Einträge der Differenz-Blöcke das geforderte Abfallverhalten, liegen also in der Klasse $\mathcal{M}_m^2$, und die Induktionsverankerung ist gezeigt.
Es sei nun $S_{k-2} = T_{k-2} - F_{k-2}$ mit $F_{k-2} \in \mathcal{M}_m^{k-2}$, und für die Einträge der Blöcke A_{k-1}^S, B_{k-1}^S und C_{k-1}^S gelte die entsprechende Abschätzung (4.27). Ist A_{k-1}^S invertierbar, so schreibt sich der Schur-Komplement Operator zur Skala $k-1$ ausgehend von der Zerlegung von S_{k-2} als

$$\begin{aligned} S_{k-1} &= T_{k-1}^S - C_{k-1}^S (A_{k-1}^S)^{-1} B_{k-1}^S \\ &= T_{k-1} - [T_{k-1}^{F_{k-2}} + C_{k-1}^S (A_{k-1}^S)^{-1} B_{k-1}^S] \\ &=: T_{k-1} - F_{k-1}, \end{aligned}$$

wobei wir mit $T_{k-1}^{F_{k-2}} := (P_{k-1}^{k-2})^t F_{k-2} P_{k-1}^{k-2}$ die Projektion von F_{k-2} auf die nächstgröbere Skala $k-1$ bezeichnen. Es gilt $F_{k-1} \in \mathcal{M}_m^{k-1}$ aufgrund der Algebra-Eigenschaften der Klasse $\mathcal{M}_m^{k-1}$. Genauso wie die Blöcke A_{k-1}^S, B_{k-1}^S und C_{k-1}^S liegt nämlich auch $T_{k-1}^{F_{k-2}}$ in der Klasse $\mathcal{M}_m^{k-1}$, da er durch einfache Zerlegung von $F_{k-2} \in \mathcal{M}_m^{k-2}$ entsteht. Hier verwenden wir die Induktionsannahme. Damit hat S_{k-1} die geforderte Gestalt (4.26) und zerfällt im Verlauf der nächsten einfachen Zerlegung in die Blöcke A_k^S, B_k^S, C_k^S und T_k^S, die sich wieder als Differenzen der jeweiligen Blöcke von T_{k-1} und F_{k-1} darstellen lassen. Die Einträge der Blöcke von T_{k-1} haben wegen Satz 11 das richtige Abfallverhalten. Die Einträge der Blöcke von F_{k-1} besitzen ebenfalls das richtige Abfallverhalten, da sie durch eine einfache Zerlegung von $F_{k-1} \in \mathcal{M}_m^{k-1}$ entstehen. Damit haben die Einträge der Differenz-Blöcke ebenfalls das geforderte Abfallverhalten und der Induktionsschritt ist gezeigt.

□

Wir diskutieren nun den Algorithmus, der zu dem in [48, 49] vorgestellten direkten Löser führt. Er kombiniert die oben beschriebene algebraische Transformation auf Nichtstandardform mit gewöhnlichen LU-Faktorisierungen und liefert eine obere und untere Nichtstandardform $\tilde{U}_{0,ns}^B$ und $\tilde{L}_{0,ns}^B$, die zusammen mit unseren Multiskalen-Vorwärts- und -Rückwärtssubstitutionen ein direktes Lösungsverfahren zur Lösung bis auf eine vorgegebene Genauigkeit ε ergeben. Obwohl das Verfahren ursprünglich zur effizienten Multiskalen-Faktorisierung von Integraloperatoren gedacht war, läßt es sich gleichermaßen auch für eine Multiskalen-Faktorisierung von Differentialoperatoren einsetzen.

Es sei ε die beabsichtigte Lösungsgenauigkeit für das Verfahren. Wir wählen dann als Bandweite $w \geq (C/\varepsilon)^{\frac{1}{m}}$. Da die Konstante C, die unter anderem die Abhängigkeit von der Kernfunktion ausdrückt, nicht bekannt ist, nehmen wir an, daß w hinreichend groß ist. Wie im Fall der algebraischen Transformation auf Nichtstandardform wird zuerst wieder eine Approximation $T_0^B = T_0^{2w}$ des Operators T_0 bezüglich der feinsten Skala aufgestellt, indem man nur Einträge von T_0 berücksichtigt, die innerhalb eines Bandes $[-w, w]$ der Weite $2w$ um die Hauptdiagonale liegen. Die Zweiskalen-Transformation von T_0^B liefert die Blöcke A_1^B, B_1^B, C_1^B sowie T_1^B, die ebenfalls Bandgestalt haben. Mittels einer gewöhnlichen LU-Faktorisierung erhalten wir obere und untere Dreiecksmatrizen $\hat{A}_1^B$ und $\check{A}_1^B$, die Bandgestalt haben und so daß $A_1^B = \hat{A}_1^B \check{A}_1^B$ gilt. In [48, 49] werden „dünne“ Vorwärts- und Rückwärtssubstitutionen zur Berechnung von Bandapproximationen $\hat{C}_1^B$ und $\check{B}_1^B$ der am Aufbau der unteren und oberen Nichtstandardform beteiligten Nebendiagonalblöcke $\hat{C}_1 := C_1^B(\check{A}_1^B)^{-1}$ und $\check{B}_1 := (\hat{A}_1^B)^{-1}B_1^B$ vorgeschlagen. Da im Verlauf der späteren Multiskalen-Vorwärts- und Rückwärtssubstitutionen jedoch nur die Anwendung von $\hat{C}_1$ und $\check{B}_1$ auf Vektoren benötigt wird, genügt es, zunächst nur C_1^B und $\check{A}_1^B$ sowie B_1^B und $\hat{A}_1^B$ abzuspeichern. In Abhängigkeit der Länge der verwendeten Skalierungsfilter verkleinert sich, wie bereits erwähnt wurde, die Ausgangsbandweite $2w$ für T_1^B auf die halbe Weite w. Um rekursiv bezüglich der gröberen Skalen mit der Zerlegung und gleichzeitigen Faktorisierung fortzufahren, berechnet man wie folgt eine Bandapproximation S_1^B an den exakten Schur-Komplement Operator $S_1 = T_1 - C_1 A_1^{-1} B_1$. Es wird $T_1^B = T_1^w$ mit Hilfe einer Interpolationsformel symmetrisch zu einer Bandmatrix T_1^{2w} der ursprünglichen Weite $2w$ erweitert und hiervon das nur bezüglich dieses Bandes berechnete Produkt $\hat{C}_1^B \check{B}_1^B \approx C_1^B (A_1^B)^{-1} B_1^B$ subtrahiert. Wir nehmen an, daß der Aufwand zum Berechnen der interpolierenden Einträge proportional zur Anzahl der Freiheitsgrade auf der betreffenden Skala ist. Damit beträgt der Aufwand zur Berechnung von S_1^B gerade $\mathcal{O}(w^2 N_1)$. (Ist T_0 die Diskretisierung eines Differentialoperators, so entfällt die Interpolation.) Ausgehend von S_1^B wird die soeben beschriebene Prozedur bis zum Erreichen der gröbsten Skala lt rekursiv fortgesetzt. Wir erhalten eine untere und obere Nichtstandardform $\tilde{L}_{0,ns}^B$ und $\tilde{U}_{0,ns}^B$ mit dem Gesamtaufwand $\mathcal{O}(w^2 N_0)$. Die dabei gemachten Bandapproximationen sind durch Satz 13 gerechtfertigt.

Abbildung 4.2 zeigt eine untere und obere Nichtstandardform, die bei einer Multiskalen-Faktorisierung des eindimensionalen diskreten Laplace-Operators auf $\Omega =]0, 1[$ mittels linearer Prewavelets entstehen. Die Gitterweite beträgt hierbei $h_0 = \frac{1}{1024}$, die Bandweite $w = 12$.

Für höherdimensionale Probleme sind Bandapproximationen nur recht kompliziert und aufwendig zu berechnen [48, 107]. Dies betrifft zum einen die Besetzungsstruktur der zu berechnenden Approximationen. Die Berücksichtigung von Kopplungen weiter entfernter Nachbarn der Grobgitterpunkte führt bei Tensorprodukt-Ansätzen zu *Multibänder* (siehe die rechte Seite von Abbildung 3.5), die sich vereinfacht gesprochen aus Produkten eindimensionaler Bänder bezüglich der jeweiligen Richtungen ergeben. Betrachtet man beispielsweise ein zweidimensionales Problem auf einem uniformen kartesischen Gitter und beachtet bei der Formulierung der reduzierten Grobgittergleichungen zusätzlich noch die jeweils übernächsten Nachbarn, so werden dadurch 25-Punkte Sterne impliziert, die in den Grobgitterfreiheitsgraden aufsitzen. Bei lexikographischer Anordnung führen diese zu einem fünfbandigen Besetzungsmuster für die Schur-Komplement Approximation, wobei jedes der fünf Bänder selbst die Bandweite 5 besitzt. Die Maschenweite muß hierbei wenigstens $h = \frac{1}{16}$ betragen, damit die zugehörigen Multibänder sich nicht berühren oder überlappen. Berühren sich Multibänder, so verlieren Datenstrukturen zur Speicherung schwachbesetzter Matrizen ihre Effizenz. Zum anderen gestaltet sich in höheren Dimension auch die Berechnung effizienter Schur-Komplement

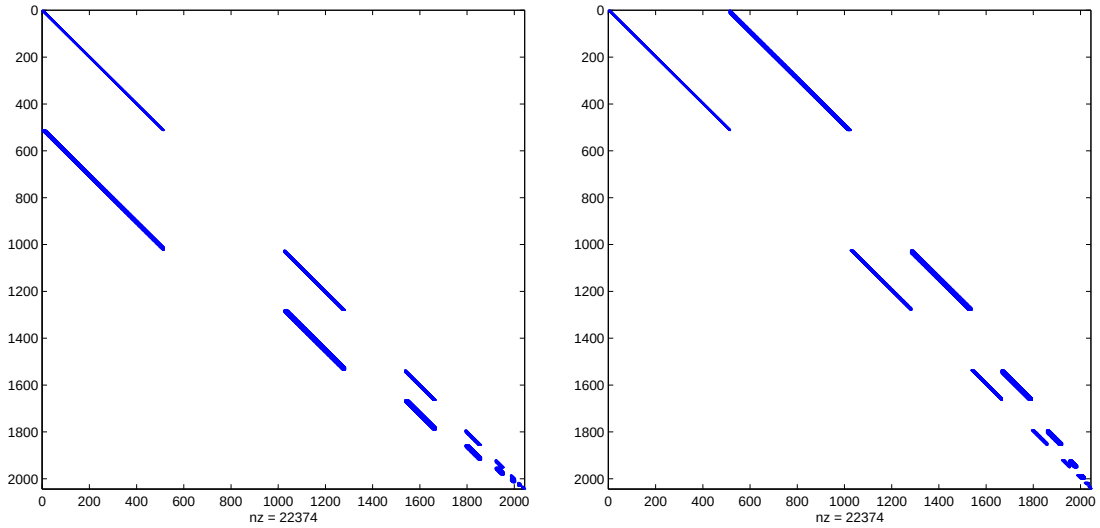


ABBILDUNG 4.2: *Untere (links) und obere (rechts) Nichtstandardformen* $\tilde{L}^B_{0,ns}$ *und* $\tilde{U}^B_{0,ns}$*, die bei einer Multiskalen-Faktorisierung des diskreten* 1D *Laplace-Operators auf* $\Omega =]0,1[$ *mittels linearer Prewavelets entstehen,* $h_0 = \frac{1}{1024}$*, Bandweite* $w = 12$*.*

Approximationen innerhalb der Multibänder wesentlich komplizierter als in einer Raumdimension. Es müssen Approximationen an die Produkte $C_k A_k^{-1} B_k$ berechnet werden, die auf subtile nichtlineare Weise den Einfluß der Detail-Anteile auf das grobskalige Verhalten der Lösung des betrachteten Problems modellieren. In zwei Dimensionen bestehen bei einem Tensorprodukt-Ansatz die Blöcke C_k und B_k aus jeweils drei Unterblöcken, die unter dem Einfluß der Inversen von A_k miteinander wechselwirken. Das Produkt $C_k A_k^{-1} B_k$ setzt sich somit aus insgesamt neun Summanden zusammen. Schließlich müssen die Schur-Komplement Approximationen so geartet sein, daß ein rekursives Fortsetzen der verfolgten Multiband-Strategie auf gröberen Skalen weiterhin möglich ist, ohne die Stabilität des resultierenden grobskaligen Problems zu verlieren oder die Komplexitätsgrenzen zu sprengen.

Für Konvektions-Diffusions Gleichungen erweist sich selbst für eindimensionale Fälle gerade die Frage nach dem Gelingen der approximativen Faktorisierungen der entsprechenden Blöcke A_k im Verlauf der weiteren Rekursion als problematisch, wie die numerischen Experimente in [107] zeigen. Ein Grund für die dort aufgezeigten Schwierigkeiten ist die mangelnde numerische Stabilität der mittels starrer Zweiskalen-Transformationen gebildeten Grobgitterprobleme. Weiterhin ist im Konvektions-Diffusions Fall für zunehmende Bandweiten und zunehmende Konvektionsstärken die Tendenz hin zu einem direkten Löser nicht mehr klar zu erkennen. Dies hängt auch mit einer Verschlechterung der Konstanten in den Abfallabschätzungen (4.22) und (4.27) zusammen, die von der zum Differentialoperator gehörenden Kernfunktion $K(x,y)$ abhängen.
Für das Lösen zweidimensionaler diskreter Konvektions-Diffusions Probleme erscheinen daher unsere auf approximativen Multiskalen-Faktorisierungen beruhenden iterativen Löser geeigneter. Unsere numerischen Beispiele in Kapitel 6 zeigen, daß man damit insbesondere durch den Einsatz von problemabhängigen Vergröberungstechniken, die in der Mehrgitter-Welt seit langem bekannt sind, und geschickten damit verbundenen Petrov–Galerkin-artigen Multiskalen-Zerlegungen zu robusten Verfahren gelangt.

AMLI-Vorkonditionierer

In [5, 6] wird mit den sogenannten AMLI-Verfahren (AMLI = Algebraic MultiLevel Iteration) eine Klasse optimaler Multilevel-Vorkonditionierer zur Lösung großer schwachbesetzter linearer Gleichungssysteme vorgestellt, die durch eine Finite-Elemente Diskretisierung elliptischer Randwertaufgaben mit symmetrisch positiv definitem Operator T und einer zugehörigen Bilinearform $a(\cdot\,,\cdot)$ entstehen. AMLI-Verfahren basieren auf der hierarchischen Basis Zerlegung [129] und können als eine algebraische Stabilisierungstechnik für die klassische HBMG-Methode [12, 16] interpretiert werden, die rekursiv zusätzliche innere Iterationen mittels geeigneter Matrixpolynome durchführt. Eine ähnliche Methode für Finite-Differenzen Diskretisierungen wird in [78] vorgeschlagen. Der Einsatz geeigneter AMLI-Vorkonditionierer in Bezug auf Krylovraum-Methoden wie etwa das CG-Verfahren führt zu Gitterweiten-unabhängigen Konvergenzraten der so vorkonditionierten Verfahren. Dies läßt sich ohne besondere Regularitätsannahmen im Rahmen einer algebraisch orientierten Konvergenztheorie zeigen, die als wesentliches mathematisches Hilfsmittel eine *verschärfte Cauchy–Schwarzsche Ungleichung* verwendet [12, 42]. Die Konstruktion der AMLI-Vorkonditionierer gründet auf dem folgenden natürlichen Aufbau der hierarchischen Basis „von grob nach fein".

Eine Funktion u werde zunächst auf einem relativ groben Gitter Ω_{lt} approximiert, indem ihre Werte in den Gitterpunkten als Koeffizienten der Entwicklung einer Approximation u_{lt} in einem grobskaligen Raum $\mathcal{V}_{lt}$, beispielsweise aufgespannt von Hutfunktionen, die in den entsprechenden Gitterpunkten verankert sind, interpretiert werden. Man gelangt so zu einem Vektor s^u_{lt}, der die nodale Darstellung der Approximation $u_{lt} \in \mathcal{V}_{lt}$ beschreibt. Will man die gegebene Funktion nun genauer bezüglich eines feineren Raumes $\mathcal{V}_{lt-1}$ mit $\mathcal{V}_{lt} \subset \mathcal{V}_{lt-1}$ approximieren, so fügt man durch (adaptive) Gitterverfeinerung weitere Diskretisierungspunkte zu den vorhandenen Punkten hinzu, in denen man sich die entsprechenden Hutfunktionen der nächstfeineren Skala verheftet denkt. Sie spannen einen Raum $\mathcal{W}_{lt}$ auf. Anstatt nun die im allgemeinen großen Werte der zu approximierenden Funktion in den neu hinzugekommenen Punkten zu speichern, kann man sie zunächst anhand von s^u_{lt} darin interpolieren und lediglich die Differenzen zwischen den interpolierten und ihren tatsächlichen Werten, sozusagen *hierarchische Überschüsse*, als $\mathcal{W}_{lt}$-Anteil d^u_{lt} abspeichern. Aus den Koeffizientenvektoren bezüglich $\mathcal{V}_{lt}$ und $\mathcal{W}_{lt}$ sowie der Kenntnis der Interpolationsvorschrift läßt sich sofort der Koeffizientenvektor der nodalen Darstellung bezüglich $\mathcal{V}_{lt-1}$, aufgespannt von den feinskaligen Hutfunktionen in sämtlichen Diskretisierungspunkten, berechnen. Dieser Basiswechsel hat bei einer entsprechenden Anordnung der Freiheitsgrade die Gestalt

$$W^{lt-1}_{lt,\mathcal{V}} = \begin{pmatrix} \mathbf{1} & J_{lt} \\ \mathbf{0} & \mathbf{1} \end{pmatrix},$$

wobei der Block J_{lt} die Interpolationsvorschrift beinhaltet. Aus einer so konstruierten hierarchischen Darstellung $\tilde{u}_{lt-1} = [d^u_{lt}; s^u_{lt}]$ einer feineren Approximation erhält man ihre nodale Darstellung bezüglich $\mathcal{V}_{lt-1}$ als

$$u_{lt-1} = W^{lt-1}_{lt,\mathcal{V}} \begin{pmatrix} d^u_{lt} \\ s^u_{lt} \end{pmatrix} = \begin{pmatrix} d^u_{lt} + J_{lt} s^u_{lt} \\ s^u_{lt} \end{pmatrix}.$$

Man erkennt hieran deutlich, daß der Anteil d^u_{lt} lediglich eine Korrektur der aus dem Anteil s^u_{lt} berechneten Interpolation $J_{lt} s^u_{lt}$ darstellt. Das rekursive Fortsetzen dieser Prozedur bis zum Erreichen einer feinsten Skala 0 führt zu einer Folge von Zweiskalen-Transformationen $\{W^k_{k+1}\}_{k=0,\ldots,lt-1}$, die eine hierarchische Zerlegung des feinsten Raums $\mathcal{V}_0$ induzieren.

Über diesen Zugang zur hierarchischen Basis definiert man rekursiv „von grob nach fein" die nachfolgende approximative Faktorisierung der zugehörigen Nichtstandardform $\tilde{T}_{0,ns}$ eines Feingitteroperators T_0. Es sei $M_{lt} := T_{lt}$. Wir definieren rekursiv für $k = lt-1, lt-2, ..., 0$

$$\tilde{M}_k := \begin{pmatrix} A_{k+1} & \mathbf{0} \\ C_{k+1} & \breve{T}_{k+1} \end{pmatrix} \begin{pmatrix} \mathbf{1} & A_{k+1}^{-1}B_{k+1} \\ \mathbf{0} & \mathbf{1} \end{pmatrix}, \quad M_k^{-1} := W_{k+1}^k \tilde{M}_k^{-1} (W_{k+1}^k)^t,$$

wobei die Matrizen $\breve{T}_{k+1}$ über ihre Inversen wie folgt erklärt sind:

$$\breve{T}_{k+1}^{-1} \quad := \quad [\mathbf{1} - p_\mu(M_{k+1}^{-1}T_{k+1})]T_{k+1}^{-1}.$$

Der Ausdruck p_μ bezeichnet hierbei ein Polynom vom Grad $\mu \geq 1$, so daß $0 \leq p_\mu(t) < 1$ für $0 < t \leq 1$ und $p_\mu(0) = 1$ ist. Ist $p_\mu \approx 0$ auf dem Spektrum von $M_{k+1}^{-1}T_{k+1}$, so gilt approximativ $\breve{T}_{k+1}^{-1} \approx T_{k+1}^{-1}$, und $\tilde{M}_k$ stellt eine approximative Zweiskalen-Faktorisierung der einfach transformierten Matrizen $\tilde{T}_k$ dar. Dies erklärt anschaulich den Stabilisierungseffekt. Man definiert schließlich als AMLI-Vorkonditionierer für T_0

$$M_{0,AMLI}^{-1} := M_0^{-1}.$$

Die Anwendung von $\tilde{M}_k^{-1}$ geschieht mittels einer Vorwärts- und Rückwärtssubstitution. Die im Verlauf einer Vorwärtssubstitution auftretenden Anwendungen von $\breve{T}_{k+1}^{-1}$ können ohne Invertierung von T_{k+1} implementiert werden. Aufgrund der Normierung $p_\mu(0) = 1$, so daß $p_\mu(t) = 1 - \eta_0 t - ... - \eta_{\mu-1}t^\mu$ mit geeigneten Koeffizienten $\eta_0, ..., \eta_{\mu-1}$ gilt, ist $q_{\mu-1}(t) := \frac{1-p_\mu(t)}{t}$ ebenfalls ein Polynom mit der Gestalt

$$q_{\mu-1}(t) = \frac{1-p_\mu(t)}{t} = \eta_0 + \eta_1 t + ... + \eta_{\mu-1}t^{\mu-1}.$$

Damit erhalten wir für das Ergebnis w_{k+1} der Anwendung von $\breve{T}_{k+1}^{-1}$ auf einen Vektor u_{k+1}

$$w_{k+1} = \breve{T}_{k+1}^{-1}u_{k+1} = q_{\mu-1}(M_{k+1}^{-1}T_{k+1})M_{k+1}^{-1}u_{k+1},$$

was mit Hilfe des Algorithmus

setze $w_{k+1}^0 := \mathbf{0}$;
`for` $z = 1 : \mu$
    löse $M_{k+1}w_{k+1}^z = \eta_{\mu-z}u_{k+1} + T_{k+1}w_{k+1}^{z-1}$;
`end`
setze $w_{k+1} := w_{k+1}^\mu$;

berechnet werden kann. Der durch den Polynomgrad bestimmte μ-fache Aufruf eines Grobgitterproblems in der obigen for-Schleife entspricht der Zyklus-Struktur, wie man sie von gewöhnlichen Mehrgitter-Verfahren her kennt oder wie sie auch in Algorithmus 2 gegeben ist. Im Fall $\mu = 1$ erhält man ein V-Zyklus-artiges Verfahren. Der Fall $\mu = 2$ führt auf eine dem W-Zyklus verwandte Struktur. Für $p_\mu(t) = 1 - t$ erhalten wir mit $M_{0,AMLI}^{-1}$ genau die gleiche approximative Multiskalen-Faktorisierung von T_0 wie diejenige, die durch die Algorithmen 1a und 1b gegeben ist.

Die Schur-Komplement Operatoren $S_{k+1} = T_{k+1} - C_{k+1}A_{k+1}^{-1}B_{k+1}$ zu den einfach hierarchisch transformierten Operatoren $\tilde{T}_k$ stimmen für $k = 0, ..., lt-1$ mit denjenigen überein,

die man mittels einer 2×2 Block-Partitionierung von T_k gemäß einer Zwei-Level Anordnung der Freiheitsgrade von Ω_k erhält [118]. Bei einer solchen Zwei-Level Anordnung werden zunächst die Unbekannten bezüglich der Fein-ohne-Grobgitterpunkte $\Omega_k \backslash \Omega_{k+1}$ und dann diejenigen bezüglich Ω_{k+1} aufgelistet. Dies erlaubt eine etwas effizientere Implementierung der AMLI-Vorkonditionierer anhand der Folge der 2×2 Block-partitionierten Operatoren T_k, ohne die Nichtstandardform $\tilde{T}_{0,ns}$ hinsichtlich der hierarchischen Zerlegung aufstellen zu müssen [3, 7]. Die Nebendiagonalblöcke B_k und C_k innerhalb von $\tilde{T}_{0,ns}$ sind nämlich etwas stärker besetzt als die ihnen entsprechenden Blöcke der Block-partitionierten Matrizen. Einen detaillierteren Vergleich unterschiedlicher Formulierungen der AMLI-Verfahren, auch im Zusammenhang mit der klassischen HBMG-Methode, findet man in [119, 120]. In [21] wird auf der Basis einer Multilevel-Anordnung der Freiheitsgrade eine verwandte Multilevel-ILU Vorkonditionierung vorgestellt, deren Multiskalen-Charakter zusammen mit einer levelabhängigen "Drop-tolerance"-Strategie ebenfalls auf der oben erwähnten Eigenschaft der rekursiv entstehenden Schur-Komplement Operatoren beruht. An dieser Stelle möchten wir auch noch auf den „Zoo" der von Saad und Zhang entwickelten Multilevel-ILU Varianten hinweisen [105, 106, 134, 136, 135].

Für ein Modellproblem mit einem symmetrisch positiv definiten elliptischen Operator zweiter Ordnung wird in [5] die Optimalität einer AMLI-Vorkonditionierung durch Nachweis der spektralen Äquivalenz

$$(M_{0,AMLI} u_0, u_0) \approx (T_0 u_0, u_0) \quad \forall u_0 \in \mathbb{R}^{N_0}$$

und Abschätzen des Aufwands für die Anwendung von $M_{0,AMLI}^{-1}$ gezeigt. Die in der obigen Normäquivalenz auftretenden Konstanten sind abhängig von der Art und dem Grad des verwendeten Polynoms p_μ sowie der Konstanten $\gamma := \sup_{k=1,\ldots,lt} \gamma_k$, die sich aus den *verschärften Cauchy–Schwarzschen Ungleichungen*

$$|a(d_k, s_k)| \leq \gamma_k [a(d_k, d_k)]^{\frac{1}{2}} [a(s_k, s_k)]^{\frac{1}{2}} \quad \forall d_k \in \mathcal{W}_k,\ s_k \in \mathcal{V}_k$$

ergibt. In [80] wird zum Beispiel gezeigt, daß im Fall der zum gewöhnlichen Laplace-Operator gehörenden Bilinearform $a(u, v) := \int_\Omega \nabla u \cdot \nabla v \, dx$ für beliebige Triangulierungen eines polygonal berandeten Gebiets $\Omega \subset \mathbb{R}^2$ und Diskretisierung mittels stückweise linearer Basisfunktionen die Abschätzung $\gamma^2 < \frac{3}{4}$ gilt. Der Rechenaufwand für die Anwendung von $M_{0,AMLI}^{-1}$ hängt maßgeblich ab vom Grad μ des Polynoms und der Art der Gitterverfeinerung. Für d-dimensionale Probleme (d=2, 3) und uniform verfeinerte Gitter erhält man durch Wahl spezieller Tschebyscheff-Polynome einen optimalen Vorkonditionierer, falls

$$(1 - \gamma^2)^{-\frac{1}{2}} < \mu < 2^d$$

gilt. Es genügt daher, $\mu = 2$ zu wählen, um zu einer optimalen AMLI-Vorkonditionierung des zweidimensionalen Laplace-Operators zu gelangen. Der Aufwand hierfür ist damit aber ebenso hoch wie der für einen hierarchischen W-Zyklus.

Im Verlauf der letzten Jahre sind verschiedene Varianten des hier vorgestellten Vorkonditionierers entwickelt worden, insbesondere um die Gesamtkosten zu verringern und auch um Robustheitsfragen etwa bei anisotropen Diffusions Problemen zu behandeln. In [119] wird die Verwendung levelabhängiger Polynomgrade μ_k untersucht. Dabei kann gezeigt werden, daß es für eine optimale Vorkonditionierung ausreicht, Polynome höheren Grades nur auf bestimmten Zwischenleveln einzusetzen. In [88] wird die AMLI-Methode für Konvektions-Diffusions

Probleme betrachtet, deren Konvektionsfelder $\vec{b}$ sich als Gradienten von Geschwindigkeitspotentialen ϕ darstellen lassen, das heißt $\vec{b} = \nabla\phi$. In solchen Fällen ist die Symmetrisierung des Problems mittels einer exponentiellen Transformation möglich und der für das symmetrisierte Gleichungssystem gewonnene AMLI-Vorkonditionierer kann innerhalb eines PCG-Verfahrens eingesetzt werden. Die zur symmetrisierten Gleichung gehörende Energienorm ist abhängig von der gewählten Transformation, in die ihrerseits ϕ und damit implizit auch $\vec{b}$ eingeht. Daher müssen Konvergenz- und Abbruchkriterien für das zugehörige PCG-Verfahren sehr behutsam und mit großer Vorsicht festgelegt werden (siehe auch unsere diesbezügliche Diskussion im Anhang sowie [59]). Ein direktes Anwenden der AMLI-Methode auf die nach Diskretisierung von Konvektions-Diffusions Gleichungen entstehenden unsymmetrischen Gleichungssysteme ist uns allerdings nicht bekannt. Einen Überblick der Entwicklungen der letzten Jahre findet man in [7, 88, 120].

Im Hinblick auf adaptive Verfahren erscheinen der Aufbau der hierarischen Zerlegung des Feingitterraums $\mathcal{V}_0$ „von grob nach fein“ und die damit verbundene Definition von $M^{-1}_{0,AMLI}$ recht natürlich. Aber für die erfolgreiche Behandlung von Problemen mit singulären Störungen wie etwa Anisotropien oder springenden Diffusionskoeffizienten muß man in der Regel jedoch problemangepaßte Multiskalen-Zerlegungen von $\mathcal{V}_0$ betrachten. Diese gewinnt man mittels problemabhängiger Interpolationen, die zum Beispiel durch das Lösen homogener lokaler Probleme auf gröberen Gittern und die darüber erhaltenen lokalen Greenschen Funktionen definiert werden können. Wir erklären das Vorgehen am Beispiel von Konvektions-Diffusions Problemen genauer im nachfolgenden Kapitel 5 und sehen dann, daß die so gewonnenen operatorabhängigen Interpolationen identisch mit speziellen matrixabhängigen Prolongationen sind. Sie werden mittels des diskreten Operators zum nächstfeineren Gitter berechnet. In der Praxis ist die Berechnung solcher matrixabhängiger Prolongationen sogar der einzig gangbare Weg, da die Integration der homogenen lokalen Grobgitterprobleme im allgemeinen nicht möglich ist. Aus diesem Blickwinkel heraus erscheint die von uns gewählte Definition der Multiskalen-Verfahren „von fein nach grob“ daher geeigneter.

Approximative Zyklische Reduktion

Die klassische Methode der Zyklischen Reduktion (ZR) [51] verwendet man beispielsweise zur direkten Lösung linearer Gleichungssysteme mit tridiagonalen Systemmatrizen T_0. Sie entstehen etwa bei der Diskretisierung eindimensionaler Probleme mittels 3-Punkte Sternen, wie wir im zweiten Kapitel gesehen haben. Wir nehmen an, daß sämtliche Diagonaleinträge der Matrix T_0 von Null verschieden sind. Ordnet man die Unbekannten gemäß einer (offensichtlichen) Rot-Schwarz-Färbung an [62], die einer Einteilung der Freiheitsgrade in Fein-ohne-Grobgitterpunkte und Grobgitterpunkte dient, dann läßt sich der Feingitteroperator nach einer entsprechenden Permutation P^0_1 exakt faktorisieren zu

$$\tilde{T}_0 := (P^0_1)^t T_0 P^0_1 = \begin{pmatrix} A_1 & B_1 \\ C_1 & T_1 \end{pmatrix} = \begin{pmatrix} A_1 & \mathbf{0} \\ C_1 & S_1 \end{pmatrix} \begin{pmatrix} \mathbf{1} & A_1^{-1}B_1 \\ \mathbf{0} & \mathbf{1} \end{pmatrix}. \tag{4.28}$$

Dabei ist $S_1 = \tilde{T}_0/T_1 = T_1 - C_1A_1^{-1}B_1$ der Schur-Komplement Operator. Die bekannte Blockgestalt von $\tilde{T}_0$ wird hier lediglich durch eine Permutation der Unbekannten entsprechend ihrer Färbung induziert und nicht durch eine zusätzliche Basistransformation W^0_1, die von einer Zweiskalen-Zerlegung des Raums der Feingitterfunktionen herrührt. Da T_0 tridiagonal ist, sind die nach Permutation entstehenden Blöcke A_1 und T_1 diagonal und S_1 hat eine tridiagonale Gestalt. Ist T_0 symmetrisch positiv definit oder allgemeiner eine M-Matrix,

so gilt dies ebenfalls für S_1 und sämtliche Diagonaleinträge von S_1 sind verschieden von Null [64]. Damit kann die soeben vorgestellte Prozedur rekursiv fortgesetzt werden, was zu einer sukzessiven Reduktion der Grobgitterunbekannten und schließlich zu einer exakten Faktorisierung des permutierten Feingitteroperators führt. Diese ist äquivalent zur klassischen Gauß-Elimination und man erhält mittels einer Vorwärts- und Rückwärtssubstitution ein direktes Lösungsverfahren.

In [96, 97] werden Verallgemeinerungen des soeben vorgestellten Verfahrens der Zyklischen Reduktion für den Fall nicht notwendig tridiagonaler schwachbesetzter Matrizen diskutiert. Da ihre Multilevel-artige Konstruktion auf Techniken algebraischer Mehrgitter-Verfahren basiert, sollen sie insbesondere der robusten Lösung linearer Gleichungssysteme dienen, die von singulär gestörten Problemen stammen. Voraussetzung ist, daß T_0 eine schwach diagonaldominante M-Matrix ist [65]. Ziel ist es, durch ein geeignetes rekursives Aufteilen der Freiheitsgrade in Fein-ohne-Grobgitter- und Grobgitterunbekannte sowie damit verbundene Schur-Komplement Approximationen eine annähernde Faktorisierung des Feingitteroperators zu berechnen. Die Aufteilung der Unbekannten (Rot-Schwarz-Färbung) geschieht dabei anhand eines algebraischen Kriteriums, das ähnlich wie bei algebraischen Mehrgitter-Verfahren [27, 52, 103, 112] nicht die Kenntnis einer zugrundeliegenden Triangulierung oder eines Gitters voraussetzt. Hierzu wird der Graph $G_{T_0}(V, E)$ der Matrix T_0 definiert, dessen Knotenmenge den Unbekannten und dessen Kantenmenge E den nicht verschwindenden Matrixeinträgen entspricht. Ignoriert man bestimmte „kleine“ Matrixeinträge, was mit Hilfe eines Parameters $0 \leq \beta < 1$ gesteuert wird, so gelangt man zum reduzierten Matrixgraphen $G_{T_0}(V, E_s)$. Ausgehend von den so erhaltenen „starken“ Kopplungen E_s wird nun anhand von $G_{T_0}(V, E_s)$ eine geeignete Rot-Schwarz-Färbung der Freiheitsgrade definiert. Die roten Knoten bilden dabei eine *maximal unabhängige Menge*, das heißt je zwei Knoten dieser Menge haben keine verbindende Kante im reduzierten Matrixgraphen und es existiert keine echte Obermenge mit dieser Eigenschaft. Dann ist der zugehörige Block A_1 des entsprechend permutierten Operators $\tilde{T}_0$ annähernd diagonal. Es gilt

$$||\mathbf{1} - A_1 \check{D}_1^{-1}||_\infty < 1, \tag{4.29}$$

wobei $\check{D}_1$ diejenige Diagonalmatrix bezeichnet, die man durch Weglassen aller Nebendiagonaleinträge von A_1 erhält. Zur Definition einer Approximation des Schur-Komplement Operators nutzt man nun aus, daß Rechtstransformationen der Form

$$\bar{T}_0 := \tilde{T}_0 \begin{pmatrix} \mathbf{1} & -G_1 \\ \mathbf{0} & \mathbf{1} \end{pmatrix} = \begin{pmatrix} A_1 & B_1 - A_1 G_1 \\ C_1 & T_1 - C_1 G_1 \end{pmatrix}$$

den Schur-Komplement Operator $S_1 = T_1 - C_1 A_1^{-1} B_1$ invariant lassen. Es ist dabei G_1 eine beliebige Matrix, die die gleiche Dimension wie B_1 besitzt. Es gilt nämlich

$$\bar{T}_0 / (T_1 - C_1 G_1) = T_1 - C_1 G_1 - C_1 A_1^{-1} (B_1 - A_1 G_1) = S_1.$$

Wir definieren nun ausgehend von

$$\check{T}_0^{(0)} := \begin{pmatrix} \check{A}_1^{(0)} & \check{B}_1^{(0)} \\ \check{C}_1^{(0)} & \check{T}_1^{(0)} \end{pmatrix} := \begin{pmatrix} A_1 & B_1 \\ C_1 & T_1 \end{pmatrix} =: \begin{pmatrix} \hat{A}_1^{(0)} & \hat{B}_1^{(0)} \\ \hat{C}_1^{(0)} & \hat{T}_1^{(0)} \end{pmatrix} =: \hat{T}_0^{(0)} \tag{4.30}$$

für $k \in \mathbb{N}$ induktiv die Block-Matrizen

$$\check{T}_0^{(k)} := \check{T}_0^{(k-1)} \begin{pmatrix} \mathbf{1} & -\check{D}_1^{-1}\check{B}_1^{(k-1)} \\ \mathbf{0} & \mathbf{1} \end{pmatrix}, \tag{4.31}$$

$$\hat{T}_0^{(k)} := \check{T}_0^{(k-1)} \begin{pmatrix} \mathbf{1} & -\hat{D}_1^{-1}\check{B}_1^{(k-1)} \\ \mathbf{0} & \mathbf{1} \end{pmatrix}, \tag{4.32}$$

wobei $\hat{D}_1 := diag(A_1 \cdot \mathbf{1})$ neben $\check{D}_1$ eine weitere Diagonalmatrix darstellt. Eine Motivation für diese hybride Konstruktion wird weiter unten gegeben. Sie läßt sich auch als Petrov–Galerkin Ansatz interpretieren. Wir erhalten für alle $k \geq 0$

$$\check{T}_0^{(k)}/\check{T}_1^{(k)} = S_1 = \hat{T}_0^{(k)}/\hat{T}_1^{(k)},$$

da die Block-Matrizen einzig durch Rechtstransformationen aus $\tilde{T}_0$ hervorgehen. Es gelten zudem die Formeln

$$\begin{aligned} \check{B}_1^{(k)} &= (\mathbf{1} - A_1\check{D}_1^{-1})^k B_1, \\ \hat{B}_1^{(k)} &= (\mathbf{1} - A_1\hat{D}_1^{-1})(\mathbf{1} - A_1\check{D}_1^{-1})^{(k-1)} B_1. \end{aligned}$$

Aufgrund von (4.29), also der Diagonaldominanz von A_1, kann man zeigen, daß die so darstellbaren Nebendiagonalblöcke $\check{B}_1^{(k)}$ und $\hat{B}_1^{(k)}$ für hinreichend großes k „klein“ sind und die Blöcke $\check{T}_1^{(k)}$ und $\hat{T}_1^{(k)}$ im starken Sinne gegen S_1 konvergieren. Es ist $\check{T}_1^{(k)}$ stets eine schwach diagonaldominante M-Matrix, und ist S_1 nichtsingulär, so trifft dies auch auf $\hat{T}_1^{(k)}$ zu. Daher ist es sinnvoll, $\check{T}_1^{(k)}$ und $\hat{T}_1^{(k)}$ als Approximationen für S_1 heranzuziehen. Um ein Auffüllen von $\check{T}_1^{(k)}$ und $\hat{T}_1^{(k)}$ zu vermeiden, sollte man ein relativ kleines k benutzen. In [96] wird vorgeschlagen, $k = 2$ und damit $\hat{T}_1^{(2)}$ als Approximationen für S_1 und als Ausgangspunkt für die weitere Reduktion zu verwenden. Ist $\check{A}_1$ eine Approximation an A_1, so wählt Reusken im Zweilevel-Fall als approximative Faktorisierung von $\tilde{T}_0$

$$\tilde{M}_{0,AZR} := \begin{pmatrix} \check{A}_1 & \mathbf{0} \\ C_1 & \hat{T}_1^{(2)} \end{pmatrix} \begin{pmatrix} \mathbf{1} & \check{A}_1^{-1}B_1 \\ \mathbf{0} & \mathbf{1} \end{pmatrix}. \tag{4.33}$$

Zwar erhält man durch $\check{T}_1^{(k)}$ ebenfalls eine stabile Approximation von S_1, doch ist die Konvergenz $\lim_{k\to\infty} \check{T}_1^{(k)} u_1 = S_1 u_1$ für „glatte“ Gitterfunktionen u_1 recht langsam [97]. Durch die Wahl $\hat{T}_1^{(2)}$ aufgrund der hybriden Konstruktion (4.30)-(4.32) kann dieser Defekt überwunden werden. Dies wird in [96] für einige Modellbeispiele mit periodischen Randbedingungen ((anisotrope) Diffusion, Konvektion-Diffusion) im Zweilevel-Fall quantitativ gezeigt. Obwohl $\hat{T}_1^{(2)}$ nach Konstruktion ein schwaches Besetzungsmuster aufweist, ist dieses aufgrund von “fill in” doch dichter als das von T_0. Um ein sukzessives Auffüllen der im Verlauf einer mehrstufigen Zerlegung auftretenden Matrizen zu vermeiden, wird darüberhinaus eine „Lumpingtechnik“ vorgeschlagen, die die betragsmäßig kleinsten Matrixeinträge der Schur-Komplement Approximation auf die Diagonale addiert und diese innerhalb der Matrix auf Null setzt. Die Eigenschaft von $\hat{T}_1^{(2)}$, eine schwach diagonaldominante M-Matrix zu sein, bleibt dabei erhalten, so daß eine Rekursion weiterhin möglich ist.

Wir wollen jetzt $\tilde{M}_{0,AZR}$ mit einer approximativen Faktorisierung $\tilde{M}_0$ vergleichen, die man mittels verallgemeinerter hierarchische Basis Transformationen erhält. Dazu bemerken wir,

daß die Schur-Komplement Approximation $\hat{T}_1^{(2)}$ auch über einen Petrov–Galerkin-artigen Vergröberungsprozeß mit matrixabhängigen Prolongationen und Restriktionen interpretiert werden kann. Wie man nach einfacher Rechnung sieht, gilt

$$\hat{T}_1^{(2)} = \begin{pmatrix} -C_1\hat{D}_1^{-1} & \mathbf{1} \end{pmatrix} \tilde{T}_0 \begin{pmatrix} -\check{D}_1^{-1}B_1 \\ \mathbf{1} \end{pmatrix}. \tag{4.34}$$

Man vergleiche dies auch mit den folgenden Darstellungen des exakten Schur-Komplement Operators, bei denen jeweils $*$ einen beliebigen Block passender Größe kennzeichnet:

$$S_1 = \begin{pmatrix} -C_1A_1^{-1} & \mathbf{1} \end{pmatrix} \tilde{T}_0 \begin{pmatrix} * \\ \mathbf{1} \end{pmatrix} = \begin{pmatrix} * & \mathbf{1} \end{pmatrix} \tilde{T}_0 \begin{pmatrix} -A_1^{-1}B_1 \\ \mathbf{1} \end{pmatrix}.$$

Den permutierten Feingitteroperator $\tilde{T}_0$ transformiert man dann ausgehend von (4.34) mittels einer zugehörigen verallgemeinerten hierarchischen Basis Transformation zu

$$\begin{pmatrix} \mathbf{1} & \mathbf{0} \\ -C_1\hat{D}_1^{-1} & \mathbf{1} \end{pmatrix} \begin{pmatrix} A_1 & B_1 \\ C_1 & T_1 \end{pmatrix} \begin{pmatrix} \mathbf{1} & -\check{D}_1^{-1}B_1 \\ \mathbf{0} & \mathbf{1} \end{pmatrix} = \begin{pmatrix} A_1 & [\mathbf{1} - A_1\check{D}_1^{-1}]B_1 \\ C_1[\mathbf{1} - \hat{D}_1^{-1}A_1] & \hat{T}_1^{(2)} \end{pmatrix}. \tag{4.35}$$

Unsere dafür im Zweilevel-Fall erklärte approximative Faktorisierung

$$\tilde{M}_0 := \begin{pmatrix} \check{A}_1 & 0 \\ C_1[\mathbf{1} - \hat{D}_1^{-1}A_1] & \hat{T}_1^{(2)} \end{pmatrix} \begin{pmatrix} \mathbf{1} & \check{A}_1^{-1}[\mathbf{1} - A_1\check{D}_1^{-1}]B_1 \\ \mathbf{0} & \mathbf{1} \end{pmatrix} \tag{4.36}$$

benutzt die leicht modifizierten Nebendiagonalblöcke $[\mathbf{1} - A_1\check{D}_1^{-1}]B_1$ und $C_1[\mathbf{1} - \hat{D}_1^{-1}A_1]$, nun allerdings zur Approximation der transformierten Steifigkeitsmatrix. Möglichkeiten $\tilde{M}_{0,AZR}$ mit Hilfe von leicht „verbogenen“ verallgemeinerten hierarchischen Basis Transformationen über unseren Ansatz zu generieren, werden in [94] diskutiert.

Reuskens Methode der Approximativen Zyklischen Reduktion definiert auf rein algebraischem Wege eine näherungsweise Inverse $M_{0,AZR}^{-1}$ von T_0. Sie läßt sich ebenfalls über eine verallgemeinerte hierarchische Basis Transformation interpretieren, wie wir gerade gesehen haben. Die Unterteilung der Freiheitsgrade in Fein-ohne-Grobgitter- und Grobgitterpunkte erfolgt dabei mit Hilfe AMG-basierter Prinzipien. Eine naheliegende Idee ist daher, die von uns in diesem Kapitel vorgestellten Multiskalen-Verfahren direkt über verallgemeinerte hierarchische Basen zu definieren, die mit Hilfe von algebraischen Mehrgitter-Verfahren konstruiert werden. Es werden dazu die rechteckigen Transformations-Matrizen in Formel (4.34) zur Berechnung der Schur-Komplement Approximationen ersetzt durch Prolongationen und Restriktionen, die mittels AMG bestimmt werden. Da die AMG-Konstruktionen im Fall des gewöhnlichen Laplace-Operators jedoch auf geometrische Vergröberungen und bilineare Interpolationen zurückführen, ist nicht zu erwarten, daß man mittels AMG-basierter hierarchischer Zerlegungen zumindest optimale Multiskalen-Löser erhält. Unter diesem Nachteil leiden ebenfalls die durch Approximative Zyklische Reduktion gewonnenen Vorkonditionierer, wie die numerischen Experimente in [98] zeigen und Reusken dort auch eingesteht. Wir zeigen in Kapitel 5.3, das sich mit algebraisch definierten Multiskalen-Zerlegungen befaßt, eine einfache Möglichkeit zur Stabilisierung AMG-basierter hierarchischer Basen. Mit ihrer Hilfe gelingt die Konstruktion sowohl Gitterweiten- als auch weitgehend störungsunabhängiger Multiskalen-Löser für schwierige Probleme, etwa wirbelbehaftete Konvektions-Diffusions Gleichungen, wie unsere numerischen Beispiele in Kapitel 6 zeigen.

4.3 Additive Multiskalen-Verfahren

In diesem Abschnitt definieren wir einen additiven Multiskalen-Vorkonditionierer für das Feingittersystem (3.5), der ebenfalls auf den Multiskalen-Zerlegungen (3.2) und (3.3) beruht. Im Unterschied zu multiplikativen Verfahren werden hierbei gleichzeitig und unabhängig voneinander Unterraumkorrekturen für alle an der Zerlegung beteiligten Skalen berechnet. Der so definierte Vorkonditionierer gehört damit zur Klasse der *parallelen Unterraumkorrekturmethoden* [127].

Äquivalente Implementierungen ergeben sich über die Multiskalen-Vorwärts- und Rückwärtssubstitutionen sowie den Multiskalen-V-Zyklus durch Weglassen der Modifikationen der grobskaligen Anteile (Schritt 2c in Algorithmus 1a bzw. Schritt 3 in Algorithmus 2) und der Nachkorrektur (Schritt 2a in Algorithmus 1b bzw. Schritt 5 in Algorithmus 2). Die Rückwärtssubstitutionsroutine sammelt dann nur noch die berechneten Korrekturen auf und transformiert sie geeignet zurück. Man erhält mit Hilfe der so modifizierten Algorithmen das Ergebnis f_0^P der Anwendung des additiven Multiskalen-Vorkonditionierers auf einen Vektor f_0 durch die Aufrufe

$$\begin{aligned} [d_1; ...; d_{lt}; s_{lt}] &= \text{ms\_vorsubst}(lt, \mathbf{0}, f_0), \\ f_0^P &= \text{ms\_rücksubst}(lt, [d_1; ...; d_{lt}; s_{lt}]) \end{aligned}$$

beziehungsweise

$$f_0^P = \text{ms\_v\_zyklus}(lt, 1, \mathbf{0}, f_0).$$

Eine effiziente Implementierung ist ebenfalls auf der Basis des Standardsystems

$$(W_{lt,\mathcal{S}}^0)^t T_0 W_{lt,\mathcal{V}}^0 \tilde{u}_{0,s} = (W_{lt,\mathcal{S}}^0)^t f_0$$

möglich. Es ergibt sich hieraus ein additiv Multiskalen-linksvorkonditioniertes System durch Links-Multiplikation mit der Blockdiagonalmatrix

$$\tilde{M}_{0,s}^{-1} := blockdiag(A_1^{-1}, ..., A_{lt}^{-1}, T_{lt}^{-1}) \tag{4.37}$$

und anschließende Rücktransformation mittels $W_{lt,\mathcal{V}}^0$. Wir erhalten

$$W_{lt,\mathcal{V}}^0 \tilde{M}_{0,s}^{-1} (W_{lt,\mathcal{S}}^0)^t T_0 u_0 = W_{lt,\mathcal{V}}^0 \tilde{M}_{0,s}^{-1} (W_{lt,\mathcal{S}}^0)^t f_0.$$

Der additive Multiskalen-Vorkonditionierer schreibt sich daher als

$$M_{0,s}^{-1} = W_{lt,\mathcal{V}}^0 \tilde{M}_{0,s}^{-1} (W_{lt,\mathcal{S}}^0)^t, \tag{4.38}$$

und es gilt $f_0^P = M_{0,s}^{-1} f_0$. Zur Implementierung von $M_{0,s}^{-1}$ braucht insbesondere die Standardform nicht vollständig bekannt zu sein. Man benötigt lediglich deren Diagonalböcke zur Definition der Blockdiagonalmatrix $\tilde{M}_{0,s}^{-1}$ sowie die Transformationen $W_{lt,\mathcal{V}}^0$ und $(W_{lt,\mathcal{S}}^0)^t$, die als schnelle Wavelettransformationen effizient realisiert werden können. Mit Hilfe der Synthese-Operatoren $Q_{k,\mathcal{V}}^0$ und $P_{lt,\mathcal{V}}^0$ sowie der transponierten Synthese-Operatoren $(Q_{k,\mathcal{S}}^0)^t$ und $(P_{lt,\mathcal{S}}^0)^t$ für $k = 1, ..., lt$ (siehe deren Definitionen (3.17) und (3.18)) kann $M_{0,s}^{-1}$ auch als Summe

$$M_{0,s}^{-1} = \sum_{k=1}^{lt} Q_{k,\mathcal{V}}^0 A_k^{-1} (Q_{k,\mathcal{S}}^0)^t + P_{lt,\mathcal{V}}^0 T_{lt}^{-1} (P_{lt,\mathcal{S}}^0)^t \tag{4.39}$$

geschrieben werden. Hierdurch wird der additive Charakter des Vorkonditionierers ebenfalls deutlich. Die darin auftretenden Inversen werden selbstverständlich für praktische Rechnungen wie schon im multiplikativen Fall durch geeignete approximative Inversen ersetzt.

Kapitel 5

Multiskalen-Zerlegungen für Konvektions-Diffusions Probleme

Im letzten Kapitel haben wir ausgehend von den recht allgemeinen Petrov–Galerkin-artigen Multiskalen-Zerlegungen (3.2) und (3.3) der Ansatz- und Testseite sowohl multiplikative als auch additive Multiskalen-Verfahren entwickelt und mit bekannten Methoden verglichen. Es ist klar, daß die Güte und insbesondere die Robustheit der Verfahren wesentlich von den Transformationen $P^k_{k+1,\mathcal{V}/\mathcal{S}}$ und $Q^k_{k+1,\mathcal{V}/\mathcal{S}}$ $(k = 0, ..., lt-1)$ abhängt, die implizit die Zerlegungen (3.2) und (3.3) festgelegen. Sie bestimmen die Blöcke der Nichtstandardform $\tilde{T}_{0,ns}$ sowie die Projektionen der Residuenvektoren auf die Räume grobskaliger Anteile. Mit ihnen wird außerdem die Rücktransformation der berechneten Korrekturen durchgeführt. Um robuste Multiskalen-Verfahren zu erhalten, sollte daher die Wahl der Transformationen $P^k_{k+1,\mathcal{V}/\mathcal{S}}$ und $Q^k_{k+1,\mathcal{V}/\mathcal{S}}$ problemangepaßt erfolgen. Dies unterscheidet unseren Ansatz wesentlich von den Wavelet-basierten Methoden [29, 30, 49, 92, 93, 100].

In diesem Kapitel diskutieren wir zunächst die problemabhängige Wahl der Prolongationen $P^k_{k+1,\mathcal{V}/\mathcal{S}}$, die dazu dient, physikalisch sinnvolle Grobgitterprobleme im Verlauf der rekursiven Zerlegung aufzustellen. Problemabhängige Prolongationen $P^k_{k+1,\mathcal{V}/\mathcal{S}}$ können über eine durch den Operator induzierte Interpolation berechnet werden, die auf der Lösung lokaler homogener *kontinuierlicher* Probleme beruht und mittels operatorabhängiger Finite-Elemente Basen (L-Splines, siehe Kapitel 2) interpretiert werden kann. Diese lokalen Greenschen Funktionen erfüllen verallgemeinerte Skalierungsgleichungen, was den Zusammenhang mit einer *Multiskalen-Analyse* herstellt [81]. Mit Hilfe der über L-Splines berechneten Steifigkeitsmatrix kann dann eine äquivalente matrixabhängige Prolongation durch die Lösung entsprechender lokaler homogener *diskreter* Probleme bestimmt werden. Letzteres entspricht gerade einer lokalen Gauß-Elimination und ist für Probleme mit variablen Koeffizientenfunktionen der einzig gangbare Weg zur Bestimmung einer solchen problemabhängigen Prolongation, da die Lösung der lokalen homogenen *kontinuierlichen* Probleme im allgemeinen viel zu aufwendig ist. Nach der Vorstellung problemangepaßter Prolongationen $P^k_{k+1,\mathcal{V}/\mathcal{S}}$ diskutieren wir ihr Zusammenspiel mit einer möglichen hierarchischen und Wavelet-artigen Wahl der Basen der zugehörigen Komplementräume, die über die Transformationen $Q^k_{k+1,\mathcal{V}/\mathcal{S}}$ festgelegt wird.

Wir besprechen zuerst eindimensionale Probleme, für deren Lösung man selbstverständlich kein Multiskalen-Verfahren mit notwendigerweise problemangepaßten Komponenten bemühen muß (siehe etwa die Methode der Zyklischen Reduktion in Kapitel 4.2.3), und untersuchen

dann, inwieweit sich die eindimensionalen Techniken auf zweidimensionale Aufgaben im Rahmen eines klassischen Tensorprodukt-Ansatzes übertragen lassen.

Die bis dahin diskutierten Multiskalen-Zerlegungen werden anhand einer a priori festgelegten Gitterhierarchie konstruiert. Wir sprechen daher auch von *geometrischen Multiskalen-Zerlegungen*. Im letzten Unterkapitel diskutieren wir noch Möglichkeiten, problemangepaßte Multiskalen-Zerlegungen auf rein *algebraischem* Wege zu erklären. Wir greifen dazu auf Vergröberungsstrategien zurück, die von algebraischen Mehrgitter-Verfahren her bekannt sind [27, 52, 112]. Darüber definierte *algebraische Multiskalen-Zerlegungen* sind insbesondere nützlich zur Konstruktion robuster Verfahren für nichtseparable Konvektions-Diffusions Probleme, etwa für Aufgaben mit wirbelbehafteter Konvektion.

5.1 Probleme in 1D

5.1.1 Operatorabhängige Prolongationen

Zur Herleitung einer problemangepaßten Prolongation $P^k_{k+1,\mathcal{V}}$ für die Ansatzseite betrachten wir das eindimensionale homogene Konvektions-Diffusions Problem mit konstanter Konvektionsstärke $b \neq 0$ und inhomogenen Dirichletschen Randbedingungen

$$-u'' + bu' = 0 \quad \text{auf} \quad]x_i^{k+1}, x_{i+1}^{k+1}[, \quad u(x_i^{k+1}) = \alpha \quad \text{und} \quad u(x_{i+1}^{k+1}) = \beta. \tag{5.1}$$

Das Intervall $]x_i^{k+1}, x_{i+1}^{k+1}[\subset \Omega =]0,1[$ sei hierbei ein finites Element der Länge $h_{k+1} = 2h_k$ bezüglich eines uniformen Gitters $\Omega_{k+1} := \{x_i^{k+1} \in \Omega : x_i^{k+1} := ih_{k+1} \text{ für } 0 < i < 1/h_{k+1}\}$. Wir nehmen ferner an, daß das Gitter Ω_{k+1} eingebettet ist in eine geschachtelte Folge uniform verfeinerter Gitter $\Omega_{lt} \subset \cdots \subset \Omega_{k+1} \subset \Omega_k \subset \cdots \subset \Omega_0$. Die Lösung des eindimensionalen Randwertproblems (5.1) lautet

$$u(x) = c_1 + c_2 \exp(bx)$$

mit den durch die Randbedingungen (5.1) festgelegten Konstanten

$$c_1 = \frac{\alpha \exp(bx_{i+1}^{k+1}) - \beta \exp(bx_i^{k+1})}{\exp(bx_{i+1}^{k+1}) - \exp(bx_i^{k+1})} \quad \text{und} \quad c_2 = \frac{\beta - \alpha}{\exp(bx_{i+1}^{k+1}) - \exp(bx_i^{k+1})}.$$

Sie ist von $x_i^{k+1}, x_{i+1}^{k+1}, \alpha, \beta$ und der Konvektion b abhängig. Mit Hilfe von u kann eine problemangepaßte Interpolation im Mittelpunkt $x_{i+1/2}^{k+1} := x_{2i+1}^k = (x_i^{k+1} + x_{i+1}^{k+1})/2 \in \Omega_k$ des betrachteten Intervalls definiert werden, die dem Transport grobskaliger Information auf die nächstfeinere Skala dient. Man berechnet dort den Wert der Lösungsfunktion und interpretiert ihn als das Ergebnis einer Interpolation

$$u(x_{i+1/2}^{k+1}) \stackrel{!}{=} p_O \alpha + p_W \beta$$

mit zu bestimmenden Interpolationsgewichten p_W und p_O. Wir erhalten

$$p_W = \frac{\exp(bx_{i+1/2}^{k+1}) - \exp(bx_i^{k+1})}{\exp(bx_{i+1}^{k+1}) - \exp(bx_i^{k+1})} \quad \text{und} \quad p_O = \frac{\exp(bx_{i+1}^{k+1}) - \exp(bx_{i+1/2}^{k+1})}{\exp(bx_{i+1}^{k+1}) - \exp(bx_i^{k+1})}.$$

Mit $x_{i+1}^{k+1} = x_i^{k+1} + 2h_k$ berechnet man hieraus die von $x_i^{k+1}, x_{i+1/2}^{k+1}$ und x_{i+1}^{k+1} freie Darstellung der Gewichte

$$p_W = \frac{1 - \exp(-bh_k)}{\exp(bh_k) - \exp(-bh_k)} \quad \text{und} \quad p_O = \frac{\exp(bh_k) - 1}{\exp(bh_k) - \exp(-bh_k)}.$$

Man erhält nach einer weiteren Rechnung

$$p_W = \frac{1}{2}\left(1 - \tanh\left(\frac{bh_k}{2}\right)\right) \quad \text{und} \quad p_O = \frac{1}{2}\left(1 + \tanh\left(\frac{bh_k}{2}\right)\right).$$

Im Hinblick auf Mehrgitter-Verfahren führt dieses Vorgehen zu dem eindimensionalen operatorabhängigen Prolongationsmolekül

$$(P_{k+1}^k)_* = [\, p_W \quad p_Z \quad p_O \,] = \left[\frac{1}{2}\left(1 - \tanh\left(\frac{bh_k}{2}\right)\right) \quad 1 \quad \frac{1}{2}\left(1 + \tanh\left(\frac{bh_k}{2}\right)\right)\right], \tag{5.2}$$

das eine problemabhängige Interpolation mit Hilfe der Lösungsfunktion u zur Randwertaufgabe (5.1) definiert. Selbstverständlich soll ein im Punkt x_i^{k+1} gegebener Wert auf feineren Skalen unverändert übernommen werden, weshalb p_Z gleich 1 zu wählen ist.

Wir geben nun eine Finite-Elemente Interpretation des Sterns (5.2). Dazu nehmen wir an, daß $\mathcal{V}_{k+1}$ ein Finite-Elemente Raum ist, der durch nodale Basisfunktionen $\varphi_{i,\mathcal{V}}^{k+1}$ aufgespannt wird, die den Gitterpunkten von Ω_{k+1} zugehören. Eine Approximation u_{k+1} der kontinuierlichen Lösung u läßt sich dann als Linearkombination

$$u_{k+1}(x) = \alpha\varphi_i^{k+1}(x) + \beta\varphi_{i+1}^{k+1}(x)$$

zweier in den Randpunkten des betrachteten Intervalls verankerten Basisfunktionen schreiben. Der lineare Ansatz bei der Wahl der Basisfunktionen $\varphi_{i,\mathcal{V}}^{k+1}$ führt zu linearen Splines, also den bekannten Hutfunktionen, und liefert auf groben Gittern und für große Werte von $|b|$ im allgemeinen keine hinreichend gute Approximation der kontinuierlichen Lösung u aufgrund des möglichen Grenzschichtverhaltens. Wählt man hingegen die operatorabhängigen Basisfunktionen

$$\varphi_{i,\mathcal{V}}^{k+1}(x) := \begin{cases} \dfrac{\exp(bx) - \exp(bx_{i-1}^{k+1})}{\exp(bx_i^{k+1}) - \exp(bx_{i-1}^{k+1})}, & x \in [x_{i-1}^{k+1}, x_i^{k+1}[, \\ \dfrac{\exp(bx_{i+1}^{k+1}) - \exp(bx)}{\exp(bx_{i+1}^{k+1}) - \exp(bx_i^{k+1})}, & x \in [x_i^{k+1}, x_{i+1}^{k+1}[, \\ 0, & \text{sonst}, \end{cases} \tag{5.3}$$

die als lokale Greensche Funktionen (L-Splines, siehe Kapitel 2) stückweise mit Hilfe der Lösungen des homogenen Problems zu den Randwerten 0 und 1 erklärt sind, so ist die Approximation sogar exakt. Es gilt

$$\begin{aligned} u(x) &= \frac{\alpha\exp(bx_{i+1}^{k+1}) - \beta\exp(bx_i^{k+1})}{\exp(bx_{i+1}^{k+1}) - \exp(bx_i^{k+1})} + \frac{(\beta - \alpha)\exp(bx)}{\exp(bx_{i+1}^{k+1}) - \exp(bx_i^{k+1})} \\ &= \frac{\alpha(\exp(bx_{i+1}^{k+1}) - \exp(bx)) + \beta(\exp(bx) - \exp(bx_i^{k+1}))}{\exp(bx_{i+1}^{k+1}) - \exp(bx_i^{k+1})} \\ &= \alpha\varphi_{i,\mathcal{V}}^{k+1}(x) + \beta\varphi_{i+1,\mathcal{V}}^{k+1}(x) = u_{k+1}(x). \end{aligned}$$

Damit lassen sich die Gewichte des soeben definierten operatorabhängigen Prolongationsmoleküls (5.2) auch als die Funktionswerte der entsprechenden operatorabhängigen Basisfunktionen an den Stellen $x_{i-1/2}^{k+1}$, x_i^{k+1} und $x_{i+1/2}^{k+1}$ interpretieren:

$$p_W = \varphi_{i,\mathcal{V}}^{k+1}(x_i^{k+1} - h_k), \quad p_Z = \varphi_{i,\mathcal{V}}^{k+1}(x_i^{k+1}) \quad \text{und} \quad p_O = \varphi_{i,\mathcal{V}}^{k+1}(x_i^{k+1} + h_k). \tag{5.4}$$

Man rechnet leicht nach, daß die durch (5.3) definierten operatorabhängigen Basisfunktionen eine *verallgemeinerte Skalierungsgleichung* [36] erfüllen:

$$\varphi_{i,\mathcal{V}}^{k+1}(x) = p_W \varphi_{2i-1,\mathcal{V}}^{k}(x) + p_Z \varphi_{2i,\mathcal{V}}^{k}(x) + p_O \varphi_{2i+1,\mathcal{V}}^{k}(x). \tag{5.5}$$

Wir werden dieses Resultat später noch in größerer Allgemeinheit beweisen. Damit kann das operatorabhängige Prolongationsmolekül (5.2) auch als verallgemeinerter Skalierungsfilter zu einer *verallgemeinerten Multiskalen-Analyse* [81] gesehen werden, die durch die geschachtelte Folge der L-Spline Finite-Elemente Räume

$$\mathcal{V}_{lt} \subset \mathcal{V}_{lt-1} \subset \cdots \subset \mathcal{V}_{k+1} \subset \mathcal{V}_k \subset \mathcal{V}_{k-1} \subset \cdots \subset \mathcal{V}_0 \subset \cdots \subset \mathcal{L}^2(\Omega)$$

gebildet wird. Ebenso wie die klassischen Filter aus der Theorie der Wavelets und Signalverarbeitung ist unser verallgemeinerter Skalierungsfilter, der den Übergang von einer gröberen Skala $k+1$ zur nächstfeineren Skala k beschreibt, örtlich konstant. Im Unterschied zu klassischen Filtern hängt er jedoch vom betrachteten Level ab. Wir verzichten der Übersichtlichkeit halber aber auf eine entsprechende Kennzeichnung der Prolongationsgewichte. Die Interpretation der Prolongationsgewichte mittels (5.4) läßt vermuten, daß man im allgemeinen Fall variabler Konvektion sogar ortsabhängige Gewichte betrachten wird. Um den Zusammenhang von $(P_{k+1}^k)_*$ mit der Sequenz der L-Spline Finite-Elemente Räume $\mathcal{V}_k$ zu verdeutlichen, schreiben wir fortan auch $(P_{k+1,\mathcal{V}}^k)_*$.

Um im nächsten Abschnitt eine zu (5.2) äquivalente matrixabhängige Prolongation herleiten zu können, berechnen wir mit Hilfe der Bilinearform

$$a(u,v) = \int_\Omega u'v' + bu'v \, dx$$

eine Steifigkeitsmatrix für den Konvektions-Diffusions Operator $Tu = -u'' + bu'$ relativ zu einer Skala k. Wir machen dazu einen Petrov–Galerkin-Ansatz, der als Ansatzraum den Raum $\mathcal{V}_k$ der Funktionen benutzt, die von den L-Splines aufgespannt werden. Für die Basis des Testraums $\mathcal{S}_k$ setzen wir lediglich voraus, daß sie aus hinreichend glatten Funktionen $\varphi_{j,\mathcal{S}}^k$ besteht, die in den ihnen zugeordneten Gitterpunkten auf 1 normiert sind und jeweils gleichen Träger mit den darin verankerten L-Splines haben. Der Testraum sei darüberhinaus eingebettet in eine entsprechende Sequenz geschachtelter Räume

$$\mathcal{S}_{lt} \subset \mathcal{S}_{lt-1} \subset \cdots \subset \mathcal{S}_{k+1} \subset \mathcal{S}_k \subset \mathcal{S}_{k-1} \subset \cdots \subset \mathcal{S}_0 \subset \cdots \subset \mathcal{L}^2(\Omega).$$

Von den zugehörigen Prolongationsmolekülen $(P_{k+1,\mathcal{S}}^k)_*$ fordern wir lediglich, daß durch sie die Werte in den Grobgitterpunkten und die oben erwähnte Trägerbedingung auf der gröberen Skala erhalten bleiben. Letzteres bedeutet, daß $(P_{k+1,\mathcal{S}}^k)_*$ maximal drei Einträge besitzt. Wir

erhalten dann durch Aufsplitten des Integrals und partielle Integration

$$\begin{aligned} a(\varphi^k_{i,\mathcal{V}}, \varphi^k_{j,\mathcal{S}}) &= \int_\Omega (\varphi^k_{i,\mathcal{V}})'[(\varphi^k_{j,\mathcal{S}})' + b\varphi^k_{j,\mathcal{S}}]\,dx \\ &= \sum_{l=1}^{1/h_k} \int_{x^k_{l-1}}^{x^k_l} (\varphi^k_{i,\mathcal{V}})'[(\varphi^k_{j,\mathcal{S}})' + b\varphi^k_{j,\mathcal{S}}]\,dx \\ &= \sum_{l=1}^{1/h_k} (\varphi^k_{i,\mathcal{V}})'\varphi^k_{j,\mathcal{S}}\big|_{x^k_{l-1}}^{x^k_l} + \int_{x^k_{l-1}}^{x^k_l} [-(\varphi^k_{i,\mathcal{V}})'' + b(\varphi^k_{i,\mathcal{V}})']\varphi^k_{j,\mathcal{S}}\,dx \\ &= \sum_{l=1}^{1/h_k} (\varphi^k_{i,\mathcal{V}})'\varphi^k_{j,\mathcal{S}}\big|_{x^k_{l-1}}^{x^k_l}, \end{aligned}$$

da die L-Splines $\varphi^k_{i,\mathcal{V}}$ die homogene Differentialgleichung im Inneren der Intervalle $]x^k_{l-1}, x^k_l[$ erfüllen. Wir berechnen damit nun die jeweiligen Kopplungen im entstehenden 3-Punkte Stern.

- Zentraler Sterneintrag:

$$\begin{aligned} a(\varphi^k_{i,\mathcal{V}}, \varphi^k_{i,\mathcal{S}}) &= (\varphi^k_{i,\mathcal{V}})'\varphi^k_{i,\mathcal{S}}\big|_{x^k_{i-1}}^{x^k_i} + (\varphi^k_{i,\mathcal{V}})'\varphi^k_{i,\mathcal{S}}\big|_{x^k_i}^{x^k_{i+1}} \\ &= (\varphi^k_{i,\mathcal{V}})'\varphi^k_{i,\mathcal{S}}\big|_{(x^k_i)^-} - (\varphi^k_{i,\mathcal{V}})'\varphi^k_{i,\mathcal{S}}\big|_{(x^k_i)^+} \\ &= b\left(\frac{1}{1-\exp(-bh_k)} - \frac{1}{1-\exp(bh_k)}\right) \\ &= b\coth\left(\frac{bh_k}{2}\right). \end{aligned}$$

- Östlicher Sterneintrag:

$$\begin{aligned} a(\varphi^k_{i+1,\mathcal{V}}, \varphi^k_{i,\mathcal{S}}) &= (\varphi^k_{i+1,\mathcal{V}})'\varphi^k_{i,\mathcal{S}}\big|_{x^k_i}^{x^k_{i+1}} = -(\varphi^k_{i+1,\mathcal{V}})'\big|_{(x^k_i)^+} \\ &= \frac{-b\exp(-bh_k)}{1-\exp(-bh_k)} = \frac{b}{2}\left(1 - \coth\left(\frac{bh_k}{2}\right)\right). \end{aligned}$$

- Westlicher Sterneintrag:

$$\begin{aligned} a(\varphi^k_{i-1,\mathcal{V}}, \varphi^k_{i,\mathcal{S}}) &= (\varphi^k_{i-1,\mathcal{V}})'\varphi^k_{i,\mathcal{S}}\big|_{x^k_{i-1}}^{x^k_i} = -(\varphi^k_{i-1,\mathcal{V}})'\big|_{(x^k_i)^-} \\ &= \frac{b\exp(bh_k)}{1-\exp(bh_k)} = -\frac{b}{2}\left(1 + \coth\left(\frac{bh_k}{2}\right)\right). \end{aligned}$$

Offensichtlich sind die Einträge unabhängig von der konkreten Gestalt der verwendeten Testfunktionen $\varphi^k_{i,\mathcal{S}}$. Man vergleiche dies auch mit der Bemerkung im Anschluß an Satz 5 in Kapitel 2. Es sei $\gamma_k := \frac{bh_k}{2}\coth\left(\frac{bh_k}{2}\right)$ ein (skalenabhängiger) Viskositätsfaktor. Wir erhalten damit bis auf den Faktor $1/h_k$ (aufgrund des Finite-Elemente Ansatzes) den Finite-Differenzen Stern des Il'in–Allen–Southwell Schemas (Kapitel 2, Schema 3)

$$(T_k)_* := [t_w \quad t_z \quad t_o] := \frac{\gamma_k}{h_k}\Big[-1 \quad 2 \quad -1\Big] + \frac{b}{2}\Big[-1 \quad 0 \quad 1\Big]. \tag{5.6}$$

Hierbei verzichten wir wie bereits im Fall der Prolongationsgewichte auf die Kennzeichnung der Skalenabhängigkeit der Sterneinträge.

5.1.2 Matrixabhängige Prolongationen

Das oben beschriebene Vorgehen zur Bestimmung einer operatorabhängigen Prolongation läßt sich nun mit Hilfe der berechneten Steifigkeitsmatrix T_k ins Diskrete übertragen. Die Forderung, daß das Ergebnis der Prolongation einer Grobgitterfunktion u_{k+1} in den Feinohne-Grobgitterpunkten $\Omega_k^f := \Omega_k \backslash \Omega_{k+1}$ durch die Lösung des homogenen lokalen Problems bestimmt ist, schreibt sich diskret als

$$(T_k P_{k+1,\mathcal{V}}^k u_{k+1})(x_i^k) = 0 \quad \text{für} \quad x_i^k \in \Omega_k^f. \tag{5.7}$$

Es sei $u_k := P_{k+1,\mathcal{V}}^k u_{k+1}$ die Interpolation einer Grobgitterfunktion aus dem Raum $\mathcal{V}_{k+1}$ in den Raum $\mathcal{V}_k$ zum nächstfeineren Gitter. Dann erhält man aus (5.7) zusammen mit den Randbedingungen $u_k(x_{i-1}^{k+1}) = 0$ und $u_k(x_{i+1}^{k+1}) = 0$

$$\begin{aligned} t_w u_k(x_{i-1}^{k+1}) + t_z u_k(x_{i-1/2}^{k+1}) + t_o u_k(x_i^{k+1}) = 0, \text{ d.h.} \quad u_k(x_{i-1/2}^{k+1}) = -\frac{t_o}{t_z} u_k(x_i^{k+1}), \\ t_w u_k(x_i^{k+1}) + t_z u_k(x_{i+1/2}^{k+1}) + t_o u_k(x_{i+1}^{k+1}) = 0, \text{ d.h.} \quad u_k(x_{i+1/2}^{k+1}) = -\frac{t_w}{t_z} u_k(x_i^{k+1}). \end{aligned}$$

Der Vergleich mit (5.2) zeigt, daß die Gewichte $p_W^m := -(t_o/t_z)$ und $p_O^m := -(t_w/t_z)$ der auf diese Weise matrixabhängig bestimmten Prolongation mit den Gewichten der operatorabhängigen Prolongation (5.2) übereinstimmen.

Für Probleme mit nichtkonstanter Konvektionsfunktion $b = b(x)$ bleiben die soeben gemachten Aussagen bestehen, wenn die Lösung der homogenen kontinuierlichen Gleichungen exakt erfolgt. Das gilt insbesondere auch für die Skalierungsgleichung (5.5), da die Finite-Elemente Basisfunktionen, die man durch exaktes Integrieren bezüglich der Elemente des groben Gitters Ω_{k+1} erhält, genauso verbogen sind wie die operatorabhängig gewichteten Summen der zugehörigen drei feinskaligen L-Spline Basisfunktionen zum Gitter Ω_k. In diesem Fall sind die Prolongationsmoleküle $(P_{k+1,\mathcal{V}}^k)_*$ und Operatorsterne $(T_k)_*$ nicht nur skalen- sondern zudem auch ortsabhängig. Man erhält so die in Punkten x_i^{k+1} des groben Gitters aufsitzenden lokalen Sterne

$$(P_{k+1,i,\mathcal{V}}^k)_* = [p_W(x_i^{k+1}) \quad p_Z(x_i^{k+1}) \quad p_O(x_i^{k+1})], \tag{5.8}$$

die beschreiben, wie beim Wechsel zur nächstfeineren Skala in x_i^{k+1} befindliche grobskalige Information an die umgebenden Feingitterpunkte x_{2i-1}^k, x_{2i}^k und x_{2i+1}^k weiterzugeben ist.

Satz 14 (Lokale Skalierungsgleichung im eindimensionalen Fall)
Die operatorabhängigen L-Splines zur eindimensionalen Konvektions-Diffusions Gleichung mit allgemeiner Konvektion $b = b(x)$ erfüllen die lokalen Skalierungsgleichungen

$$\varphi_{i,\mathcal{V}}^{k+1}(x) = p_W(x_i^{k+1})\varphi_{2i-1,\mathcal{V}}^k(x) + p_Z(x_i^{k+1})\varphi_{2i,\mathcal{V}}^k(x) + p_O(x_i^{k+1})\varphi_{2i+1,\mathcal{V}}^k(x). \tag{5.9}$$

Beweis:
Um die Notation einfach zu halten, verzichten wir im folgenden auf die Kennzeichnung $\mathcal{V}$. Der Beweis ergibt sich durch explizites Nachprüfen der behaupteten Identität in den vier relevanten Teilintervallen, in denen die feinskaligen L-Splines φ_{2i-1}^k, φ_{2i}^k und φ_{2i+1}^k einen gemeinsamen Träger mit dem Grobgitter L-Spline φ_i^{k+1} haben.

i) Nachprüfen auf $]x_{2i-2}^k, x_{2i-1}^k[$:

Es sei $\tilde{\varphi}_{2i-1}^k$ der skalierte L-Spline in x_{2i-1}^k, den man über die elementweise Konstruktion erhält, wenn man dort statt der Randbedingung 1 den Wert $\tilde{\varphi}_{2i-1}^k(x_{2i-1}^k) = p_W(x_i^{k+1})$ $(= \varphi_i^{k+1}(x_{2i-1}^k))$ fordert. Die dazu verwendeten Randwertaufgaben skalieren aufgrund der Linearität des Problems mit einem beliebigen Faktor ungleich 0 und es gilt

$$\tilde{\varphi}_{2i-1}^k = p_W(x_i^{k+1})\varphi_{2i-1}^k.$$

Da $\tilde{\varphi}_{2i-1}^k$ auf $]x_{2i-2}^k, x_{2i-1}^k[$ über die gleiche Differentialgleichung wie φ_i^{k+1} bestimmt wurde und die Randwerte 0 und $\varphi_i^{k+1}(x_{2i-1}^k)$ annimmt, erhalten wir

$$\varphi_i^{k+1}|_{]x_{2i-2}^k, x_{2i-1}^k[} = \tilde{\varphi}_{2i-1}^k|_{]x_{2i-2}^k, x_{2i-1}^k[} = p_W(x_i^{k+1})\varphi_{2i-1}^k|_{]x_{2i-2}^k, x_{2i-1}^k[}.$$

ii) Nachprüfen auf $]x_{2i-1}^k, x_{2i}^k[$:

Es sei jetzt $\varphi_{2i-1,2i}^k$ die Lösung des kontinuierlichen homogenen Problems auf dem Teilintervall $]x_{2i-1}^k, x_{2i}^k[$ mit den linken und rechten Randwerten $\varphi_{2i-1,2i}^k(x_{2i-1}^k) = p_W(x_i^{k+1})$ und $\varphi_{2i-1,2i}^k(x_{2i}^k) = 1$. Aufgrund der Linearität des Problems und der Tatsache, daß wir jeweils nur die homogenen Probleme betrachten, können wir $\varphi_{2i-1,2i}^k$ durch Superposition als Summe

$$\varphi_{2i-1,2i}^k = p_W(x_i^{k+1})\varphi_{2i-1}^k|_{]x_{2i-1}^k, x_{2i}^k[} + 1 \cdot \varphi_{2i}^k|_{]x_{2i-1}^k, x_{2i}^k[}$$

schreiben. Da $\varphi_{2i-1,2i}^k$ auf $]x_{2i-1}^k, x_{2i}^k[$ über die gleiche Differentialgleichung wie φ_i^{k+1} bestimmt wurde und die Randwerte $\varphi_i^{k+1}(x_{2i-1}^k)$ und 1 annimmt, erhalten wir

$$\varphi_i^{k+1}|_{]x_{2i-2}^k, x_{2i-1}^k[} = \varphi_{2i-1,2i}^k = p_W(x_i^{k+1})\varphi_{2i-1}^k|_{]x_{2i-1}^k, x_{2i}^k[} + 1 \cdot \varphi_{2i}^k|_{]x_{2i-1}^k, x_{2i}^k[}.$$

Das Nachprüfen der Behauptung für die verbleibenden beiden Intervallen geschieht vollkommen analog.

□

Für variable Konvektion $b = b(x)$ ist die exakte Integration der homogenen kontinuierlichen Gleichungen im allgemeinen nicht möglich. Ausgehend von einer Feingitterdiskretisierung T_k wählt man in der Praxis daher stets die soeben eingeführte matrixabhängige Prolongation. Wir bemerken ausdrücklich, daß dann in Bedingung (5.7) die lokalen Operatorsterne

$$(T_{k,i})_* = [t_w(x_i^k) \quad t_z(x_i^k) \quad t_o(x_i^k)]$$

zu beachten sind, die zu den lokalen Prolongationsmolekülen (5.8) führen.

Der mittels der Prolongationen $P_{k+1,\mathcal{V}/\mathcal{S}}^k$ durch eine Petrov–Galerkin-Vergröberung entstehende Grobgitteroperator stimmt wegen der Skalierungsbeziehung für die L-Splines mit dem direkt bezüglich der gröberen Skala diskretisierten Operator überein:

$$T_{k+1} = (P_{k+1,\mathcal{S}}^k)^t T_k P_{k+1,\mathcal{V}}^k. \tag{5.10}$$

Aufgrund ihrer besonderen Interpolationseigenschaften ist die operatorabhängige Prolongation $P_{k+1,\mathcal{V}}^k$ also mit dem gewählten Diskretisierungsverfahren kompatibel. Die Prolongationen $P_{k+1,\mathcal{S}}^k$ für die Testseite können etwa linear oder angepaßt an das transponierte Problem gewählt werden. Die zweite Wahl wird im folgenden Abschnitt motiviert.

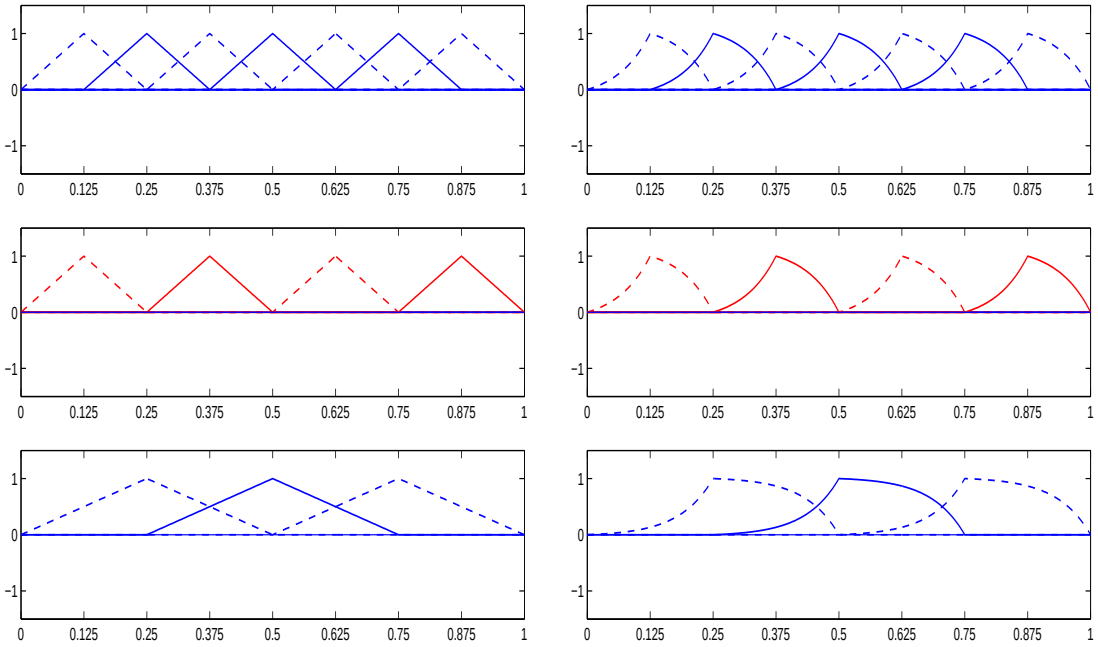


ABBILDUNG 5.1: *Klassische (links) und mittels L-Splines (rechts) gebildete einfache eindimensionale hierarchische Zerlegung auf* $\Omega =]0,1[$*, obere Bilder* $\mathcal{V}_0$*, mittlere Bilder* $\mathcal{W}_1$*, untere Bilder* $\mathcal{V}_1$*,* $h_0 = \frac{1}{8}$.

5.1.3 Zur Wahl geeigneter Basen der Komplementräume

Wir beschäftigen uns nunmehr mit der Frage der Wahl geeigneter Basen der Komplementräume $\mathcal{W}_{k+1}$ und $\mathcal{F}_{k+1}$, so daß $\mathcal{V}_k = \mathcal{W}_{k+1} \oplus \mathcal{V}_{k+1}$ und $\mathcal{S}_k = \mathcal{F}_{k+1} \oplus \mathcal{S}_{k+1}$ gelten. Dies ist gleichbedeutend mit der Auswahl geeigneter Matrizen $Q^k_{k+1,\mathcal{V}/\mathcal{S}}$, deren Spalten angeben, wie sich Basisfunktionen der jeweiligen Komplementräume mittels Basisfunktionen der zugehörigen Feingitterräume darstellen lassen. Es wird lediglich gefordert, daß die Matrizen $Q^k_{k+1,\mathcal{V}/\mathcal{S}}$ die bisherigen Prolongationsmatrizen $P^k_{k+1,\mathcal{V}/\mathcal{S}}$ zu Basiswechseln der Form $W^k_{k+1,\mathcal{V}/\mathcal{S}} = [Q^k_{k+1,\mathcal{V}/\mathcal{S}}, P^k_{k+1,\mathcal{V}/\mathcal{S}}]$ ergänzen.

Hierarchische Wahl

Die einfachste Möglichkeit für die Wahl einer Basis von $\mathcal{F}_{k+1}$ auf der Testseite besteht aus den in den Fein-ohne-Grobgitterpunkten $x^k_i \in \Omega^f_k$ verankerten Feingitterfunktionen $\varphi^k_{i,\mathcal{S}}$, was mit Hilfe des trivialen in den Fein-ohne-Grobgitterpunkten aufsitzenden Moleküls

$$(Q^k_{k+1,\mathcal{S}})_* = [\quad 1 \quad]$$

realisiert werden kann. Wir nennen diese Wahl fortan auch die *hierarchische Wahl*. Abbildung 5.1 zeigt links eine klassische und rechts eine mittels L-Splines ($b = 15$) gebildete einfache eindimensionale hierarchische Zerlegung auf dem Intervall $]0,1[$. Die oberen Bilder stellen die feinskaligen Basisfunktionen des Raums $\mathcal{V}_0$ für $h_0 = \frac{1}{8}$, dar. Die unteren Bilder zeigen die Basisfunktionen des nächstgrobskaligeren Raums $\mathcal{V}_1$ und die mittleren Bilder die Basen der hierarchisch aufgespannten Komplementräume. Zur besseren Unterscheidung werden benachbarte Basisfunktionen im Wechsel als durchgezogene und unterbrochene Kurven gezeichnet.

Mit der hierarchischen Wahl verschwinden die Einträge im Block $B_{k+1} = (Q^k_{k+1,\mathcal{S}})^t T_k P^k_{k+1,\mathcal{V}}$ der Zweiskalen-transformierten Steifigkeitsmatrix $\tilde{T}_k = (W^k_{k+1,\mathcal{S}})^t T_k W^k_{k+1,\mathcal{V}}$. Man erhält für die relevanten Einträge von B_{k+1} durch Einsetzen der Skalierungsgleichung (5.9)

$$\begin{aligned}(B_k)_{i+1,i} = a(\varphi^{k+1}_{i,\mathcal{V}}, \varphi^k_{2i+1,\mathcal{S}}) &= a(p_W \varphi^k_{2i-1,\mathcal{V}} + 1 \cdot \varphi^k_{2i,\mathcal{V}} + p_O \varphi^k_{2i+1,\mathcal{V}}, \varphi^k_{2i+1,\mathcal{S}}) \\ &= 1 \cdot a(\varphi^k_{2i,\mathcal{V}}, \varphi^k_{2i+1,\mathcal{S}}) + p_O a(\varphi^k_{2i+1,\mathcal{V}}, \varphi^k_{2i+1,\mathcal{S}}) \\ &= t_w(x^k_{2i+1}) - \frac{t_w(x^k_{2i+1})}{t_z(x^k_{2i+1})} t_z(x^k_{2i+1}) = 0,\end{aligned}$$

$$\begin{aligned}(B_k)_{i-1,i} = a(\varphi^{k+1}_{i,\mathcal{V}}, \varphi^k_{2i-1,\mathcal{S}}) &= a(p_W \varphi^k_{2i-1,\mathcal{V}} + 1 \cdot \varphi^k_{2i,\mathcal{V}} + p_O \varphi^k_{2i+1,\mathcal{V}}, \varphi^k_{2i-1,\mathcal{S}}) \\ &= p_W a(\varphi^k_{2i-1,\mathcal{V}}, \varphi^k_{2i-1,\mathcal{S}}) + 1 \cdot a(\varphi^k_{2i,\mathcal{V}}, \varphi^k_{2i-1,\mathcal{S}}) \\ &= -\frac{t_o(x^k_{2i-1})}{t_z(x^k_{2i-1})} t_z(x^k_{2i-1}) + t_o(x^k_{2i-1}) = 0.\end{aligned}$$

Durch die hierarchische Zerlegung des Testraums sind die Grobgitterunbekannten also aus den transformierten Gleichungen für die Fein-ohne-Grobgitter-Freiheitsgrade eliminiert worden.

Das folgende Argument zeigt, daß durch eine geeignete Wahl der noch nicht bestimmten Transfomationen $Q^k_{k+1,\mathcal{V}}$ und $P^k_{k+1,\mathcal{S}}$ zusätzlich auch der Block $C_{k+1} := (P^k_{k+1,\mathcal{S}})^t T_k Q^k_{k+1,\mathcal{V}}$ von $\tilde{T}_k$ eliminiert werden kann. Es hat dann $\tilde{T}_k$ eine blockdiagonale Gestalt. Dies entspricht der Eliminierung der Fein-ohne-Grobgitterunbekannten aus den transformierten Grobgittergleichungen und man erreicht damit eine Entkopplung des Systems. Wir wählen dazu $P^k_{k+1,\mathcal{S}}$ matrixabhängig von T^t_k, dem transponierten Feingitteroperator, sowie $Q^k_{k+1,\mathcal{V}}$ hierarchisch. Der Komplementraum $\mathcal{W}_{k+1}$ von $\mathcal{V}_{k+1}$ relativ zu $\mathcal{V}_k$ wird also von den in den Fein-ohne--Grobgitterpunkten befindlichen feinskaligen L-Splines zum ursprünglichen Operator T aufgespannt. Wir zeigen später bei der Besprechung zweidimensionaler Probleme, daß diese Wahl von $P^k_{k+1,\mathcal{S}}$ auch durch die zu (5.7) duale Forderung

$$((P^k_{k+1,\mathcal{S}})^t T_k u_k)(x^k_i) = 0 \quad \text{für } x^k_i \in \Omega^f_k \tag{5.11}$$

gegeben ist. Vertauschen wir die Ansatz- und Testseite — wir machen dies durch das Symbol $\mathcal{V}\circlearrowleft\mathcal{S}$ kenntlich — und transformieren die transponierte Steifigkeitsmatrix $T_{k,\mathcal{V}\circlearrowleft\mathcal{S}} = T^t_k$ mit Hilfe der vertauschten Zweiskalen-Transformationen, so erhalten wir

$$\begin{aligned}\tilde{T}_{k,\mathcal{V}\circlearrowleft\mathcal{S}} &:= (W^k_{k+1,\mathcal{V}})^t T^t_k W^k_{k+1,\mathcal{S}} \\ &= \begin{pmatrix} (Q^k_{k+1,\mathcal{V}})^t T^t_k Q^k_{k+1,\mathcal{S}} & (Q^k_{k+1,\mathcal{V}})^t T^t_k P^k_{k+1,\mathcal{S}} \\ (P^k_{k+1,\mathcal{V}})^t T^t_k Q^k_{k+1,\mathcal{S}} & (P^k_{k+1,\mathcal{V}})^t T^t_k P^k_{k+1,\mathcal{S}} \end{pmatrix},\end{aligned}$$

Hierbei ist $(Q^k_{k+1,\mathcal{V}})^t T^t_k P^k_{k+1,\mathcal{S}} = \mathbf{0}$. Durch Transponieren von $\tilde{T}_{k,\mathcal{V}\circlearrowleft\mathcal{S}}$ erhalten wir dann wegen $B_{k+1} = (Q^k_{k+1,\mathcal{S}})^t T_k P^k_{k+1,\mathcal{V}} = \mathbf{0}$

$$\begin{aligned}\tilde{T}_k &= ((W^k_{k+1,\mathcal{V}})^t T^t_k W^k_{k+1,\mathcal{S}})^t \\ &= \begin{pmatrix} (Q^k_{k+1,\mathcal{S}})^t T_k Q^k_{k+1,\mathcal{V}} & \mathbf{0} \\ \mathbf{0} & (P^k_{k+1,\mathcal{S}})^t T_k P^k_{k+1,\mathcal{V}} \end{pmatrix},\end{aligned}$$

also die gewünschte blockdiagonale Form. Aufgrund der hierarchischen Wahl der Basen von $\mathcal{W}_{k+1}$ und $\mathcal{F}_{k+1}$ ist $A_{k+1} = (Q^k_{k+1,\mathcal{S}})^t T_k Q^k_{k+1,\mathcal{V}}$ diagonal. Da der untere Hauptdiagonalblock

$T_{k+1} = (P^k_{k+1,\mathcal{S}})^t T_k P^k_{k+1,\mathcal{V}}$ wieder durch einen 3-Punkte Stern gegeben ist, kann die matrixabhängige Bestimmung der Prolongationsmatrizen in Bezug auf die Ansatz- und Testräume rekursiv fortgesetzt werden. Die Standardform $\tilde{T}_{0,s}$ des Feingitteroperators T_0 zu den so erhaltenen problemabhängigen Multiskalen-Zerlegungen der Ansatz- und Testräume $\mathcal{V}_0$ und $\mathcal{S}_0$ ist damit diagonal und das transformierte System kann trivial gelöst werden. Für die Nichtstandardform $\tilde{T}_{0,ns}$ ergibt sich entsprechend, daß sämtliche Nebendiagonalblöcke verschwinden. Beide Implementierungen des additiven Verfahrens aus Kapitel 4.3 (mittels Standard- und Nichtstandardsystem) entarten hier also zu einem direkten Löser.

Werden die Prolongationen $P^k_{k+1,\mathcal{S}}$ für den Testraum starr gewählt, zum Beispiel als lineare Interpolationen, so verschwinden die Blöcke C_k von $\tilde{T}_{0,ns}$ für $k = 1, ..., lt$ im allgemeinen nicht mehr. Ein Schritt einer Multiskalen-Vorwärtsubstitution liefert hier dennoch die exakte Lösung (aus Sicht des Standardsystems). In gleicher Weise besitzt $\tilde{T}_{0,s}$ untere Dreiecksgestalt und eine gewöhnliche Vorwärtssubstitution des auf Standardform transformierten Systems ergibt dessen exakte Lösung.

Wavelet-artige Wahl

In höheren Raumdimensionen erweist sich leider die eben besprochene hierarchische Wahl als ungünstig, denn

- die Elimination der Nebendiagonalblöcke ist nur unvollständig [54] und
- die Stabilitätseigenschaften der dadurch erzeugten Zerlegungen sind nicht mehr optimal wie im eindimensionalen Fall [91].

Benutzt man zur Konstruktion der Komplementräume in höheren Dimensionen (Pre-) Wavelets statt der feinskaligen Basisfunktionen, so kann man zeigen, daß für symmetrisch positiv definite Probleme die Standardform $\tilde{T}_{0,s}$ spektral äquivalent zu einer Diagonalmatrix ist [18, 33, 75, 91]. Wir beschäftigen uns daher in diesem Abschnitt mit der Konstruktion eindimensionaler Wavelet-artiger Komplementräume. Mit ihrer Hilfe lassen sich über einen Tensorprodukt-Ansatz Wavelet-artige Multiskalen-Zerlegungen in höheren Raumdimensionen einfach konstruieren. Wie wir in Kapitel 4.2 gezeigt haben, erzeugt man damit innerhalb von Multiskalen-Verfahren Unterraumkorrekturen, die aus Mehrgitter-Sicht verbesserte Glättungseigenschaften gegenüber solchen aufweisen, die durch hierarchische Zerlegungen entstehen.

Es seien $\mathcal{V}_k$ und $\mathcal{V}_{k+1}$ zunächst die Räume, die von eindimensionalen linearen Splines bezüglich Ω_k und Ω_{k+1} aufgespannt werden. Dann gilt nach Definition für den Raum $\mathcal{W}_{k+1}$ der zugehörigen linearen Prewavelets [91]

$$\mathcal{V}_k = \mathcal{W}_{k+1} \oplus_{\mathcal{L}_2} \mathcal{V}_{k+1}.$$

Lineare Prewavelets zu eindimensionalen linearen Splines werden als lokale Basisfunktionen von $\mathcal{W}_{k+1}$ durch die Forderung bestimmt, daß sie $\mathcal{L}_2(\Omega)$-orthogonal zu den in den Grobgitterpunkten befindlichen grobskaligen Hutfunktionen sind. Sie werden in nicht unmittelbar randnahen Fein-ohne-Grobgitterpunkten durch das zentral darin aufsitzende Molekül

$$(Q^k_{k+1,\mathcal{V}})_* = [1 \quad -6 \quad \underline{10} \quad -6 \quad 1]/10$$

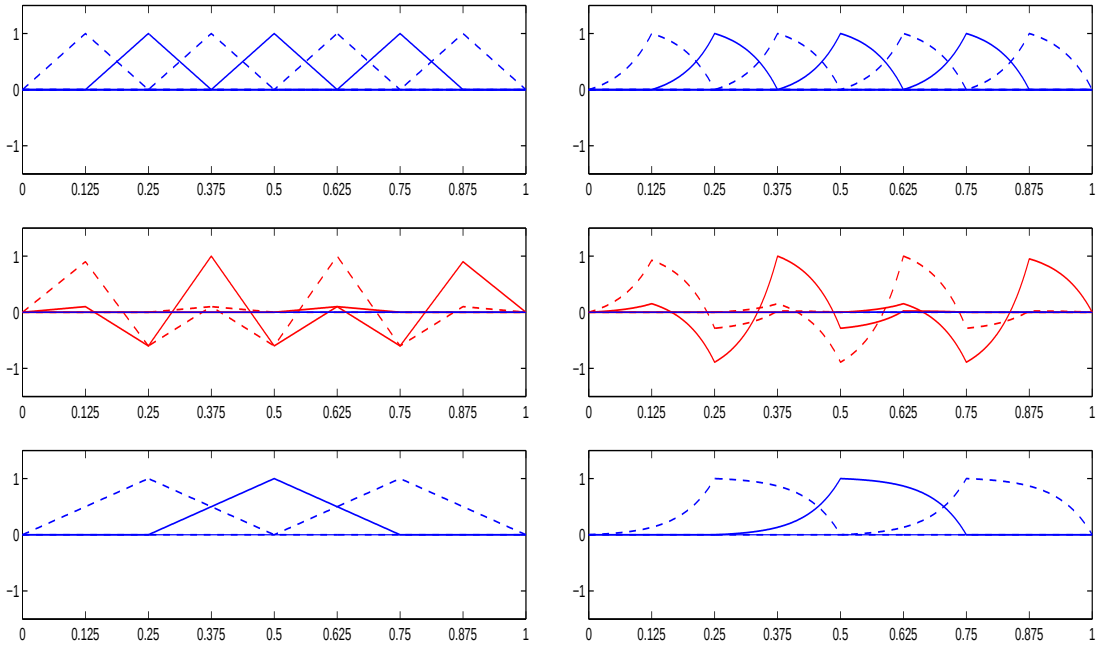


ABBILDUNG 5.2: *Lineare (links) und mittels L-Spline Prewavelets (rechts, $b = 15$) gebildete einfache eindimensionale Prewavelet-Zerlegung auf $\Omega =]0,1[$, obere Bilder $\mathcal{V}_0$, mittlere Bilder $\mathcal{W}_1$, untere Bilder $\mathcal{V}_1$, $h_0 = \frac{1}{8}$.*

beschrieben. Das zum Aufsatzpunkt gehörende Gewicht ist unterstrichen. Für den linken und rechten randnahen Punkt erhält man die links- beziehungsweise rechtsbündig aufsitzenden Moleküle

$$(Q^k_{k+1,\mathcal{V},l})_* = [\underline{9} \quad -6 \quad 1]/10 \quad \text{und} \quad (Q^k_{k+1,\mathcal{V},r})_* = [1 \quad -6 \quad \underline{9}]/10.$$

Die Sterne geben an, wie sich ein im betreffenden Fein-ohne-Grobgitterpunkt befindliches Prewavelet als Linearkombination der umliegenden feinskaligen Hutfunktionen darstellen läßt. Es werden dadurch Verfeinerungsgleichungen definiert, die die Teilmatrix $Q^k_{k+1,\mathcal{V}}$ des Basiswechsels $W^k_{k+1,\mathcal{V}} : \mathcal{W}_{k+1} \oplus_{\mathcal{L}_2} \mathcal{V}_{k+1} \longrightarrow \mathcal{V}_k$ eindeutig festlegen. Abbildung 5.2 zeigt links (analog zur linken Seite von Abbildung 5.1) eine mittels linearer Prewavelets gebildete einfache eindimensionale Zerlegung. Die rechten Bilder zeigen eine Prewavelet-Zerlegung, die zu L-Splines gehört und auf die wir gleich zu sprechen kommen. Im Vergleich zu Abbildung 5.1 haben sich nur die mittleren Bilder geändert, die die Basen der neu gewählten Komplementräume darstellen.

Wir zeigen nun, wie man die Filterkoeffizienten in den nicht unmittelbar randnahen Fein-ohne-Grobgitterpunkten berechnen kann. Die beiden Randfilter werden auf analoge Weise bestimmt. Man setzt dazu ein in einem Punkt $x^k_{2i-1} \in \Omega^f_k$ zu konstruierendes Prewavelet $\psi^{k+1}_{i,\mathcal{V}}$ als eine Linearkombination von fünf feinskaligen Hutfunktionen an, die sich in den Gitterpunkten x^k_{2i-3}, x^k_{2i-2}, x^k_{2i-1}, x^k_{2i} und x^k_{2i+1}, befinden:

$$\psi^{k+1}_{i,\mathcal{V}} := q_1\varphi^k_{2i-3,\mathcal{V}} + q_2\varphi^k_{2i-2,\mathcal{V}} + 1 \cdot \varphi^k_{2i-1,\mathcal{V}} + q_3\varphi^k_{2i,\mathcal{V}} + q_4\varphi^k_{2i+1,\mathcal{V}}. \tag{5.12}$$

Wir haben dabei den Beitrag von $\varphi^k_{2i-1,\mathcal{V}}$ durch 1 fixiert, was einer Skalierung des Moleküls entspricht. Es hat ψ^{k+1}_i dann nur mit den in x^k_{2i-4}, x^k_{2i-2}, x^k_{2i} und x^k_{2i+2} verankerten grobskaligen Hutfunktionen $\varphi^{k+1}_{i-2,\mathcal{V}}$, $\varphi^{k+1}_{i-1,\mathcal{V}}$, $\varphi^{k+1}_{i,\mathcal{V}}$ und $\varphi^{k+1}_{i+1,\mathcal{V}}$, einen gemeinsamen Träger. Durch

sie erhalten wir dann die folgenden vier Orthogonalitätsbedingungen zur Bestimmung der unbekannten Gewichte $q_1, ..., q_4$:

$$(\varphi^{k+1}_{i-2,\mathcal{V}}, \psi^{k+1}_{i,\mathcal{V}}) = 0, \qquad (\varphi^{k+1}_{i-1,\mathcal{V}}, \psi^{k+1}_{i,\mathcal{V}}) = 0,$$
$$(\varphi^{k+1}_{i,\mathcal{V}}, \psi^{k+1}_{i,\mathcal{V}}) = 0, \qquad (\varphi^{k+1}_{i+1,\mathcal{V}}, \psi^{k+1}_{i,\mathcal{V}}) = 0.$$

Ersetzt man hierbei die jeweiligen Basisfunktionen aus $\mathcal{V}_{k+1}$ durch die über das Prolongationsmolekül $(P^k_{k+1,\mathcal{V}})_*$ gegebene gewichtete Summe (5.12) feinskaliger Basisfunktionen aus $\mathcal{V}_k$, so erhält man ein lineares Gleichungssystem für die vier Unbekannten $q_1, ..., q_4$, dessen Koeffizienten sich aus $\mathcal{L}_2(\Omega)$-Integralen koppelnder feinskaliger Hutfunktionen ergeben.

Diese Konstruktion läßt sich sofort auf den Fall übertragen, daß $\mathcal{V}_k$ und $\mathcal{V}_{k+1}$ von uniformen eindimensionalen L-Splines bezüglich Ω_k und Ω_{k+1} aufgespannt werden. Es müssen dann die $\mathcal{L}_2(\Omega)$-Integrale koppelnder feinskaliger L-Splines sowie die zugehörigen operatorabhängigen Prolongationsmoleküle $(P^k_{k+1,\mathcal{V}})_*$ beim Aufstellen der linearen Gleichungssysteme zur Berechnung der Detailfilterkoeffizienten berücksichtigt werden. Wir erhalten dadurch problemangepaßte Filter $(Q^k_{k+1,\mathcal{V}})_*$ und damit problemangepaßte Prewavelets, kurz *L-Spline Prewavelets*, die eine problemabhängige Basis des Komplementraums $\mathcal{W}_{k+1}$ bilden, so daß wiederum $\mathcal{V}_k = \mathcal{W}_{k+1} \oplus_{\mathcal{L}_2} \mathcal{V}_{k+1}$ gilt. Abbildung 5.2 zeigt rechts (analog zur rechten Seite von Abbildung 5.1) die entsprechende mittels L-Spline Prewavelets gebildete einfache eindimensionale Zerlegung zur moderaten Konvektionsstärke $b = 15$. Im Fall extremer Konvektion mit

$$(P^{k,\infty}_{k+1,\mathcal{V}})_* := \lim_{b\to\infty} (P^k_{k+1,\mathcal{V}})_* = [0 \quad 1 \quad 1]$$

entarten die darüber gebildeten L-Splines zur bekannten Haar-Funktion [36]. Als zugehörige L-Spline Prewavelets erhält man wegen

$$(Q^{k,\infty}_{k+1,\mathcal{V}})_* = [0 \quad -1 \quad \underline{1} \quad 0 \quad 0], \tag{5.13}$$
$$(Q^{k,\infty}_{k+1,\mathcal{V},l})_* = [\underline{1} \quad 0 \quad 0] \quad \text{und} \quad (Q^{k,\infty}_{k+1,\mathcal{V},r})_* = [0 \quad -1 \quad \underline{1}] \tag{5.14}$$

genau das Haar-Wavelet zusammen mit den entsprechenden Randmodifikationen. Der Übergang innerhalb unserer operatorabhängigen Skalierungs- und Waveletfilter-Konstruktion zwischen dem Fall $b = 0$ (reiner Laplace-Operator) und dem Fall $b = \infty$ (reiner Konvektions-Operator) ist stetig. Dies wird durch Abbildung 5.3 deutlich. Sie zeigt für zunehmende Konvektionsstärken L-Spline Prewavelets im Innern des Gebiets sowie die zugehörigen Randmodifikationen in den unmittelbar randbenachbarten Feingitterpunkten relativ zu den umgebenden L-Spline Basisfunktionen (mit unterbrochenen Kurven gezeichnet).

Für Probleme mit nichtkonstanter Konvektionsfunktion lassen sich ausgehend von lokalen in Punkten x^{k+1}_j des groben Gitters befindlichen Molekülen $(P^k_{k+1,j,\mathcal{V}})_*$ lokale Filter $(Q^k_{k+1,i,\mathcal{V}})_*$ berechnen, die in Fein-ohne-Grobgitterpunkten $x^k_i \in \Omega^f_k$ verankert sind. Man erhält so *lokale L-Spline Prewavelets* und den von ihnen aufgespannten Komplementraum $\mathcal{W}_{k+1}$, der wiederum $\mathcal{L}_2(\Omega)$-orthogonal zu dem von lokal definierten L-Splines aufgespannten Raum $\mathcal{V}_{k+1}$ ist. Wir bemerken schließlich, daß eine Basis des Komplementraums $\mathcal{W}_{k+1}$, die man über feste Filtermasken $(Q^k_{k+1,\mathcal{V}})_*$ im Zusammenhang mit einer problemabhängigen Wahl der Räume $\mathcal{V}_k$ und $\mathcal{V}_{k+1}$ konstruiert, trotzdem problemabhängig ist. Es werden hierbei Basisfunktionen von $\mathcal{W}_{k+1}$ nach einer festen Vorschrift aus problemabhängigen feinskaligen Basisfunktionen von $\mathcal{V}_k$ aufgebaut. Dann gilt allerdings nur noch $\mathcal{V}_k = \mathcal{W}_{k+1} \oplus \mathcal{V}_{k+1}$. Aus unseren numerischen Beispielen in Kapitel 6 wird deutlich, daß gerade die Kombination problemabhängiger

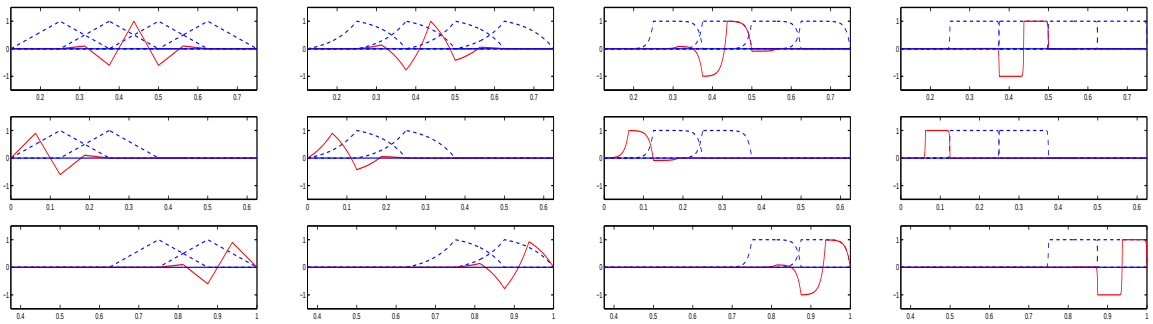


ABBILDUNG 5.3: *Entwicklung der eindimensionalen (Rand-) L-Spline Prewavelets für zunehmende Konvektionsstärke* ($b = 0, 15, 100, 1000$) *von links nach rechts.*

Skalierungsfilter mit fest gewählten Detailfiltern, etwa

$$(Q_{k,\mathcal{V}}^{k-1})_* = [-1 \quad \underline{2} \quad -1]/2 \qquad (\textit{Pre-Prewavelet Filter}), \tag{5.15}$$

$$(Q_{k,\mathcal{V}}^{k-1})_* = [1 \quad -6 \quad \underline{10} \quad -6 \quad 1]/10 \qquad (\textit{Prewavelet Filter}) \tag{5.16}$$

und zugehörige Randmodifikationen, im Rahmen eines Tensorprodukt-Ansatzes die besten Resultate liefert.

5.2 Probleme in 2D: Tensorprodukt-Konstruktionen

5.2.1 Operatorabhängige Prolongationen

Im folgenden erweitern wir die Überlegungen des letzten Abschnitts auf den zweidimensionalen Fall. Wir betrachten dazu zuerst den formalen zweidimensionalen Konvektions-Diffusions Operator mit konstanter Konvektion $\vec{b} = (b_1, b_2) \neq \mathbf{0}$ bezüglich $\Omega =]0,1[^2$

$$T = -\triangle + b_1\partial_x + b_2\partial_y. \tag{5.17}$$

Es läßt sich dann T additiv in die Richtungsanteile

$$T_x := -\partial_x^2 + b_1\partial_x \quad \text{und} \quad T_y := -\partial_y^2 + b_2\partial_y$$

aufspalten, so daß formal gilt

$$T = T_x + T_y. \tag{5.18}$$

Es sei $\Omega_{k+1} := \{(x_i^{k+1}, y_j^{k+1}) \in \Omega : x_i^{k+1} := ih_{k+1} \text{ und } y_j^{k+1} := jh_{k+1} \text{ für } 0 < i, j < 1/h_{k+1}\}$ ein uniformes quadratisches Gitter mit Maschenweite $h_{k+1} = 2h_k$. Wir nehmen an, daß Ω_{k+1} in eine geschachtelte Folge $\Omega_{lt} \subset \cdots \subset \Omega_{k+1} \subset \Omega_k \subset \cdots \subset \Omega_0$ dyadisch verfeinerter quadratischer Gitter eingebettet ist. Mittels einer klassischen Vierfarbenunterteilung [62] kann das nächstfeinere Gitter Ω_k in vier disjunkte Mengen unterteilt werden:

$$\begin{aligned}
\Omega_k^{(1,1)} &:= \{(x_i^k, y_j^k) \in]0,1[^2 : x_i^k = ih_k, y_j^k = jh_k, 0 < i, j < 1/h_k, i \text{ ungerade}, j \text{ ungerade}\},\\
\Omega_k^{(0,1)} &:= \{(x_i^k, y_j^k) \in]0,1[^2 : x_i^k = ih_k, y_j^k = jh_k, 0 < i, j < 1/h_k, i \text{ gerade}, j \text{ ungerade}\},\\
\Omega_k^{(1,0)} &:= \{(x_i^k, y_j^k) \in]0,1[^2 : x_i^k = ih_k, y_j^k = jh_k, 0 < i, j < 1/h_k, i \text{ ungerade}, j \text{ gerade}\},\\
\Omega_k^{(0,0)} &:= \{(x_i^k, y_j^k) \in]0,1[^2 : x_i^k = ih_k, y_j^k = jh_k, 0 < i, j < 1/h_k, i \text{ gerade}, j \text{ gerade}\},\\
&= \Omega_{k+1}.
\end{aligned}$$

Die Punkte aus $\Omega_k^f := \Omega_k^{(1,1)} \cup \Omega_k^{(0,1)} \cup \Omega_k^{(1,0)}$ bilden die Fein-ohne-Grobgitterpunkte des Gitters Ω_k, wohingegen $\Omega_k^c := \Omega_k^{(0,0)}$ die Menge der Grobgitterpunkte ist, die aufgrund der uniformen dyadischen Verfeinerung das nächstgröbere Gitter Ω_{k+1} darstellt. Wir betrachten nun einen Gitterpunkt $(x_i^{k+1}, y_j^{k+1}) \in \Omega_{k+1}$. Löst man dort mittels der eindimensionalen Technik aus dem vorherigen Unterkapitel separat die zugehörigen homogenen Gleichungen bezüglich T_x und T_y zu allgemeinen Randwerten, so ergeben sich für beide Richtungen die eindimensionalen Prolongationsmoleküle (siehe Formel (5.2))

$$(P_{k+1}^{k,x})_* = [p_W \quad p_{Z,x} \quad p_O] = \left[\frac{1}{2}\left(1-\tanh\left(\frac{b_1 h_k}{2}\right)\right) \quad 1 \quad \frac{1}{2}\left(1+\tanh\left(\frac{b_1 h_k}{2}\right)\right)\right],$$

$$(P_{k+1}^{k,y})_* = [p_N \quad p_{Z,y} \quad p_S] = \left[\frac{1}{2}\left(1-\tanh\left(\frac{b_2 h_k}{2}\right)\right) \quad 1 \quad \frac{1}{2}\left(1+\tanh\left(\frac{b_2 h_k}{2}\right)\right)\right].$$

Sie hängen nur von T_x respektive T_y ab und legen die Werte einer von uns zu bestimmenden operatorabhängigen Prolongation mit 9-Punkte Sternmuster in den Gitterpunkten aus $\Omega_k^{(0,1)}$ und $\Omega_k^{(1,0)}$ fest:

$$P_{k+1}^k = \begin{bmatrix} p_{NW} & p_N & p_{NO} \\ p_W & p_Z & p_O \\ p_{SW} & p_S & p_{SO} \end{bmatrix}.$$

Die eindimensionalen operatorabhängigen Prolongationssterne $(P_{k+1}^{k,x})_*$ und $(P_{k+1}^{k,y})_*$ können mit Hilfe eindimensionaler L-Splines $\varphi_{(i,j),\mathcal{V}}^{k+1,x}$ und $\varphi_{(i,j),\mathcal{V}}^{k+1,y}$ interpretiert werden, die jeweils eindimensionale operatorabhängige Finite-Elemente Räume $\mathcal{V}_{k+1}^{x,j}$ und $\mathcal{V}_{k+1}^{y,i}$ entlang der Gitterlinien $y = y_j^{k+1} = const$ und $x = x_i^{k+1} = const$ definieren. Legt man eine Anordnung der Gitterlinien in x- und y-Richtung fest, so können diese Räume zu abstrakten eindimensionalen Räumen $\mathcal{V}_{k+1}^x$ und $\mathcal{V}_{k+1}^y$ zusammengefaßt werden.
Die so gefundenden eindimensionalen L-Splines liefern die erforderlichen Randbedingungen zur Bestimmung der verbleibenden Gewichte p_{NW}, p_{NO}, p_{SW} und p_{SO} in den Punkten $(x_{i\pm 1/2}^{k+1}, y_{j\pm 1/2}^{k+1}) \in \Omega_k^f$ über homogene lokale Probleme. Wir illustrieren das Vorgehen anhand der Berechnung von p_{SW} in $(x_{i-1/2}^{k+1}, y_{j-1/2}^{k+1})$. Wir lösen dazu das homogene Dirichletproblem

$$-\triangle u + b_1 \partial_x u + b_2 \partial_y u = 0 \tag{5.19}$$

auf dem Element $]x_{i-1}^{k+1}, x_i^{k+1}[\times]y_{j-1}^{k+1}, y_j^{k+1}[$ mit den Randbedingungen

$$\begin{aligned}
u|_{[(x_{i-1}^{k+1}, y_{j-1}^{k+1}),(x_i^{k+1}, y_{j-1}^{k+1})]} &= 0, \\
u|_{[(x_{i-1}^{k+1}, y_{j-1}^{k+1}),(x_{i-1}^{k+1}, y_j^{k+1})]} &= 0, \\
u|_{[(x_i^{k+1}, y_{j-1}^{k+1}),(x_i^{k+1}, y_j^{k+1})]} &= \varphi_{(i,j),\mathcal{V}}^{k+1,y}|_{[(x_i^{k+1}, y_{j-1}^{k+1}),(x_i^{k+1}, y_j^{k+1})]}, \\
u|_{[(x_{i-1}^{k+1}, y_j^{k+1}),(x_i^{k+1}, y_j^{k+1})]} &= \varphi_{(i,j),\mathcal{V}}^{k+1,x}|_{[(x_{i-1}^{k+1}, y_j^{k+1}),(x_i^{k+1}, y_j^{k+1})]}
\end{aligned} \tag{5.20}$$

und definieren

$$p_{SW} := u(x_{i-1/2}^{k+1}, y_{j-1/2}^{k+1}).$$

In analoger Weise werden p_{NW}, p_{NO} und p_{SO} berechnet, so daß zusammen mit $p_Z := 1$ schließlich P_{k+1}^k vollständig bestimmt ist.

Durch Tensorproduktbildung erhält man aus den eindimensionalen Räumen $\mathcal{V}^x_{k+1}$ und $\mathcal{V}^y_{k+1}$ den zweidimensionalen operatorabhängigen Finite-Elemente Raum

$$\begin{aligned} \mathcal{V}_{k+1} &:= \mathcal{V}^x_{k+1} \otimes \mathcal{V}^y_{k+1} \\ &:= \left\langle \varphi^{k+1}_{(i,j),\mathcal{V}} \,:\, \varphi^{k+1}_{(i,j),\mathcal{V}}(x,y) = \varphi^{k+1,x}_{(i,j),\mathcal{V}}(x)\varphi^{k+1,y}_{(i,j),\mathcal{V}}(y) \text{ für } 0 < i,j < 1/h_{k+1} \right\rangle. \end{aligned} \tag{5.21}$$

Hierbei bezeichnet $\langle \quad \rangle$ den Span der entsprechenden Basisfunktionen. Die Werte des zweidimensionalen Prolongationsmoleküls $(P^k_{k+1})_*$ in den (x^{k+1}_i, y^{k+1}_j) unmittelbar umgebenden Punkten des feineren Gitters Ω_k können dann direkt als die dortigen Funktionswerte der zweidimensionalen Finite-Elemente Basisfunktion $\varphi^{k+1}_{(i,j),\mathcal{V}}$, die wir im folgenden auch *Tensorprodukt-L-Splines* nennen, abgelesen werden:

$$\begin{aligned} p_N &= \varphi^{k+1}_{(i,j),\mathcal{V}}(x^{k+1}_i, y^{k+1}_j + h_k), & p_S &= \varphi^{k+1}_{(i,j),\mathcal{V}}(x^{k+1}_i, y^{k+1}_j - h_k), \\ p_W &= \varphi^{k+1}_{(i,j),\mathcal{V}}(x^{k+1}_i - h_k, y^{k+1}_j), & p_O &= \varphi^{k+1}_{(i,j),\mathcal{V}}(x^{k+1}_i + h_k, y^{k+1}_j), \\ p_Z &= \varphi^{k+1}_{(i,j),\mathcal{V}}(x^{k+1}_i, y^{k+1}_j) & &\text{sowie} \\ p_{NW} &= \varphi^{k+1}_{(i,j),\mathcal{V}}(x^{k+1}_i - h_k, y^{k+1}_j + h_k), & p_{NO} &= \varphi^{k+1}_{(i,j),\mathcal{V}}(x^{k+1}_i + h_k, y^{k+1}_j + h_k), \\ p_{SW} &= \varphi^{k+1}_{(i,j),\mathcal{V}}(x^{k+1}_i - h_k, y^{k+1}_j - h_k), & p_{SO} &= \varphi^{k+1}_{(i,j),\mathcal{V}}(x^{k+1}_i + h_k, y^{k+1}_j - h_k). \end{aligned} \tag{5.22}$$

Das gilt insbesondere auch für die Funktionswerte in den Punkten $x^k_i \in \Omega^{(1,1)}_k$. Da T nach Voraussetzung additiv in die eindimensionalen Richtungsanteile T_x und T_y zerfällt, erfüllen die Einschränkungen der zweidimensionalen Tensorprodukt-L-Splines auf die Elemente von Ω_{k+1} dort die homogene kontinuierliche Differentialgleichung zusammen mit den oben gestellten Randbedingungen. Aufgrund der Eindeutigkeit ihrer Lösung stimmen sie also genau mit der Lösung u von (5.19), (5.20) überein. Man beachte weiterhin, daß die Eckeinträge aufgrund der Tensorprodukt-Struktur von $\mathcal{V}_{k+1}$ als Produkte geschrieben werden können:

$$\begin{aligned} p_{NW} &= p_N \cdot p_W, & p_{NO} &= p_N \cdot p_O, \\ p_{SW} &= p_S \cdot p_W, & p_{SW} &= p_S \cdot p_O. \end{aligned} \tag{5.23}$$

Die zweidimensionalen Tensorprodukt-L-Splines genügen wiederum einer verallgemeinerten Skalierungsgleichung und die operatorabhängigen zweidimensionalen Prolongationsmoleküle $(P^k_{k+1})_*$ können als verallgemeinerte Skalierungsfilter zu einer verallgemeinerten Multiskalen-Analyse angesehen werden, die durch eine geschachtelte Sequenz eben dieser operatorabhängigen Tensorprodukt-Finite-Elemente Räume gebildet wird. Um den Zusammenhang von $(P^k_{k+1})_*$ mit der Sequenz der Tensorprodukt-L-Spline Finite-Elemente Räume $\mathcal{V}_k$ zu verdeutlichen, schreiben wir fortan wieder $(P^k_{k+1,\mathcal{V}})_*$ sowie auch $(P^{k,x}_{k+1,\mathcal{V}})_*$ und $(P^{k,y}_{k+1,\mathcal{V}})_*$.

Zur Herleitung einer entsprechenden matrixabhängigen Prolongation berechnen wir mittels der Bilinearform

$$a(u,v) = \int_\Omega \nabla u \cdot \nabla v + b_1(\partial_x u)v + b_2(\partial_y u)v \, dxdy$$

eine Steifigkeitsmatrix für den Konvektions-Diffusions Operator $T = -\triangle + b_1\partial_x + b_2\partial_y$ relativ zu einer Skala k. Wir machen dazu wie im eindimensionalen Fall einen Petrov–Galerkin-Ansatz mit einem von $\mathcal{V}_k$ verschiedenen Testraum $\mathcal{S}_k$, der ebenfalls über das Tensorprodukt eindimensionaler Räume $\mathcal{S}^x_k$ und $\mathcal{S}^y_k$ entstanden und in eine entsprechende Sequenz geschachtelter Räume eingebettet sei. Die am Aufbau von $\mathcal{S}^x_k$ und $\mathcal{S}^y_k$ beteiligten eindimensionalen

Räume $\mathcal{S}_k^{x,j}$ und $\mathcal{S}_k^{y,i}$ entlang der Gitterlinien mögen die gleichen Voraussetzungen wie im eindimensionalen Fall erfüllen. Es ergeben sich dann ausgehend von den eindimensionalen Räumen $\mathcal{V}_k^{x,j}$ und $\mathcal{S}_k^{x,j}$ sowie $\mathcal{V}_k^{y,i}$ und $\mathcal{S}_k^{y,i}$ als Diskretisierungen der Richtungsanteile T_x und T_y sowie der eindimensionalen schwachen Identitäten bezüglich beider Richtungen die Sterne

$$\begin{array}{rclcrcl} (T_k^x)_* & := & [t_w \quad t_{z,x} \quad t_o], & & (T_k^y)_* & := & [t_n \quad t_{z,y} \quad t_s], \\ (M_k^x)_* & := & [m_w \quad m_{z,x} \quad m_o], & & (M_k^y)_* & := & [m_n \quad m_{z,y} \quad m_s]. \end{array}$$

Man erhält damit aufgrund des Tensorprodukt-Ansatzes und der Tatsache, daß T separabel ist, als zweidimensionalen Operatorstern die folgende Summe von dyadischen Produkten eindimensionaler Sterne:

$$(T_k)_* = (M_k^y)_*^t \cdot (T_k^x)_* + (T_k^y)_*^t \cdot (M_k^x)_*. \tag{5.24}$$

5.2.2 Matrixabhängige Prolongation

Mit Hilfe der gerade berechneten Steifigkeitsmatrix T_k und einer speziellen „Plättungstechnik“ läßt sich das oben beschriebene Vorgehen zur Bestimmung einer operatorabhängigen Prolongation ins Diskrete übertragen. In den Punkten aus $\Omega_k^{(0,0)}$ verwenden wir selbstverständlich $p_Z^m := 1$ als Wert der matrixabhängigen Prolongation. Zur Berechnung der Gewichte p_W^m, p_O^m, p_N^m und p_S^m in den Punkten aus $\Omega_k^{(1,0)}$ und $\Omega_k^{(0,1)}$ definiert man eindimensionale 3-Punkte Sterne $(\breve{T}_k^x)_*$ und $(\breve{T}_k^y)_*$ aus den beiden Richtungsanteilen von (5.24) durch diagonales „Plätten“ der Massensterne $(M_k^x)_*$ und $(M_k^y)_*$:

$$\begin{array}{rcl} (M_k^y)_*^t \cdot (T_k^x)_* & \longrightarrow & (\breve{T}_k^x)_* := m_y \cdot (T_k^x)_* \quad \text{mit} \quad m_y := m_n + m_{z,y} + m_s, \\ (T_k^y)_*^t \cdot (M_k^x)_* & \longrightarrow & (\breve{T}_k^y)_* := m_x \cdot (T_k^y)_* \quad \text{mit} \quad m_x := m_w + m_{z,x} + m_o. \end{array}$$

Aufgrund des Verschwindens der Sternsummen von $(T_k^x)_*$ und $(T_k^y)_*$ läßt sich dies auch als ein spalten- und zeilenweises Aufsummieren der Einträge des 9-Punkte Sterns $(T_k)_*$ interpretieren:

$$\left[\begin{array}{ccc} t_{nw} & t_n & t_{no} \\ t_w & t_z & t_o \\ t_{sw} & t_s & t_{so} \end{array}\right] \longrightarrow \left[\begin{array}{r|r|r} 0 & 0 & 0 \\ \hline t_{nw} & t_n & t_{no} \\ +\,t_w & +\,t_z & +\,t_o \\ +\,t_{sw} & +\,t_s & +\,t_{so} \\ \hline 0 & 0 & 0 \end{array}\right],$$

$$\left[\begin{array}{ccc} t_{nw} & t_n & t_{no} \\ t_w & t_z & t_o \\ t_{sw} & t_s & t_{so} \end{array}\right] \longrightarrow \left[\begin{array}{c|ccccc|c} 0 & t_{nw} & + & t_n & + & t_{no} & 0 \\ \hline 0 & t_w & + & t_z & + & t_o & 0 \\ \hline 0 & t_{sw} & + & t_s & + & t_{so} & 0 \end{array}\right].$$

Die durch Plättung gewonnenen eindimensionalen Sterne $(\breve{T}_k^x)_*$ und $(\breve{T}_k^y)_*$ sind lediglich skalierte Versionen von $(T_k^x)_*$ und $(T_k^y)_*$. Die mit ihrer Hilfe definierten Gewichte $p_W^m := -\breve{t}_o/\breve{t}_{z,x}$, $p_O^m := -\breve{t}_w/\breve{t}_{z,x}$, $p_N^m := -\breve{t}_s/\breve{t}_{z,y}$ und $p_S^m := -\breve{t}_n/\breve{t}_{z,y}$ in den Punkten aus $\Omega_k^{(1,0)}$ und $\Omega_k^{(0,1)}$ sind daher identisch mit den oben erklärten operatorabhängigen Werten, da sich die Skalierungsfaktoren bei der Quotientenbildung herauskürzen.
Zur Bestimmung der Gewichte p_{NW}^m, p_{NO}^m, p_{SW}^m und p_{SO}^m für die verbleibenden Punkte aus

$\Omega_k^{(1,1)}$ addiert man die geplätteten Sterne und gelangt so zu dem 5-Punkte Stern

$$(\breve{T}_k)_* := (\breve{T}_k^x)_* + (\breve{T}_k^y)_*^t = \begin{bmatrix} & \breve{t}_n & \\ \breve{t}_w & \breve{t}_{z,x} + \breve{t}_{z,y} & \breve{t}_o \\ & \breve{t}_s & \end{bmatrix}. \tag{5.25}$$

Man löst nun in den Punkten aus $\Omega_k^{(1,1)}$ die durch $\breve{T}_k$ gegebenen diskreten homogenen Probleme zusammen mit den bereits bestimmten Gewichten als den diskreten Randwerten. Wir erhalten beispielsweise für p_{SW}^m die Bestimmungsgleichung

$$\breve{t}_z p_{SW}^m + \breve{t}_n p_W^m + \breve{t}_o p_S^m = 0.$$

Durch Umformen und Einsetzen der bekannten Gewichte $p_W^m = -\breve{t}_o/\breve{t}_{z,x}$ und $p_S^m = -\breve{t}_n/\breve{t}_{z,y}$ ergibt sich

$$p_{SW}^m = \frac{1}{\breve{t}_z}\left[\breve{t}_n \frac{\breve{t}_o}{\breve{t}_{z,x}} + \breve{t}_o \frac{\breve{t}_n}{\breve{t}_{z,y}}\right] = \frac{\breve{t}_o}{\breve{t}_{z,x}} \frac{\breve{t}_n}{\breve{t}_{z,y}} = p_S^m p_W^m.$$

Analog erhält man $p_{SO}^m = p_S^m p_O^m$, $p_{NW}^m = p_N^m p_W^m$ und $p_{NO}^m = p_N^m p_O^m$. Da sich somit die Eckgewichte der matrixabhängig bestimmten Prolongation ebenfalls als Produkte der bereits berechneten Gewichte schreiben lassen, ist wegen (5.23) die matrixabhängige Prolongation gleich der operatorabhängigen Prolongation. Die Wahl des Testraums $\mathcal{S}_k$ beeinflußt unter den gegebenen Voraussetzungen lediglich die Einträge der eindimensionalen Massensterne $(M_k^x)_*$ und $(M_k^y)_*$ und verändert damit höchstens die Skalierung der Sterne $(\breve{T}_k^x)_*$ und $(\breve{T}_k^y)_*$. Daher ist die matrixabhängige Prolongation unabhängig von $\mathcal{S}_k$.

Wir betrachten nun den Fall variabler separabler Konvektion $\vec{b} = (b_1(x), b_2(y))$. Auch hier ist eine additive Zerlegung des Konvektions-Diffusions Operators T wie in (5.18) immer noch möglich. Exaktes Lösen der eindimensionalen Gleichungen für beide Richtungsanteile T_x und T_y entlang der Elementkanten hinsichtlich des groben Gitters Ω_{k+1} mit den Randwerten 0 und 1 liefert verbogene eindimensionale L-Splines. Diese werden zusammen mit homogenen Nullrandwerten entlang der äußeren Kanten der Grobgitterelemente als Randwerte zur Lösung der homogenen zweidimensionalen Gleichung auf diesen Elementen gesetzt. Aufgrund der Tatsache, daß T separabel ist, ergeben sich deren Lösungen als die Produkte der verbogenen L-Spline Anteile, die als Randwerte dienten. Setzt man die so erhaltenen Lösungen bezüglich der vier an einen Grobgitterpunkt (x_i^{k+1}, y_j^{k+1}) angrenzenden Grobgitterelemente stückweise zusammen, so erhält man lokale zweidimensionale Basisfunktionen relativ zu Ω_{k+1} und die operatorabhängigen Prolongationsgewichte können wieder als deren Funktionswerte in den umliegenden Feingitterpunkten von Ω_k interpretiert werden. In diesem Fall sind die Prolongationssterne $(P_{k+1,\mathcal{V}}^k)_*$ und Operatorsterne $(T_k)_*$ nicht nur skalen- sondern zudem auch ortsabhängig. Man erhält die in Punkten (x_i^{k+1}, y_j^{k+1}) des groben Gitters aufsitzenden lokalen Sterne

$$(P_{k+1,(i,j),\mathcal{V}}^k)_* = \begin{bmatrix} p_{NW}((x_i^{k+1}, y_j^{k+1})) & p_N((x_i^{k+1}, y_j^{k+1})) & p_{NO}((x_i^{k+1}, y_j^{k+1})) \\ p_W((x_i^{k+1}, y_j^{k+1})) & p_Z((x_i^{k+1}, y_j^{k+1})) & p_O((x_i^{k+1}, y_j^{k+1})) \\ p_{SW}((x_i^{k+1}, y_j^{k+1})) & p_S((x_i^{k+1}, y_j^{k+1})) & p_{SO}((x_i^{k+1}, y_j^{k+1})) \end{bmatrix}. \tag{5.26}$$

Sie beschreiben, wie beim Wechsel zur nächstfeineren Skala in (x_i^{k+1}, y_j^{k+1}) befindliche grobskalige Information an die umgebenden neun Feingitterpunkte weiterzugeben ist. Die lokalen Grobgitter-Basisfunktionen können als Tensorprodukte der stückweise zusammengesetzten

verbogenen eindimensionalen L-Splines entlang der Elementkanten geschrieben werden. Daher haben die Eckeinträge der lokalen operatorabhängigen Prolongation (5.26) wieder Produktstruktur. Es gelten damit dann auch verallgemeinerte lokale Skalierungsgleichungen.

Satz 15 (Lokale Skalierungsgleichung im zweidimensionalen separablen Fall) *Die operatorabhängigen L-Splines zur zweidimensionalen Konvektions-Diffusions Gleichung mit separabler Konvektion $\vec{b} = (b_1(x), b_2(y))$ erfüllen die verallgemeinerten lokalen Skalierungsgleichungen*

$$\begin{aligned}\varphi^{k+1}_{(i,j),\mathcal{V}}(x,y) \;=\; & \Big[p_{NW}\varphi^{k}_{(2i-1,2j+1),\mathcal{V}} + p_{N}\varphi^{k}_{(2i,2j+1),\mathcal{V}} + p_{NO}\varphi^{k}_{(2i-1,2j+1),\mathcal{V}} \\ & + p_{W}\varphi^{k}_{(2i-1,2j),\mathcal{V}} + p_{Z}\varphi^{k}_{(2i,2j),\mathcal{V}} + p_{O}\varphi^{k}_{(2i-1,2j),\mathcal{V}} \\ & + p_{SW}\varphi^{k}_{(2i-1,2j+1),\mathcal{V}} + p_{S}\varphi^{k}_{(2i,2j+1),\mathcal{V}} + p_{SO}\varphi^{k}_{(2i-1,2j+1),\mathcal{V}}\Big](x,y).\end{aligned} \tag{5.27}$$

Bei den Gewichten handelt es sich um die Einträge des in (x_i^{k+1}, y_j^{k+1}) aufsitzenden lokalen Prolongationsmoleküls $(P^k_{k+1,(i,j),\mathcal{V}})_*$.

Beweis:
Die Behauptung ergibt sich sofort aus der Tensorprodukt-Darstellung von $\varphi^{k+1}_{(i,j),\mathcal{V}}$ und der verallgemeinerten Skalierungsgleichung (5.9) im eindimensionalen Fall.

□

Wie bereits festgestellt wurde, ist für variable Konvektionsfunktionen die exakte Integration der kontinuierlichen homogenen eindimensionalen Gleichungen zur Bestimmung der nichtverschwindenden Randwerte entlang der Elementkanten im allgemeinen nicht möglich. Ausgehend von einer Feingitterdiskretisierung T_k wählt man in der Praxis daher stets die soeben eingeführte matrixabhängige Prolongation. Hat T_k eine Tensorprodukt-Struktur ähnlich wie (5.24), was beispielsweise auf die einfache 5-Punkte Upwind-Diskretisierung (2.14) aus Kapitel 2 zutrifft, so liefert die matrixabhängige Prolongation auch in diesem Fall die gleichen Gewichte wie die operatorabhängige Prolongation. Dabei müssen bei der Berechnung dann lediglich die lokalen Operatorsterne $(T_{k,(i,j)})_*$ beachtet werden. Der mittels der Prolongationsoperatoren $P^k_{k+1,\mathcal{V}}$ und $P^k_{k+1,\mathcal{S}}$ durch eine Petrov–Galerkin-Vergröberung entstehende Grobgitteroperator stimmt aufgrund der Skalierungsbeziehung für die L-Spline Ansatzräume mit dem direkt über den Petrov–Galerkin-Ansatz bezüglich der gröberen Skala diskretisierten Operator überein:

$$T_{k+1} = (P^k_{k+1,\mathcal{S}})^t T_k P^k_{k+1,\mathcal{V}}. \tag{5.28}$$

Aufgrund ihrer besonderen Interpolationseigenschaften ist die operatorabhängige Prolongation $P^k_{k+1,\mathcal{V}}$ also mit dem gewählten Diskretisierungsverfahren kompatibel.

Zur Bestimmung einer matrixabhängigen Prologation im allgemeinen Fall nichtseparabler Konvektion $\vec{b} = (b_1(x,y), b_2(x,y))$ verwendet man anstelle der geplätteten die ursprünglichen Operatorsterne. Die damit berechneten Eckgewichte von $(P^k_{k+1,(i,j),\mathcal{V}})_*$ tragen dann ebenfalls zur lokalen Gauß-Elimination bei. Eine verallgemeinerte lokale Skalierungsgleichung wie (5.27) ist dann aber nicht mehr erfüllt und man gelangt insbesondere zu nichtkonformen Grobgitterräumen. Darüberhinaus existieren verfeinerte Techniken von de Zeeuw, die zur Definition matrixabhängiger Prolongationen die Aufspaltung der lokalen Operatorsterne in ihren symmetrischen und schiefsymmetrischen Anteil verwenden. Für Einzelheiten ihrer Konstruktion verweisen wir auf die Arbeiten [132, 133].

5.2.3 Zur Wahl geeigneter Basen der Komplementräume

Wir untersuchen im folgenden die Frage nach der Wahl geeigneter Basen für die Komplementräume $\mathcal{W}_{k+1}$ und $\mathcal{F}_{k+1}$, so daß $\mathcal{V}_k = \mathcal{W}_{k+1} \oplus \mathcal{V}_{k+1}$ und $\mathcal{S}_k = \mathcal{F}_{k+1} \oplus \mathcal{S}_{k+1}$ gelten. Dies ist wiederum gleichbedeutend mit der Auswahl geeigneter Matrizen $Q^k_{k+1,\mathcal{V}/\mathcal{S}}$, deren Spalten angeben, wie sich Punkten $(x^k_i, y^k_j) \in \Omega^f_k$ zugeordnete Basisfunktionen der jeweiligen Komplementräume mittels der Basisfunktionen der zugehörigen Feingitterräume darstellen lassen. Matrixabhängige Prolongationen $P^k_{k+1,\mathcal{V}/\mathcal{S}}$ für die Ansatz- und Testseite können zusammen mit einer hierarchischen Wahl der Prolongationen $Q^k_{k+1,\mathcal{V}/\mathcal{S}}$ für die Komplementräume als Transformationen interpretiert werden, die eine approximative Block-Gauß-Elimination des einfach zerlegten Operators $\tilde{T}_k$ bewirken. Wir illustrieren diesen Zusammenhang, indem wir entsprechend dem eindimensionalen Fall nachweisen, daß *ideale matrixabhängige Prolongationen* für die Ansatz- und Testseite zusammen mit einer hierarchischen Wahl der Komplementräume wieder zu einer Block-Diagonalisierung der Zweiskalen-transformierten Steifigkeitsmatrix führen. Die ideale matrixabhängige Prolongation für die Ansatzseite erhält man durch exaktes Lösen homogener Probleme, die ideale matrixabhängige Prolongation für die Testseite durch exaktes Lösen entsprechender dualer Probleme. Dies führt aber im allgemeinen zu vollbesetzten Matrizen $P^k_{k+1,\mathcal{V}/\mathcal{S}}$ und damit zu nicht mehr effizient durchführbaren Verfahren.

Verwendet man statt der idealen die im letzten Abschnitt berechneten matrixabhängigen Prolongationen, so ist die Block-Gauß-Elimination nur noch unvollständig. Es ergibt sich außerdem die folgende Schwierigkeit. In höheren Raumdimensionen besitzen die über die hierarchische Wahl gebildeten Zerlegungen nicht mehr die gleichen optimalen Stabilitätseigenschaften wie im eindimensionalen Fall. Betrachtet man anstelle der hierarchischen Zerlegungen Wavelet-artige Zerlegungen, so zeigen die darüber gewonnenen Multiskalen-Zyklen optimale Konvergenzeigenschaften [121]. Dies liegt insbesondere an dem stärkeren Einfluß der zugehörigen Multiskalen-Glätter. Wir beschäftigen uns deshalb mit der Konstruktion Wavelet-artiger Basisfunktionen für die Komplementräume mittels eines Tensorprodukt-Ansatzes ausgehend von den besprochenen eindimensionalen Wavelet-artigen Zerlegungen. Aufgrund unseres allgemeinen Petrov–Galerkin-Ansatzes ist es dann möglich, Wavelet-artige Multiskalen-Zerlegungen mit hierarchischen zu kombinieren, um dadurch in einem multiplikativen Verfahren eine approximative Elimination beispielsweise der Grobgitterunbekannten aus den transformierten Feingittergleichungen unter verbesserten Stabilitäts- und Glättungseigenschaften zu erhalten.

Hierarchische Wahl und ideale Prolongationen

Für die folgende algebraisch gehaltene Betrachtung ordnen wir die Gitterpunkte von Ω_k so, daß wir zunächst die Punkte aus $\Omega^f_k = \Omega_k \backslash \Omega_{k+1}$ und danach die aus $\Omega^c_k = \Omega_{k+1}$ aufzählen. Dies erreichen wir durch eine einfache Permutation und wir erhalten als permutierten Operator

$$T_k := \begin{pmatrix} A_{k+1} & B_{k+1} \\ C_{k+1} & T_{k+1} \end{pmatrix}. \tag{5.29}$$

Die permutierten Matrizen werden mit den gleichen Ausdrücken wie die ursprünglichen bezeichnet. Die Blockgestalt (5.29) wird hierbei also nur durch die Permutation der Unbekannten induziert und nicht durch einen zusätzlichen Basiswechsel. Die Prolongationsmatrizen

$P^k_{k+1,\mathcal{V}}$ und $P^k_{k+1,\mathcal{S}}$ haben bezüglich der neuen Anordnung die Blockstrukturen

$$P^k_{k+1,\mathcal{V}} = \begin{pmatrix} P^{k,f}_{k+1,\mathcal{V}} \\ P^{k,c}_{k+1,\mathcal{V}} \end{pmatrix} \quad \text{und} \quad P^k_{k+1,\mathcal{S}} = \begin{pmatrix} P^{k,f}_{k+1,\mathcal{S}} \\ P^{k,c}_{k+1,\mathcal{S}} \end{pmatrix}.$$

Es seien ferner $n_f = |\Omega^f_k|$ und $n_c = |\Omega^c_k|$ die Anzahlen der Fein-ohne-Grob- sowie Grobgitterpunkte. Der Übersichtlichkeit halber kennzeichnen wir hiermit im folgenden die Größen von Null- und Einheitsmatrizen. Der Wert der Prolongationen $P^k_{k+1,\mathcal{V}}$ und $P^k_{k+1,\mathcal{S}}$ in den Punkten aus $\Omega^{(0,0)}_k$ betrage 1, so daß also $P^{k,c}_{k+1,\mathcal{V}} = \mathbf{1}_{n_c \times n_c}$ und $P^{k,c}_{k+1,\mathcal{S}} = \mathbf{1}_{n_c \times n_c}$ sind. Wir erhalten gemäß einem hierarchischen Ansatz für die Komplementräume

$$Q^k_{k+1,\mathcal{V}} = \begin{pmatrix} \mathbf{1}_{n_f \times n_f} \\ \mathbf{0}_{n_c \times n_f} \end{pmatrix} \quad \text{und} \quad Q^k_{k+1,\mathcal{S}} = \begin{pmatrix} \mathbf{1}_{n_f \times n_f} \\ \mathbf{0}_{n_c \times n_f} \end{pmatrix}.$$

Die Forderung, daß das Ergebnis der Prolongation $P^k_{k+1,\mathcal{V}}$ in den Feingitterpunkten durch die Lösung des homogenen lokalen Problems bestimmt ist, formulierten wir diskret in (5.7) als

$$(T_k P^k_{k+1,\mathcal{V}} u_{k+1})(x^k_i) = 0 \quad \text{für } x^k_i \in \Omega^f_k.$$

Die duale Forderung (5.11) zur Bestimmung der Prolongation $P^k_{k+1,\mathcal{S}}$ lautete

$$((P^k_{k+1,\mathcal{S}})^t T_k u_k)(x^k_i) = 0 \quad \text{für } x^k_i \in \Omega^f_k.$$

Hierbei sind $u_k \in \mathcal{V}_k$ und $u_{k+1} \in \mathcal{V}_{k+1}$ beliebige Funktionen. Unter Berücksichtigung der neuen Anordnung kann man diese beiden Bedingungen äquivalent in Matrixform schreiben:

$$A_{k+1} P^{k,f}_{k+1,\mathcal{V}} + B_{k+1} = \mathbf{0}_{n_f \times n_c}, \quad \text{d.h.} \quad P^{k,f}_{k+1,\mathcal{V}} = -A^{-1}_{k+1} B_{k+1}, \tag{5.30}$$

$$(P^{k,f}_{k+1,\mathcal{S}})^t A_{k+1} + C_{k+1} = \mathbf{0}_{n_c \times n_f}, \quad \text{d.h.} \quad P^{k,f}_{k+1,\mathcal{S}} = -A^{-t}_{k+1} C^t_{k+1}. \tag{5.31}$$

Definieren wir mit Hilfe der so bestimmten *idealen Prolongationen* für den Ansatz- und den Testraum die Zweiskalen-Transformationen

$$W^k_{k+1,\mathcal{V}} = \begin{pmatrix} \mathbf{1}_{n_f \times n_f} & -A^{-1}_{k+1} B_{k+1} \\ \mathbf{0}_{n_c \times n_f} & \mathbf{1}_{n_c \times n_c} \end{pmatrix} \quad \text{und} \quad W^k_{k+1,\mathcal{S}} = \begin{pmatrix} \mathbf{1}_{n_f \times n_f} & -A^{-t}_{k+1} C^t_{k+1} \\ \mathbf{0}_{n_c \times n_f} & \mathbf{1}_{n_c \times n_c} \end{pmatrix}, \tag{5.32}$$

so erhalten wir

$$\begin{aligned} \tilde{T}_k &= (W^k_{k+1,\mathcal{S}})^t T_k W^k_{k+1,\mathcal{V}} \\ &= \begin{pmatrix} \mathbf{1}_{n_f \times n_f} & \mathbf{0}_{n_f \times n_c} \\ -C_{k+1} A^{-1}_{k+1} & \mathbf{1}_{n_c \times n_c} \end{pmatrix} \begin{pmatrix} A_{k+1} & \mathbf{0}_{n_f \times n_c} \\ C_{k+1} & T_{k+1} - C_{k+1} A^{-1}_{k+1} B_{k+1} \end{pmatrix} \\ &= \begin{pmatrix} A_{k+1} & \mathbf{0}_{n_f \times n_c} \\ \mathbf{0}_{n_c \times n_f} & T_{k+1} - C_{k+1} A^{-1}_{k+1} B_{k+1} \end{pmatrix}. \end{aligned}$$

Der Zweiskalen-transformierte Operator $\tilde{T}_k$ ist also wie schon im eindimensionalen Fall blockdiagonal. Für eindimensionale Probleme stimmen die mittels (5.30) und (5.31) algebraisch definierten idealen Prolongationen mit den in Kapitel 5.1 erklärten operator- beziehungsweise matrixabhängigen Prolongationen überein.

Im zweidimensionalen Fall sind Fein-ohne-Grobgitterpunkte anders als im Eindimensionalen nicht nur von Punkten des gröberen Gitters sondern auch von Fein-ohne-Grobgitterpunkten umgeben. Beispielsweise haben für ein Gitter Ω_k Punkte aus der Menge $\Omega_k^{(1,1)}$ auch Gitterpunkte aus $\Omega_k^{(0,1)}$ und $\Omega_k^{(1,0)}$ als direkte Nachbarn. Diese topologische Beziehung führt ausgehend von einem Operator T_k mit 5-Punkte oder 9-Punkte Sternmuster dazu, daß der Block A_{k+1} des permutierten Operators nicht mehr diagonal und damit dessen Inverse A_{k+1}^{-1} nicht lokal ist. Der mit Hilfe idealer Prolongationen $P_{k+1,\mathcal{V}}^k$ und $P_{k+1,\mathcal{S}}^k$ für den Ansatz- und Testraum berechnete Grobgitteroperator $T_{k+1} - C_{k+1}A_{k+1}^{-1}B_{k+1}$ ist dann im allgemeinen vollbesetzt und daher ein rekursives Fortführen der idealen Vergröberung auf effiziente Weise nicht möglich. Approximationen $\breve{P}_{k+1,\mathcal{V}}^k$ und $\breve{P}_{k+1,\mathcal{S}}^k$ an die idealen Prolongationen können nach dem folgenden allgemeinen Prinzip gefunden werden. Man ersetze die Blöcke A_{k+1} und B_{k+1} sowie A_{k+1}^t und C_{k+1}^t durch $\breve{A}_{k+1}$ und $\breve{B}_{k+1}$ respektive $\breve{A}_{k+1}^t$ und $\breve{C}_{k+1}^t$, so daß die beiden Forderungen

- $\breve{A}_{k+1}$ und $\breve{A}_{k+1}^t$ sind einfach zu invertieren und
- $\breve{A}_{k+1}^{-1}\breve{B}_{k+1}$ und $\breve{A}_{k+1}^{-t}\breve{C}_{k+1}^t$ sind lokale Approximationen an $A_{k+1}^{-1}B_{k+1}$ sowie $A_{k+1}^{-t}C_{k+1}^t$

erfüllt sind. Man definiert damit

$$\breve{P}_{k+1,\mathcal{V}}^k := \begin{pmatrix} -\breve{A}_{k+1}^{-1}\breve{B}_{k+1} \\ \mathbf{1}_{n_c \times n_c} \end{pmatrix} \quad \text{und} \quad \breve{P}_{k+1,\mathcal{S}}^k := \begin{pmatrix} -\breve{A}_{k+1}^{-t}\breve{C}_{k+1}^t \\ \mathbf{1}_{n_c \times n_c} \end{pmatrix} \tag{5.33}$$

als Approximationen der idealen Prolongationen für den Ansatz- und Testraum. Sie können als ideale Prolongationen zu den modifizierten Diskretisierungen des ursprünglichen und dualen Problems

$$\begin{pmatrix} \breve{A}_{k+1} & \breve{B}_{k+1} \\ C_{k+1} & T_{k+1} \end{pmatrix} \quad \text{und} \quad \begin{pmatrix} \breve{A}_{k+1}^t & \breve{C}_{k+1}^t \\ B_{k+1}^t & T_{k+1}^t \end{pmatrix}$$

angesehen werden.

Wir zeigen abschließend noch, wie die im vorherigen Abschnitt konstruierte matrixabhängige Prolongation für den Ansatzraum hierüber verstanden werden kann. Ausgehend von der Vierfarbenunterteilung von Ω_k besitzen die Blöcke A_{k+1}, B_{k+1} und C_{k+1} der permutierten Steifigkeitsmatrix (5.29) eine innere Blockstruktur, so daß

$$T_k = \begin{pmatrix} A_{k+1}^{11} & A_{k+1}^{12} & A_{k+1}^{13} & B_{k+1}^1 \\ A_{k+1}^{21} & A_{k+1}^{22} & A_{k+1}^{23} & B_{k+1}^2 \\ A_{k+1}^{31} & A_{k+1}^{32} & A_{k+1}^{33} & B_{k+1}^3 \\ C_{k+1}^1 & C_{k+1}^2 & C_{k+1}^3 & T_{k+1} \end{pmatrix}$$

gilt. Die erste Blockzeile beschreibt die Kopplungen der Freiheitsgrade aus $\Omega_k^{(1,1)}$, die zweite diejenigen aus $\Omega_k^{(0,1)}$, die dritte die aus $\Omega_k^{(1,0)}$ und schließlich die vierte die Kopplungen der Grobgitterfreiheitsgrade. Der durch Plätten der Massensterne enstehende Operator $\breve{T}_k$ mit dem 5-Punkte Sternmuster (5.25) hat dann die Gestalt

$$\breve{T}_k = \begin{pmatrix} \breve{A}_{k+1}^{11} & \breve{A}_{k+1}^{12} & \breve{A}_{k+1}^{13} & \mathbf{0} \\ \breve{A}_{k+1}^{21} & \breve{A}_{k+1}^{22} & \mathbf{0} & \breve{B}_{k+1}^2 \\ \breve{A}_{k+1}^{31} & \mathbf{0} & \breve{A}_{k+1}^{33} & \breve{B}_{k+1}^3 \\ \mathbf{0} & \breve{C}_{k+1}^2 & \breve{C}_{k+1}^3 & \breve{T}_{k+1} \end{pmatrix}.$$

Hierbei sind die durch **0** gekennzeichneten Blöcke Nullblöcke passender Dimension und die Hauptdiagonalblöcke $\breve{A}^{ii}_{k+1}$ $(i = 1, 2, 3)$ jeweils diagonal. Die Bestimmung der Gewichte p^m_W, p^m_O, p^m_N und p^m_S in den Punkten aus $\Omega^{(1,0)}_k$ und $\Omega^{(0,1)}_k$ mittels der eindimensionalen Richtungsanteile $(\breve{T}^x_k)_*$ und $(\breve{T}^y_k)_*$ kann als das Vernachlässigen der Kopplungen $\breve{A}^{21}_{k+1}$ und $\breve{A}^{31}_{k+1}$ verstanden werden, so daß sich die im vorherigen Abschnitt konstruierte matrixabhängige Prolongation für den Ansatzraum mittels (5.33) über die Approximationen

$$\breve{A}_{k+1} = \begin{pmatrix} \breve{A}^{11}_{k+1} & \breve{A}^{12}_{k+1} & \breve{A}^{13}_{k+1} \\ \mathbf{0} & \breve{A}^{22}_{k+1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \breve{A}^{33}_{k+1} \end{pmatrix} \quad \text{und} \quad \breve{B}_{k+1} = \begin{pmatrix} \mathbf{0} \\ \breve{B}^{2}_{k+1} \\ \breve{B}^{3}_{k+1} \end{pmatrix}$$

der Blöcke A_{k+1} und B_{k+1} ergibt. Es hat $\breve{A}_{k+1}$ insbesondere obere Dreiecksgestalt und ist damit einfach zu invertieren.

Wavelet-artige Wahl

Wir beschreiben jetzt für den Fall einer Wavelet-artigen Wahl der Komplementraum-Basen auf der Testseite den allgemeinen Aufbau der zugehörigen Transformationsmatrix $Q^k_{k+1,\mathcal{V}}$. Die Konstruktion der entsprechenden Transformation $Q^k_{k+1,\mathcal{S}}$ für die Ansatzseite geschieht vollkommen analog. Anhand der eindimensionalen Finite-Elemente Räumen $\mathcal{V}^{x,j}_k$ und $\mathcal{V}^{y,i}_k$ zur Skala k entlang der Gitterlinien $y = y^k_j = const$ und $x = x^k_i = const$ erhalten wir durch Tensorprodukt-Bildung der abstrakten eindimensionalen Räume $\mathcal{V}^x_{k+1}$ und $\mathcal{V}^y_{k+1}$ den zweidimensionalen Finite-Elemente Raum

$$\mathcal{V}_k = \mathcal{V}^x_k \otimes \mathcal{V}^y_k \tag{5.34}$$

mit der natürlichen Tensorprodukt-Basis

$$\{\varphi^k_{(i,j),\mathcal{V}} \,:\, \varphi^k_{(i,j),\mathcal{V}}(x,y) = \varphi^{k,x}_{(i,j),\mathcal{V}}(x)\varphi^{k,y}_{(i,j),\mathcal{V}}(y) \quad \text{für } 0 < i,j < 1/h_k\}.$$

Wir kennzeichnen im folgenden eindimensionale Basisfunktionen mit Hilfe von Indizes zugehöriger Fein-ohne-Grobgitterpunkte. Wir betrachten die Zerlegungen

$$\mathcal{V}^{x,j}_k = \mathcal{W}^{x,j}_{k+1} \oplus \mathcal{V}^{x,j}_{k+1}, \tag{5.35}$$
$$\mathcal{V}^{y,i}_k = \mathcal{W}^{y,i}_{k+1} \oplus \mathcal{V}^{y,i}_{k+1}, \tag{5.36}$$

mit den Basisdarstellungen

$$\mathcal{W}^{x,j}_{k+1} := \left\langle \psi^{k+1,x}_{(i,j),\mathcal{V}} : 0 < i < 1/h_k, i \text{ ungerade} \right\rangle,\ \mathcal{V}^{x,j}_{k+1} := \left\langle \varphi^{k+1,x}_{(i,j),\mathcal{V}} : 0 < i < 1/h_k, i \text{ gerade} \right\rangle,$$
$$\mathcal{W}^{y,i}_{k+1} := \left\langle \psi^{k+1,y}_{(i,j),\mathcal{V}} : 0 < j < 1/h_k, j \text{ ungerade} \right\rangle,\ \mathcal{V}^{y,i}_{k+1} := \left\langle \varphi^{k+1,y}_{(i,j),\mathcal{V}} : 0 < j < 1/h_k, j \text{ gerade} \right\rangle.$$

Dabei bezeichnen $\psi^{k+1,x}_{(i,j),\mathcal{V}}$ und $\psi^{k+1,y}_{(i,j),\mathcal{V}}$ Basisfunktionen der entsprechenden Komplementräume. Die Zerlegungen (5.35) und (5.36) führen zu abstrakten eindimensionalen Zerlegungen

$$\mathcal{V}^x_k = \mathcal{W}^x_{k+1} \oplus \mathcal{V}^x_{k+1},$$
$$\mathcal{V}^y_k = \mathcal{W}^y_{k+1} \oplus \mathcal{V}^y_{k+1}.$$

Durch Einsetzen dieser in die Tensorprodukt-Darstellung (5.34) gelangen wir zu der folgenden Zerlegung des feinskaligen Ansatzraums $\mathcal{V}_k$:

$$\begin{aligned}\mathcal{V}_k &= (\mathcal{W}_{k+1}^y \oplus \mathcal{V}_{k+1}^y) \otimes (\mathcal{W}_{k+1}^x \oplus \mathcal{V}_{k+1}^x) \\ &= (\mathcal{W}_{k+1}^y \otimes \mathcal{W}_{k+1}^x) \oplus (\mathcal{W}_{k+1}^y \otimes \mathcal{V}_{k+1}^x) \oplus (\mathcal{V}_{k+1}^y \otimes \mathcal{W}_{k+1}^x) \oplus (\mathcal{V}_{k+1}^y \otimes \mathcal{V}_{k+1}^x) \\ &=: \quad \mathcal{W}_{k+1}^{(1,1)} \quad \oplus \quad \mathcal{W}_{k+1}^{(0,1)} \quad \oplus \quad \mathcal{W}_{k+1}^{(1,0)} \quad \oplus \quad \mathcal{V}_{k+1}.\end{aligned}$$

Wir erhalten dabei als natürliche Basisdarstellungen der drei unterschiedenen Anteile des Komplementraums $\mathcal{W}_{k+1} = \mathcal{W}_{k+1}^{(1,1)} \oplus \mathcal{W}_{k+1}^{(0,1)} \oplus \mathcal{W}_{k+1}^{(1,0)}$

$$\begin{aligned}\mathcal{W}_{k+1}^{(1,1)} &= \Big\langle \psi_{(i,j),\mathcal{V}}^{k+1,(1,1)} \,:\, \psi_{(i,j),\mathcal{V}}^{k+1,(1,1)}(x,y) := \psi_{(i,j),\mathcal{V}}^{k+1,x}(x)\psi_{(i,j),\mathcal{V}}^{k+1,y}(y) \\ &\qquad \text{für } 0 < i,j < 1/h_k, i \text{ ungerade, } j \text{ ungerade} \Big\rangle, \\ \mathcal{W}_{k+1}^{(0,1)} &= \Big\langle \psi_{(i,j),\mathcal{V}}^{k+1,(0,1)} \,:\, \psi_{(i,j),\mathcal{V}}^{k+1,(0,1)}(x,y) := \psi_{(i,j),\mathcal{V}}^{k+1,x}(x)\varphi_{(i,j),\mathcal{V}}^{k+1,y}(y) \\ &\qquad \text{für } 0 < i,j < 1/h_k, i \text{ gerade, } j \text{ ungerade} \Big\rangle, \\ \mathcal{W}_{k+1}^{(1,0)} &= \Big\langle \psi_{(i,j),\mathcal{V}}^{k+1,(1,0)} \,:\, \psi_{(i,j),\mathcal{V}}^{k+1,(1,0)}(x,y) := \varphi_{(i,j),\mathcal{V}}^{k+1,x}(x)\psi_{(i,j),\mathcal{V}}^{k+1,y}(y) \\ &\qquad \text{für } 0 < i,j < 1/h_k, i \text{ ungerade, } j \text{ gerade} \Big\rangle.\end{aligned}$$

Die Basisfunktionen der Detail-Unterräume können auch über lokale zweidimensionale Sterne charakterisiert werden, die sich aus den Produkten der berechneten eindimensionalen lokalen Skalierungs- beziehungsweise Waveletfilter ergeben. Wir gehen davon aus, daß es möglich ist, für jeden inneren Gitterpunkt in jeder Gitterrichtung entweder eine *Grobfiltermaske* der maximalen Länge 3 oder eine *Detailfiltermaske* der maximalen Länge 5 anzugeben. Für dem Rand nächstgelegene innere Punkte hat die Detailfiltermaske in der entsprechenden Richtung jedoch die maximale Länge 3. Wir erhalten als lokale zweidimensionale Sterne in

$$\begin{aligned}(x_i^k, y_j^k) \in \Omega_k^{(1,1)} &: (Q_{k+1,(i,j),\mathcal{V}}^{k,y})_*^t \cdot (Q_{k+1,(i,j),\mathcal{V}}^{k,x})_*, \\ (x_i^k, y_j^k) \in \Omega_k^{(0,1)} &: (Q_{k+1,(i,j),\mathcal{V}}^{k,y})_*^t \cdot (P_{k+1,(i,j),\mathcal{V}}^{k,x})_*, \\ (x_i^k, y_j^k) \in \Omega_k^{(1,0)} &: (P_{k+1,(i,j),\mathcal{V}}^{k,y})_*^t \cdot (Q_{k+1,(i,j),\mathcal{V}}^{k,x})_*.\end{aligned}$$

Hierbei bezeichnen $Q_{k+1,(i,j),\mathcal{V}}^{k,y}$ und $P_{k+1,(i,j),\mathcal{V}}^{k,y}$ lokale eindimensionale Filter. Für die Punkte aus $\Omega_k^{(0,1)}$ wird dem Tensorprodukt-Ansatz entsprechend in y-Richtung ein Detailfilter gewählt, da sie (mit Ausnahme der randnahen Punkte) auf einer y-Gitterlinie von Ω_{k+1} genau zwischen zwei Grobgitterpunkten sitzen. Für die Punkte aus $\Omega_k^{(1,0)}$ wird aus dem gleichen Grund in x-Richtung ein Detailfilter gewählt. Die vollständige Transformationsmatrix, die den Komplementraum $\mathcal{W}_{k+1}$v charakterisiert, kann nun mit Hilfe der resultierenden Teilmatrizen $Q_{k+1,\mathcal{V}}^{k,(1,1)}$, $Q_{k+1,\mathcal{V}}^{k,(0,1)}$ und $Q_{k+1,\mathcal{V}}^{k,(1,0)}$ zu

$$Q_{k+1,\mathcal{V}}^k = [Q_{k+1,\mathcal{V}}^{k,(1,1)}, Q_{k+1,\mathcal{V}}^{k,(0,1)}, Q_{k+1,\mathcal{V}}^{k,(1,0)}]$$

zusammengesetzt werden.

Ausgehend von eindimensionalen hierarchischen Zerlegungen ergibt sich mit der soeben vorgestellten Tensorprodukt-Konstruktion *nicht* die im letzten Abschnitt diskutierte zweidimensionale hierarchische Zerlegung. Der Grund hierfür ist, daß am Aufbau der Transformationsmatrizen $Q^{k,(0,1)}_{k+1,\mathcal{V}}$ und $Q^{k,(1,0)}_{k+1,\mathcal{V}}$ noch die eindimensionalen Skalierungsfilter $(P^{k,x/y}_{k+1,(i,j),\mathcal{V}})_*$ beteiligt sind.

Will man über mehrere Skalen hinweg matrixabhängige Filter in Punkten aus $\Omega^{(0,1)}_k$ und $\Omega^{(1,0)}_k$ bestimmen, so verlangt dies — selbst im Fall homogener Dirichletscher Randbedingungen — eine besondere Berücksichtigung der Randgitterpunkte. Wir sprechen im folgenden der Einfachheit halber direkt von Kopplungen der Gitterpunkte anstatt von Kopplungen der darin aufsitzenden Basisfunktionen. Betrachten wir zum Beispiel einen dem westlichen Rand von $\bar{\Omega}_k$ nächstgelegenen Punkt aus $\Omega^{(1,0)}_k$. Dann sitzt gemäß der oben beschriebenen Vorgehensweise darin in x-Richtung ein Detail- und in y-Richtung ein Grobfilter auf. Zur korrekten Berechnung des (matrixabhängigen) Grobfilters benutzt man die Kopplungen des nächsten nördlichen und südlichen Feingitterfreiheitsgrades inklusive der Kopplungen durch die Randgitterpunkte. Ein „abgeschnittener“ Dirichlet-Stern liefert hier falsche Ergebnisse.
Selbst wenn auf dem feinsten Gitter die Kopplungen der randnahen Freiheitsgrade durch die Randgitterpunkte berücksichtigt werden, indem man die vollständigen Operatorsterne verwendet, müssen, damit dies auf den gröberen Skalen richtig fortgeführt wird, die jeweiligen Grobgitteroperatoren ebenfalls inklusive der Kopplungen durch die Randgrobgitterpunkte richtig aufgestellt werden.
Dies erreicht man, indem man von vornherein den Operator auch in den Randpunkten aufstellt und mittels (matrixabhängiger) Prolongationen in den Randgrobgitterpunkten, die Randkopplungen der randnahen Grobgitterpunkte berechnet. Man bestimmt ebenfalls die Kopplungen der Randgrobgitterpunkte durch die randnahen Grobgitterpunkte und erhält damit Grobgitteroperator-Sterne in den Randgrobgitterpunkten, so daß sich die Prozedur auf gröbere Skalen fortsetzen läßt.

5.3 Probleme in 2D: Algebraische Konstruktionen

5.3.1 Algebraische Multiskalen-Zerlegungen

Die im letzten Unterkapitel besprochenen problemangepaßten Tensorprodukt-artigen Multiskalen-Zerlegungen sind über eine fest vorgegebene geometrische Vergröberung der zugrundeliegenden Gitter entstanden (Vierfarbenzerlegung). Wir bezeichnen sie daher auch als *geometrische Multiskalen-Zerlegungen*. Für Randwertprobleme in kompliziert berandeten Gebieten ist ein in diesem Sinne geometrisch orientiertes Multiskalen-Konzept, zumindest was einfache ungestörte Probleme betrifft, zwar immer noch möglich, es führt jedoch schon in diesen einfachen Fällen zu recht komplizierten Konstruktionen [34, 46, 77, 110, 111]. Eine Frage, die in dem Zusammenhang ebenfalls auftritt, ist, ob sinnvolle Multiskalen-Zerlegungen auch dann noch definierbar sind, wenn zu dem zu lösenden linearen Gleichungssystem kein physikalisches Gitter mehr vorhanden ist.

Eine Mehrgitter-Variante, die mit Erfolg Multilevel-Löser für nur schwer oder tatsächlich unzugängliche Geometrien konstruiert, ist die Klasse der *algebraischen Mehrgitter-Verfahren* (AMG) [27, 52, 103, 112]. Im Unterschied zu geometrischen Mehrgitter-Verfahren, bei denen die Folge der Grobgitterpunkte vorgegeben ist, werden bei algebraischen Mehrgitter-Verfahren zunächst geeignete Grobgitterfreiheitsgrade ausgewählt, was mit Hilfe zweier re-

eller Parameter α und β über das Konzept *algebraisch glatter Fehler* und die Betrachtung *starker* und *schwacher Kopplungen* geschieht. Wir bemerken ausdrücklich, daß durch diese Auswahl die Fein-ohne-Grobgitterfreiheitsgrade automatisch festgelegt sind. Danach werden zugehörige Prolongationen und Restriktionen problemabhängig berechnet, was die sogenannte *Setup-Phase* abschließt. Sowohl für die Auswahl der entsprechenden Freiheitsgrade als auch für die Berechnung der Prolongationen und Restriktionen wird ausschließlich die Systemmatrix des zu lösenden Gleichungssystems benutzt. Die auf diese Weise algebraisch bestimmten Prolongationen und Restriktionen können dann innerhalb gewöhnlicher Mehrgitter-Löser eingesetzt werden.
Algebraische Mehrgitter-Verfahren haben sich insbesondere aber auch als sehr robuste Löser für Probleme mit kompliziert gearteten Koeffizientenfunktionen des Differentialoperators erwiesen, etwa bei der Behandlung von wirbelbehafteten Konvektions-Diffusions Problemen [52]. Wirbelbehaftete Strömungen und die daraus abgeleiteten Konvektions-Diffusions Gleichungen sind nach unserem Wissen eine große Herausforderung für alle zur Zeit existierenden Mehrgitter-Löser, da sie zu hochgradig nichtseparablen Problemen führen [76, 79, 94, 128]. Aufgrund der einschränkenden Mehrgitter-Interpretation unserer Multiskalen-Zyklen erscheint es aussichtslos, solch schwierige Probleme mit den bisherigen Tensorprodukt-Techniken erfolgreich angehen zu können. Sämtliche Versuche, etwa eine Kreisströmung auf der Basis geometrisch definierter Multiskalen-Zerlegungen und matrixabhängiger Multiskalen-Transformationen zu behandeln, führten zu nicht zufriedenstellenden Ergebnissen. Da das algebraische Mehrgitter-Konzept bewußt auf speziell definierte Glättungsprozeduren verzichtet, stellt sich die Frage, ob im Fall einer Kreisströmung mit Hilfe AMG-basierter Zerlegungen und Komponenten ein robuster Multiskalen-Löser gefunden werden kann, der auf direkten *algebraischen Multiskalen-Zerlegungen* beruht.

Eine naheliegende Unterraumzerlegung ist in diesem Zusammenhang durch eine AMG-basierte hierarchische Basis Zerlegung gegeben. Hierbei werden auf jedem Level nach Festlegen der Basen der nächstgrobskaligeren Räume, was mittels der AMG-Prolongationen geschieht, die zugehörigen Komplemente durch die bisherigen in den Fein-ohne-Grobgitterpunkten befindlichen Basisfunktionen aufgespannt. Für den gewöhnlichen Laplace-Operator ergeben sich mit AMG jedoch (approximativ) wieder geometrische Vergröberungen und bilineare Interpolationen. Deshalb ist nicht zu erwarten, daß man mittels AMG-basierter *hierarchischer* Zerlegungen einen zumindest optimalen Multiskalen-Löser erhält. Dies ist ein Nachteil, unter dem auch die Approximative Zyklische Reduktion Reuskens leidet, wie in Kapitel 4 bereits bemerkt wurde. Mit einem AMG-hierarchischen W-Zyklus würde man ähnlich wie mit AMLI-Verfahren, die höhere Polynomgrade verwenden (siehe Kapitel4), zwar eine Stabilisierung des Lösers hinsichtlich seiner Gitterweitenabhängigkeit erreichen, für komplizierte Probleme kann ein solcher W-Zyklus aber sehr aufwendig sein. Dies liegt daran, daß die Zahl der ausgewählten Grobgitterfreiheitsgrade aufgrund der algebraischen Auswahl nicht notwendig mit einer geometrischen Rate abfällt. Ein entsprechender W-Zyklus ist dann im allgemeinen nicht mehr mit einem Aufwand von $\mathcal{O}(N_0)$ implementierbar, wenn N_0 die Dimension des Gleichungssystems ist.

Man gelangt so zur Frage nach möglichen Wavelet-artigen Stabilisierungen AMG-basierter hierarchischer Basen. Die Konstruktion AMG-basierter Prewavelets über eine explizite Orthogonalisierung jeweils relativ zum nächstgrobskaligeren Raum führt — wenn überhaupt möglich — zu äußerst komplizierten und unüberschaubaren Bestimmungsgleichungen für die jeweiligen Filterkoeffizienten. Ihr Erfolg ist zudem fraglich, wie unsere Erfahrung mit den

explizit konstruierten L-Spline Prewavelets zeigt. Auch Lifting-Konstruktionen [113] leiden in unserem Fall unter ähnlichen Schwierigkeiten. Daher haben wir uns für eine wesentlich einfachere Möglichkeit entschieden, die sich auf die folgende elementare Beobachtung stützt.

Betrachtet man den trivialen Fall eindimensionaler Konvektions-Diffusions Probleme, so fällt auf, daß bei geometrischer Vergröberung und verschwindender Konvektion der 3-Punkte Operatorstern, also die zweite Ableitung mittels zentraler Differenzen, bis auf eine Skalierung mit dem Pre-Prewavelet Filter (5.15) übereinstimmt. Im umgekehrten Grenzfall verschwindender Diffusion ist der einfache Upwind-Stern mit dem berechneten L-Spline Prewavelet Filter (5.13) identisch.

Wir schlagen deshalb vor, AMG-basierte Detailfilter als Baupläne der Basisfunktionen der Komplementräume einfach über lokale Operatorsterne zu konstruieren, die sich in den entsprechenden Fein-ohne-Grobgitterpunkten befinden. Wir setzen für einen Punkt $(x_i^k, y_j^k) \in \Omega_k^f$

$$(Q^k_{k+1,(i,j),\mathcal{V}})_* := (T_{k,(i,j)})_*. \tag{5.38}$$

Die resultierenden Basisfunktionen der Komplementräume bezeichnen wir als *AMGlets*.

Eine ähnliche Idee wird in [73] zur matrixabhängigen Bestimmung von Prolongationsmatrizen $P^k_{k+1,\mathcal{V}}$ ausgehend von einer Feingitterdiskretisierung und einer Galerkin-Vergröberung verwendet. Es wird dabei $P^k_{k+1,\mathcal{V}}$ bis auf Skalierung durch $(P^k_{k+1,(i,j),\mathcal{V}})_* := |(T_{k,(i,j)})_*|$, das heißt über die Beträge der Einträge der Operatorsterne in den entsprechenden Grobgitterpunkten $(x_i^k, y_j^k) \in \Omega_k^c$ gebildet.

Im Verlauf des algebraischen Vergröberungsprozesses verbreitern sich bekanntlich zunehmend die über die Galerkin-Produkte $(P^k_{k+1,\mathcal{V}})^t T_k P^k_{k+1,\mathcal{V}}$ bestimmten Grobgitteroperator-Sterne. Die Träger der so konstruierten AMGlets vergrößern sich natürlich entsprechend. Man kann versuchen, dem Verbreitern der Grobgitteroperator-Sterne entgegenzuwirken, indem man mittels eines Paramaters γ_P die Zeilen der an dem Galerkin-Produkt beteiligten AMG-Prolongationen abschneidet und eine Reskalierung durchführt [112]. Ferner werden von uns danach relativ zum betragsmäßig größten nicht zentralen Eintrag der resultierenden Operatorsterne die kleinsten Sterneinträge mit Hilfe eines weiteren Parameters γ_Q abgeschnitten und zum zentralen Sterneintrag hinzuaddiert. Wir erhalten damit die Matrizen $Q^k_{k+1,\mathcal{V}}$ bis auf die Skalierung über die abgeschnittenen Sterne

$$(Q^k_{k+1,(i,j),\mathcal{V}})_* = [(T_{k,(i,j)})_*]_{\gamma_Q}. \tag{5.39}$$

Ein solches algebraisches Abschneiden entspricht genau der klassischen AMG-Philosophie und bewirkt anschaulich eine Konzentration und Lokalisierung der AMGlets.

Zusammenfassend erhalten wir die nachfolgende Prozedur zur Bestimmung der Transformationsmatrizen $P^k_{k+1,\mathcal{V}}$ und $Q^k_{k+1,\mathcal{V}}$ für eine Wavelet-artige algebraische Multiskalen-Zerlegung der Ansatzseite anhand von T_k und AMG(α,β):

1. Auswahl von Ω_k^c und Ω_k^f durch AMG(α,β);
2. Berechnen der AMG-basierten Prolongation $P^k_{k+1,\mathcal{V}}$;
3. Abschneiden von $P^k_{k+1,\mathcal{V}}$ mittels γ_P und Reskalierung;
4. Berechnen von $Q^k_{k+1,\mathcal{V}}$ mittels (5.39);

Die Zweiskalen-Transformationen bezüglich gröberer Skalen können dann mit Hilfe der über Galerkin-Produkte $T_{k+1} = (P^k_{k+1,\mathcal{V}})^t T_k P^k_{k+1,\mathcal{V}}$ bestimmten Grobgitteroperatoren weiter berechnet werden.

Kapitel 6

Numerische Beispiele

In unseren numerischen Beispielen betrachten wir die Konvektions-Diffusions Gleichung

$$Tu := -\triangle u + b_1 \partial_x u + b_2 \partial_y u = f \quad \text{auf} \quad \Omega =]0;1[^2$$

zusammen mit homogenen Dirichletschen Randbedingungen. Ausgehend von einer geeigneten Diskretisierung bezüglich unterschiedlich feiner uniformer Tensorproduktgitter $\Omega_0 \subset \Omega$ untersuchen wir numerisch das Konvergenzverhalten der vorgestellten multiplikativen Multiskalen-Verfahren. Es soll dabei für verschiedene Testprobleme die Auswirkung unterschiedlicher Kombinationen von Petrov–Galerkin-artigen Multiskalen-Zerlegungen der Feingitter Ansatz- und Testräume auf die Konvergenz der Verfahren untersucht werden. Wir studieren die folgenden vier Klassen von Konvektions-Diffusions Problemen, die aus Sicht der meisten derzeit bekannten iterativen Löser einen zunehmenden Grad an Schwierigkeit aufweisen:

$$\begin{aligned} T^\alpha &:= -\Delta + b\,\partial_x, & T^\beta &:= -\Delta + \frac{b}{\sqrt{2}}\left(\partial_x + \partial_y\right), \\ T^\gamma &:= -\Delta + b\cos\omega\,\partial_x + b\sin\omega\,\partial_y, & T^\delta &:= -\Delta + 4b\,x(x-1)(1-2y)\,\partial_x \\ & & & \quad -4b\,y(y-1)(1-2x)\,\partial_y. \end{aligned}$$

Hierbei ist b eine Konstante ($b = 0, 10^1, ..., 10^6, 10^8$) und der Winkel $\omega \in [0, \pi]$ beschreibt für das Beispiel T^γ die Richtung der Konvektion relativ zur x-Achse.

6.1 Beschreibung der Experimente

6.1.1 Testumgebung

Zur Diskretisierung der kontinuierlichen Operatoren verwenden wir durchweg die 5-Punkte Sterne (2.14), die durch Finite-Differenzen und einfaches Upwinding entstehen (Kapitel 2, Schema 4). Dies geschieht mit der Absicht, schon bezüglich des feinsten Gitters Ω_0 jeweils eine stabile Diskretisierung zu erhalten [62]. Wir setzen voraus, daß Ω_0 über eine Hierarchie von geschachtelten Gittern durch uniforme dyadische Verfeinerung entstanden ist. Hierauf interpretieren wir die Upwind-Diskretisierung als ein Schema, das näherungsweise mittels einer Petrov–Galerkin Finite-Elemente Methode mit operatorabhängigen Ansatz- und Testräumen $\mathcal{V}_0$ und $\mathcal{S}_0$ entsteht. Gemäß der Bemerkung zu Beginn von Kapitel 3 sprechen wir dann ebenfalls von Multiskalen-Zerlegungen zugehöriger Feingitter Ansatz- und Testräume. Um numerisch stabile Verfahren zu erhalten, skalieren wir die Operatoren der obigen Testbeispiel mit einem Faktor $\varepsilon = \frac{1}{b}$ für $b \neq 0$. Wird nicht ausdrücklich auf andere Zerlegungstiefen hingewiesen, so betrachten wir für die von uns verwendeten gewöhnlichen und verbesserten Multiskalen-Zyklen jeweils vollständige Zerlegungen der Räume $\mathcal{V}_0$ und $\mathcal{S}_0$. Die gröbsten

Gitter bestehen also aus nur einem einzigen inneren Gitterpunkt und im Fall geometrischer Multiskalen-Zerlegungen gilt $lt = -\log h_0 - 1$. Die Korrekturen bezüglich der Komponenten der Detailunterräume werden, sofern nicht eine andere Methode genannt wird, über eine unvollständige Zerlegung (ILU(0)) der jeweiligen Blöcke A_k $(k = 1, ..., lt)$ berechnet.

In unseren Experimenten bezüglich geometrisch definierter Multiskalen-Zerlegungen hat sich die klassische hierarchische Basis Transformation zur bilinearen Hut-Funktion als die günstigste Wahl für die Transformationen $W^k_{k+1,\mathcal{S}}$ und die dadurch induzierten Multiskalen-Zerlegungen der Testseite erwiesen. Diese wird daher standardmäßig in den zugehörigen Experimenten betrachtet. Im Fall algebraischer Multiskalen-Zerlegungen verwenden wir entsprechend eine AMG-basierte hierarchische Zerlegung für die Testseite. Die Zerlegungen der Ansatzseite ergeben sich durch die Kombination bekannter Prolongationstechniken aus dem Gebiet der Mehrgitter-Verfahren mit hierarchisch und Wavelet-artig aufgespannten Komplementräumen wie in Kapitel 5 beschrieben.

Zur praktischen Beurteilung der Konvergenzgeschwindigkeit unserer Verfahren wird jeweils die durchschnittliche Reduktionsrate des Iterationsfehler bezüglich der l_2-Norm

$$\varrho_{it,1} := \left(\frac{||e_0^{it}||}{||e_0^1||} \right)^{\frac{1}{it-1}} \tag{6.1}$$

während einer vorgegebenen Anzahl $it = 30$ von Iterationen gemessen [64]. Um frei von Starteffekten zu sein, wird die Messung erst nach dem ersten Iterationsschritt begonnen. Aufgrund der Linearität der Verfahren kann man sich dabei ohne Einschränkung auf den Nullvektor als Lösung zur rechten Seite $\mathbf{0}$ und einen nichtverschwindenden Startvektor x_0^0 beschränken [64]. Da die Fehlervektoren $e_0^i = x_0 - x_0^i = -x_0^i$ dann bis auf das Vorzeichen mit den Iterierten übereinstimmen, lassen sich so auf einfache Weise die Iterationsfehler verfolgen. Die Iteration wird vorzeitig abgebrochen, wenn der relative Fehler kleiner als 10^{-10} ist. Die geometrischen Mittel $\varrho_{it,1}$ werden dann bezüglich der bis dahin gemessenen Fehler berechnet. Um einen Eindruck vom asymptotischen Verhalten des Verfahrens zu bekommen, mitteln wir gegebenenfalls auch nur über die letzten fünf durchgeführten Iterationen. Die nachfolgenden Tabellen beziehen sich auf einen normierten und gleichverteilten Zufallsvektor als Startvektor x_0^0, so daß für den Startfehler immer $||e_0^0|| = 1$ gilt. Die Tabellen beschreiben, wenn nicht ausdrücklich auf etwas anderes hingewiesen wird, das Verhalten der durchschnittlichen Reduktionsraten $\varrho_{it,1}$ in Abhängigkeit von der Problemgröße, die über die Gitterweite h_0 des Anfangsgitters Ω_0 gegeben ist, und dem Störungsparameter b. Die ersten Spalten beziehen sich dabei immer auf das ungestörte Poissonproblem mit $b = 0$. Zudem werden noch teilweise die Normen der Iterationsfehler gezeichnet. Um die Abhängigkeit der Verfahren vom Störungsparameter b zu studieren, werden für feste Gitterweiten h_0 die Fehlerkurven zu unterschiedlichen Werten von b im gleichen Plot nebeneinander dargestellt. Um die Abhängigkeit der Verfahren von der Gitterweite h_0 zu untersuchen, werden für festes b die Fehlerkurven zu unterschiedlichen Gitterweiten h_0 nebeneinander gezeichnet.

Die Verfahren wurden im Rahmen eines Computerprogramms mit Hilfe der Programmbibliothek PETSc (Version 2.0.28) [9, 10, 11] unter ANSI C implementiert. Die Testbeispiele T^α, T^β und T^γ wurden jeweils auf einem PC (Dell Precision 620 mit zwei Pentium III Xeon CPUs, 933 MHz) mit zwei GigaByte Hauptspeicher gerechnet. Als Compiler wurde hierbei der GNU C-Compiler in der Version 2.95.2 verwendet. Für das Testbeispiel T^δ wurde eine SGI Origin 200 (vier MIPS R10000 CPUs, 180 MHz, 1 MB Cache) mit insgesamt vier GigaByte Hauptspeicher benutzt. Der C-Compiler war diesmal der des Herstellers.

6.1.2 Rechenaufwand

Entscheidend für die Beurteilung eines Iterationsverfahrens ist neben den Konvergenzraten, die wir anhand numerischer Experimente bestimmen, auch der Rechenaufwand $\mathcal{C}_0$, der pro Iteration zu leisten ist. Diesen bestimmt man durch Abzählen der Rechenoperationen. Aufgrund der rekursiven Struktur unserer Verfahren ist dies jedoch nicht offensichtlich. Wie wir in Kapitel 4 gesehen haben, können unsere Multiskalen-Zyklen mit Hilfe eines speziell definierten Glätters auch als gewöhnliche Mehrgitter-Verfahren interpretiert und implementiert werden. Für solche Implementierungen dürfen wir zur Aufwandsabschätzung direkt auf die entsprechenden Abschätzungen bei Mehrgitter-Verfahren zurückgreifen.

Die in Algorithmus 3 relevanten Schritte sind die Glättungsiteration, die Restriktion des nach dem Vorglättungsschritt berechneten Residuums sowie das Aufaddieren der prolongierten Grobgitterkorrektur. Im Fall von gewöhnlichen Multiskalen-Zyklen mit $\breve{S}_{k-1} = T_{k-1}$ für $k = 1, ...lt$ bezeichnen wir ihren Aufwand mit

- $\mathcal{C}_S N_{k-1}$ Operationen für $x = x + Q_{k,\mathcal{V}}^{k-1} \breve{A}_k^{-1} (Q_{k,\mathcal{S}}^{k-1})^t (f - \breve{S}_{k-1} x)$,
- $\mathcal{C}_R N_{k-1}$ Operationen für $s^{res} = (P_{k,\mathcal{S}}^{k-1})^t (f - \breve{S}_{k-1} x)$,
- $\mathcal{C}_C N_{k-1}$ Operationen für $x = x + P_{k,\mathcal{V}}^{k-1} s_v$.

Im Fall der verbesserten Multiskalen-Zyklen, die für $k > 1$ die Schur-Komplement-Approximationen $\breve{S}_{k-1} = T_{k-1} - C_{k-1} \breve{A}_{k-1}^{-1} B_{k-1}$ verwenden, erhalten wir entsprechend $\mathcal{C}_{S,imp}$ und $\mathcal{C}_{R,imp}$. Für zweidimensionale Probleme und Standardvergröberung gilt

$$N_k \leq \frac{N_{k-1}}{4}. \tag{6.2}$$

Ist $\mathcal{C}_{0,std} N_0$ der Aufwand für einen Schritt von Algorithmus 3 im Fall eines gewöhnlichen Multiskalen-Zyklus mit einem Multiskalen-Vor- und Nachglättungsschritt, so läßt sich dieser abschätzen durch

$$\mathcal{C}_{0,std} N_0 \leq (2\mathcal{C}_S + \mathcal{C}_R + \mathcal{C}_C) N_0 + \mu \mathcal{C}_{1,std} N_1.$$

Wir erhalten mit (6.2) $\mathcal{C}_{0,std} \leq (2\mathcal{C}_S + \mathcal{C}_R + \mathcal{C}_C) + \frac{\mu}{4} \mathcal{C}_{1,std}$ und daher

$$\mathcal{C}_{0,std} \leq (2\mathcal{C}_S + \mathcal{C}_R + \mathcal{C}_C) \left[1 + \frac{\mu}{4} + ... + \left(\frac{\mu}{4} \right)^{lt-1} \right] + \mu^{lt} \frac{\mathcal{C}_{lt,std}}{N_0},$$

wobei $\mathcal{C}_{lt,std}$ den Aufwand für die grobskalige Korrektur auf der Stufe lt bezeichnet. Abschätzen der geometrischen Reihe ergibt wegen $4^{lt} N_{lt} \leq N_0$ schließlich

$$\mathcal{C}_{0,std} \leq \frac{2\mathcal{C}_S + \mathcal{C}_R + \mathcal{C}_C}{1 - \frac{\mu}{4}} + \mathcal{O}(\left(\frac{\mu}{4} \right)^{lt}).$$

Definiert man $\mathcal{C}_{std} := 2\mathcal{C}_S + \mathcal{C}_R + \mathcal{C}_C$ und $\mathcal{C}_{imp} := 2\mathcal{C}_{S,imp} + \mathcal{C}_{R,imp} + \mathcal{C}_C$, so erhalten wir damit die folgenden Aufwandsabschätzungen für die von uns betrachteten Multiskalen-Zyklen:

- Gewöhnlicher Multiskalen-V(1,1)-Zyklus: $\mathcal{C}_{0,std} \leq \frac{4}{3} \mathcal{C}_{std} + \mathcal{O}(\left(\frac{1}{4} \right)^{lt})$,
- verbesserter Multiskalen-V(1,1)-Zyklus: $\mathcal{C}_{0,imp} \leq \mathcal{C}_{std} + \frac{1}{3} \mathcal{C}_{imp} + \mathcal{O}(\left(\frac{1}{4} \right)^{lt})$,
- gewöhnlicher Multiskalen-W(1,1)-Zyklus: $\mathcal{C}_{0,std} \leq 2\mathcal{C}_{std} + \mathcal{O}(\left(\frac{1}{2} \right)^{lt})$.

Benutzen wir im Fall der verbesserten Varianten für $k = 2, ..., lt$ die Darstellungen

$$\breve{S}_{k-1} \quad = \quad T_{k-1} - (P^{k-2}_{k-1,\mathcal{S}})^t T_{k-2} Q^{k-2}_{k-1,\mathcal{V}} \breve{A}^{-1}_{k-1} (Q^{k-2}_{k-1,\mathcal{S}})^t T_{k-2} P^{k-2}_{k-1,\mathcal{V}},$$

so ist die Anwendung von $\breve{S}_{k-1}$ im Vergleich zu der von T_{k-1} zumindest um einen Faktor neun teurer. Wir rechnen dabei lediglich die Kosten der Anwendungen von T_{k-1} sowie T_{k-2} und gehen von 9-Punkte Sternen aus. Eine solche Implementierung der verbesserten Multiskalen-Zyklen ist damit sogar aufwendiger als ein gewöhnlicher Multiskalen-W-Zyklus.

Durch Abzählen der Rechenoperationen ist der Rechenaufwand für Multiskalen-Verfahren im Fall geometrischer Multiskalen-Zerlegungen mit den gerade beschriebenen Formeln unmittelbar vorhersagbar. Gleiches gilt auch für den Speicheraufwand. Das ist anders im Fall algebraischer Multiskalen-Zerlegungen, da die Dimensionen der zerlegenden Unterräume grobskaliger Anteile nicht von vornherein feststehen und ein Gesetz wie (6.2) im allgemeinen nicht gilt. Die Dimensionen der grobskaligen Unterräume ergeben sich erst im Verlauf des Algorithmus und sind von Problem zu Problem verschieden. Es ist nicht einmal die Gesamtzahl der mittels AMG bestimmten Skalen zu Anfang festgelegt. Wir definieren daher zwei Kennzahlen, die *Gitterkomplexität* $\mathcal{C}^g$ und die auf eine AMG-basierte Nichtstandardform ausgerichtete *Nichtstandard-Operatorkomplexität* $\mathcal{C}^o_{ns}$, mit denen sich a posteriori Aussagen über die Kosten AMG-basierter Multiskalen-Verfahren treffen lassen:

$$\mathcal{C}^g := \frac{\sum_{k=0}^{lt} \dim(\mathcal{V}_k)}{\dim(\mathcal{V}_0)} \quad \text{und} \quad \mathcal{C}^o_{ns} := \frac{\mathrm{nnz}(T_0) + \sum_{k=1}^{lt} \left[\mathrm{nnz}(A_k) + \mathrm{nnz}(T_k)\right]}{\mathrm{nnz}(T_0)},$$

wobei $\mathrm{nnz}(M)$ für eine Matrix M die Anzahl ihrer Einträge ungleich Null bezeichnet. Die Gitterkomplexität $\mathcal{C}^g$ gibt eine Vorstellung von dem Speicheraufwand für die im Multiskalen-Verfahren verwendeten (Hilfs-)Vektoren. Mit der Nichtstandard-Operatorkomplexität $\mathcal{C}^o_{ns}$ erhält man ein Maß für den Speicheraufwand hinsichtlich der für ein Multiskalen-Verfahren abzuspeichernden Operatoren. Nimmt man an, daß die Anzahl der Fließkommazahl-Operationen auf jeder Stufe des Verfahrens proportional zur Zahl der Einträge der Grobgitteroperatoren ist, kann man mit Hilfe von $\mathcal{C}^o_{ns}$ ebenfalls den Rechenaufwand von V-Zyklus-artigen Multiskalen-Verfahren sinnvoll abschätzen.

6.2 Testbeispiel T^α: Konvektion in x-Richtung

6.2.1 Hierarchische Zerlegungen des Ansatzraums

Betrachtet man als Lösungsverfahren einen gewöhnlichen Multiskalen-V(1,1)-Zyklus zusammen mit klassischen hierarchische Basis Transformationen für die Ansatz- und Testseite, so bricht das darüber definierte Galerkin-artige Multiskalen-Verfahren mit zunehmender Konvektionsstärke sehr bald zusammen und divergiert. Das ist nicht weiter verwunderlich, da bekanntlich die durch diesen Vergröberungsprozeß gebildeten Grobgitteroperatoren mit wachsendem b zunehmend instabil werden und die darüber berechneten „Korrekturen“ die Konvergenz des Verfahrens zerstören [54, 132].

Deshalb gehen wir auf der Ansatzseite zur Verwendung matrixabhängiger Prolongationen und zugehöriger hierarchischer Zerlegungen über. Tabelle 6.1 zeigt die durchschnittlichen Fehlerreduktionsraten $\varrho_{it,1}$ für den Fall, daß die Zerlegung des Ansatzraums über die matrixabhängige hierarchische Transformation wie in Kapitel 5.2 definiert wird. Die Raten steigen mit zunehmender Problemgröße, was schon durch das Verhalten der klassischen Hierarchische Basis Mehrgitter-Methode in Bezug auf das konvektionsfreie Problem mit reinem

$h_0^{-1} \setminus b$	0.0	10^1	10^2	10^3	10^4	10^5	10^6	10^8
8	0.29	0.38	0.58	0.56	0.58	0.63	0.64	0.64
16	0.36	0.51	0.71	0.78	0.79	0.80	0.78	0.79
32	0.52	0.57	0.81	0.87	0.87	0.87	0.88	0.88
64	0.60	0.67	0.85	0.91	0.93	0.93	0.93	0.94
128	0.64	0.69	0.89	0.93	0.93	0.95	0.96	0.96
256	0.69	0.69	0.90	0.93	0.94	0.96	0.97	0.97
512	0.72	0.75	0.86	0.93	0.94	0.96	0.98	0.98
1024	0.75	0.76	0.88	0.92	0.96	0.95	0.98	0.99

$P^k_{k+1,\mathcal{V}}$: matrixabhängig $Q^k_{k+1,\mathcal{V}}$: hierarchischer Filter [1]
$P^k_{k+1,\mathcal{S}}$: bilinear $Q^k_{k+1,\mathcal{S}}$: hierarchischer Filter [1]

TABELLE 6.1: *Beispiel T^α: Gewöhnlicher Multiskalen-V(1,1)-Zyklus, $\varrho_{it,1}$.*

$h_0^{-1} \setminus b$	0.0	10^1	10^2	10^3	10^4	10^5	10^6	10^8
8	0.26	0.37	0.59	0.61	0.64	0.64	0.65	0.66
16	0.37	0.50	0.73	0.79	0.79	0.83	0.83	0.80
32	0.50	0.56	0.78	0.87	0.90	0.88	0.90	0.90
64	0.59	0.66	0.85	0.88	0.92	0.93	0.93	0.93
128	0.63	0.69	0.86	0.88	0.92	†	0.94	0.94
256	0.68	0.68	0.87	0.90	0.92	0.94	0.95	0.95
512	0.71	0.75	0.86	0.89	0.92	0.94	0.95	†
1024	0.73	0.78	0.87	0.89	0.92	0.83	0.95	†

$P^k_{k+1,\mathcal{V}}$: matrixabhängig $Q^k_{k+1,\mathcal{V}}$: hierarchischer Filter [1]
$P^k_{k+1,\mathcal{S}}$: bilinear $Q^k_{k+1,\mathcal{S}}$: hierarchischer Filter [1]

TABELLE 6.2: *Beispiel T^α: Gewöhnlicher Multiskalen-V(1,1)-Zyklus, $\varrho_{it,1}$, exakte Glättung.*

$h_0^{-1} \setminus b$	0.0	10^1	10^2	10^3	10^4	10^5	10^6	10^8
8	0.21	0.28	0.42	0.49	0.50	0.52	0.52	0.52
16	0.23	0.27	0.55	0.60	0.63	0.65	0.63	0.64
32	0.24	0.27	0.55	0.72	0.76	0.77	0.77	0.77
64	0.24	0.25	0.50	0.78	0.86	0.89	0.90	0.90
128	0.25	0.25	0.43	0.80	0.93	0.95	0.97	0.96
256	0.25	0.25	0.33	0.77	†	†	†	†
512	0.25	0.25	0.26	0.74	†	†	†	†
1024	0.25	0.25	0.25	0.65	†	†	†	†

$P^k_{k+1,\mathcal{V}}$: matrixabhängig $Q^k_{k+1,\mathcal{V}}$: hierarchischer Filter [1]
$P^k_{k+1,\mathcal{S}}$: bilinear $Q^k_{k+1,\mathcal{S}}$: hierarchischer Filter [1]

TABELLE 6.3: *Beispiel T^α: Gewöhnlicher Multiskalen-W(1,1)-Zyklus, $\varrho_{it,1}$.*

Laplace-Operator nahegelegt wird (siehe dazu die erste Spalte von Tabelle 6.1). Sie steigen darüberhinaus stark für feste Problemgröße mit wachsender Konvektionsstärke. Dies überrascht, da hierbei die gleichen stabilen Grobgitteroperatoren gebildet werden, die in der entsprechenden Mehrgitter-Variante mit matrixabhängigen Transferoperatoren schon zusammen mit einfachen punktweisen Glättern zu robusten Verfahren führen [54].

Die Einträge der Blöcke A_k stimmen mit entsprechenden Einträgen der Matrizen T_{k-1} überein, da sie lediglich über hierarchische Transformationen berechnet werden. Sie tragen daher noch vollständig den konvektiven Anteil von T_{k-1} bezüglich der Fein-ohne-Grobgitter Freiheitsgrade in sich. Um die Auswirkungen der unvollständigen Zerlegung der Blöcke A_k bei zunehmender Konvektion von dem Effekt der matrixabhängigen Bestimmung der Grobgitteroperatoren zu trennen, benutzen wir einen iterativen Löser (BICGSTAB mit ILU-Vorkonditionierer), um jeweils die Anwendung von A_k^{-1} zu realisieren. Damit berechnen wir

$h_0^{-1} \setminus b$	0.0	10^1	10^2	10^3	10^4	10^5	10^6	10^8
8	0.13	0.08	0.03	0.07	0.07	0.07	0.07	0.07
16	0.15	0.10	0.05	0.13	0.14	0.14	0.14	0.14
32	0.16	0.11	0.08	0.16	0.19	0.20	0.20	0.20
64	0.17	0.12	0.10	0.17	0.23	0.25	0.25	0.25
128	0.17	0.13	0.12	0.18	0.26	0.28	0.29	0.29
256	0.17	0.14	0.12	0.18	0.27	0.31	0.32	0.32
512	0.17	0.14	0.13	0.18	0.27	0.32	0.34	0.35
1024	0.17	0.14	0.13	0.18	0.27	0.33	0.36	0.37

$P^k_{k+1,\mathcal{V}}$: matrixabhängig $Q^k_{k+1,\mathcal{V}}$: Pre-Prewavelet Filter
$P^k_{k+1,\mathcal{S}}$: bilinear $Q^k_{k+1,\mathcal{S}}$: hierarchischer Filter [1]

TABELLE 6.4: *Beispiel T^α: Gewöhnlicher Multiskalen-V(1,1)-Zyklus, $\varrho_{it,1}$.*

$h_0^{-1} \setminus b$	0.0	10^1	10^2	10^3	10^4	10^5	10^6	10^8
8	0.17	0.13	0.08	0.07	0.07	0.07	0.07	0.07
16	0.20	0.16	0.11	0.08	0.08	0.08	0.08	0.08
32	0.24	0.19	0.14	0.11	0.08	0.07	0.07	0.08
64	0.25	0.21	0.17	0.13	0.09	0.08	0.07	0.07
128	0.26	0.23	0.19	0.17	0.11	0.08	0.08	0.08
256	0.27	0.24	0.20	0.19	0.16	0.10	0.08	0.08
512	0.27	0.25	0.21	0.21	0.21	0.13	0.09	0.08
1024	0.28	0.26	0.21	0.22	0.24	0.19	0.11	0.08

$P^k_{k+1,\mathcal{V}}$: matrixabhängig $Q^k_{k+1,\mathcal{V}}$: Prewavelet Filter
$P^k_{k+1,\mathcal{S}}$: bilinear $Q^k_{k+1,\mathcal{S}}$: hierarchischer Filter [1]

TABELLE 6.5: *Beispiel T^α: Gewöhnlicher Multiskalen-V(1,1)-Zyklus, $\varrho_{it,1}$.*

die *bestmöglichen* Detailkorrekturen. Tabelle 6.2 zeigt, daß hierdurch, wenn überhaupt, dann nur eine marginale Verbesserung der durchschnittlichen Reduktionsraten erreicht wird. In einigen Fällen divergierte das Verfahren sogar. Wir sprechen dabei von Divergenz (gekennzeichnet in den Tabellen durch †), falls die zugehörige gemittelte Rate größer gleich Eins ist. Die Detailkorrekturen mittels der unvollständigen Zerlegungen sind offenbar „ausreichend“ und damit ebenfalls die Zahl der Multiskalen-Glättungsschritte, also der ILU-Schritte während der Vor- und Nachkorrekturen. Aus Sicht der Mehrgitter-Interpretation des Multiskalen-Verfahrens kommt der Verlust an Robustheit daher durch einen prinzipiellen Defekt des hierarchischen Multiskalen-Glätters zustande.

Untersucht man die Auswirkungen, die sich durch die Berechnung der hierarchischen Residuen über die Schur-Komplement Approximation wie in Algorithmus 2’ ergeben, so stellt man fest, daß im Zusammenhang mit hierarchischen Zerlegungen durch solche verbesserten Multiskalen-V(1,1)-Zyklen kaum Verbesserungen der Reduktionsraten im Vergleich zu den gewöhnlichen Verfahren erreicht werden. Wir verzichten auf die Angabe der Ergebnisse.

Betrachtet man einen gewöhnlichen Multiskalen-W(1,1)-Zyklus, so erhält man wie erwartet zwar im niederkonvektiven Bereich ein optimales Verfahren mit $\varrho_{it,1} = 0.25$, für große Konvektionsstärken bei kleinen Gitterweiten jedoch Divergenz, wie man direkt Tabelle 6.3 entnimmt. Man gelangt trotz der durch den mehrfachen Aufruf von Grobgitterkorrekturen zusätzlich geleisteten hierarchischen Glättungsarbeit zu keinem robusten Löser.

6.2.2 Wavelet-artige Zerlegungen des Ansatzraums

Wir betrachten nun Wavelet-artige Zerlegungen des Ansatzraums, die über Zweiskalen-Transformationen $W^k_{k+1,\mathcal{V}}$ erzeugt werden, deren Teilmatrizen $P^k_{k+1,\mathcal{V}}$ wie schon im hierarchischen

$h_0^{-1} \setminus b$	0.0	10^1	10^2	10^3	10^4	10^5	10^6	10^8
8	0.10	0.10	0.07	0.06	0.06	0.06	0.06	0.06
16	0.17	0.17	0.12	0.09	0.07	0.07	0.07	0.07
32	0.22	0.23	0.16	0.11	0.08	0.07	0.07	0.08
64	0.24	0.25	0.19	0.14	0.09	0.08	0.07	0.07
128	0.25	0.27	0.23	0.16	0.11	0.08	0.08	0.07
256	0.26	0.28	0.25	0.17	0.14	0.10	0.08	0.08
512	0.27	0.28	0.27	0.19	0.17	0.13	0.09	0.08
1024	0.27	0.29	0.28	0.21	0.19	0.16	0.12	0.08

$P^k_{k+1,\mathcal{V}}$: matrixabhängig $Q^k_{k+1,\mathcal{V}}$: L-Spline Prewavelet Filter
$P^k_{k+1,\mathcal{S}}$: bilinear $Q^k_{k+1,\mathcal{S}}$: hierarchischer Filter [1]

TABELLE 6.6: *Beispiel T^α: Gewöhnlicher Multiskalen-V(1,1)-Zyklus, $\varrho_{it,1}$, $lt = -\log h_0 - 2$.*

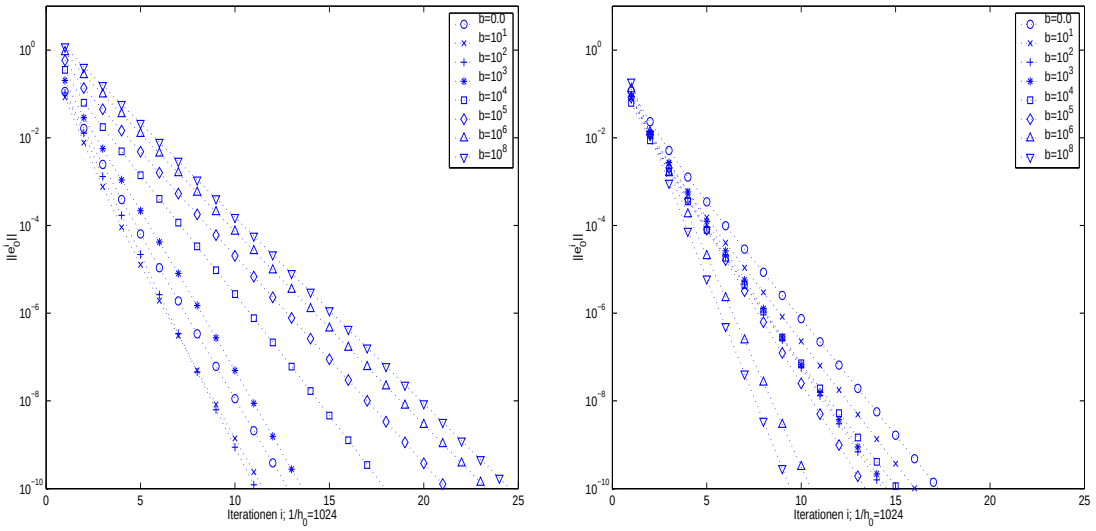


ABBILDUNG 6.1: *Beispiel T^α: Entwicklung der Fehlernormen bei zunehmender Konvektionsstärke für die feste Gitterweite $h_0 = \frac{1}{1024}$, gewöhnlicher Multiskalen-V(1,1)-Zyklus, jeweils klassische hierarchische Zerlegung des Testraums, Pre-Prewavelet-basierte Zerlegung des Ansatzraums (links), Prewavelet-basierte Zerlegung des Ansatzraums (rechts).*

Fall matrixabhängig sind. Die Teilmatrizen $Q^k_{k+1,\mathcal{V}}$ werden nun allerdings über die Tensorprodukt-Konstruktion aus Kapitel 5.2 mittels eindimensionaler Skalierungs- und Waveletfilter konstruiert.

In Tabelle 6.4 sieht man deutlich den stabilisierenden Einfluß, der durch eine solche Wavelet-artige Zerlegung des Ansatzraums erreicht wird. Hierbei werden das feste eindimensionale Molekül $[-1 \quad \underline{2} \quad -1]/2$ und seine Randmodifikationen $[\underline{2} \quad -1 \quad 0]/2$ sowie $[0 \quad -1 \quad \underline{2}]/2$ zusammen mit matrixabhängig bestimmten eindimensionalen Skalierungsfiltern zum Aufbau der Matrizen $Q^k_{k+1,\mathcal{V}}$ eingesetzt. Man stellt eine deutliche Stabilisierung sowohl in Bezug auf die Gitterweite h_0 als auch auf den Störungsparameter b fest. Verwenden wir das Molekül $[1 \quad -6 \quad \underline{10} \quad -6 \quad 1]/10$ und die Randmodifikationen $[\underline{9} \quad -6 \quad 1]/10$ sowie $[1 \quad -6 \quad \underline{9}]/10$ als eindimensionale Detailfilter, so erhält man offensichtlich ein robustes Verfahren, wie Tabelle 6.5 zeigt. Mit zunehmender Konvektionsstärke werden für feste Gitterweite h_0 die zugehörigen Raten sogar besser. Matrixabhängig berechnete L-Spline Prewavelets führen zum gleichen Erfolg, was man anhand von Tabelle 6.6 erkennt. Die Zerlegungstiefe beträgt hierbei jedoch nur $lt = -\log h_0 - 2$, so daß das gröbste Gitter aus insgesamt neun inneren Punkten besteht.

$h_0^{-1} \setminus b$	0.0	10^1	10^2	10^3	10^4	10^5	10^6	10^8
8	0.29	0.37	0.55	0.56	0.54	0.58	0.58	0.59
16	0.36	0.48	0.71	0.73	0.74	0.75	0.75	0.73
32	0.52	0.57	0.79	0.83	0.83	0.80	0.81	0.83
64	0.60	0.67	0.85	0.81	0.88	0.87	0.85	0.87
128	0.64	0.69	0.86	0.88	0.87	0.90	0.88	0.88
256	0.69	0.71	0.86	0.89	0.90	0.90	0.89	0.89
512	0.72	0.76	0.86	0.90	0.91	0.91	0.91	0.90
1024	0.75	0.77	0.87	0.90	0.91	0.91	0.91	0.92

$P^k_{k+1,\mathcal{V}}$: matrixabhängig $Q^k_{k+1,\mathcal{V}}$: hierarchischer Filter [1]
$P^k_{k+1,\mathcal{S}}$: bilinear $Q^k_{k+1,\mathcal{S}}$: hierarchischer Filter [1]

TABELLE 6.7: *Beispiel T^β: Gewöhnlicher Multiskalen-V(1,1)-Zyklus, $\varrho_{it,1}$.*

$h_0^{-1} \setminus b$	0.0	10^1	10^2	10^3	10^4	10^5	10^6	10^8
8	0.26	0.36	0.55	0.53	0.56	0.57	0.57	0.59
16	0.37	0.46	0.71	0.72	0.74	0.75	0.75	0.72
32	0.50	0.58	0.79	0.83	0.84	0.78	0.81	0.84
64	0.59	0.67	0.85	0.83	0.88	0.87	0.85	0.87
128	0.63	0.69	0.85	0.87	0.87	0.90	0.88	0.87
256	0.68	0.71	0.85	0.88	0.89	0.89	0.88	0.88
512	0.71	0.76	0.86	0.88	0.90	0.90	0.90	0.89
1024	0.73	0.77	0.85	0.89	0.90	0.89	0.90	0.90

$P^k_{k+1,\mathcal{V}}$: matrixabhängig $Q^k_{k+1,\mathcal{V}}$: hierarchischer Filter [1]
$P^k_{k+1,\mathcal{S}}$: bilinear $Q^k_{k+1,\mathcal{S}}$: hierarchischer Filter [1]

TABELLE 6.8: *Beispiel T^β: Gewöhnlicher Multiskalen-V(1,1)-Zyklus, $\varrho_{it,1}$, exakte Glättung.*

$h_0^{-1} \setminus b$	0.0	10^1	10^2	10^3	10^4	10^5	10^6	10^8
8	0.21	0.23	0.36	0.42	0.39	0.44	0.42	0.42
16	0.23	0.24	0.38	0.49	0.50	0.48	0.50	0.49
32	0.24	0.25	0.40	0.50	0.55	0.53	0.54	0.54
64	0.24	0.24	0.38	0.54	0.56	0.58	0.57	0.57
128	0.25	0.25	0.32	0.54	0.58	0.59	0.59	0.60
256	0.25	0.25	0.26	0.53	0.61	0.62	0.61	0.62
512	0.25	0.25	0.25	0.50	0.62	0.64	0.65	0.66
1024	0.25	0.25	0.25	0.43	0.65	0.68	0.69	0.69

$P^k_{k+1,\mathcal{V}}$: matrixabhängig $Q^k_{k+1,\mathcal{V}}$: hierarchischer Filter [1]
$P^k_{k+1,\mathcal{S}}$: bilinear $Q^k_{k+1,\mathcal{S}}$: hierarchischer Filter [1]

TABELLE 6.9: *Beispiel T^α: Gewöhnlicher Multiskalen-W(1,1)-Zyklus, $\varrho_{it,1}$.*

Dies erklärt die leichten Unterschiede der ersten Spalten der Tabellen 6.5 und 6.6.

Die Graphen in Abbildung 6.1 zeigen schließlich das Verhalten der Fehlernormen, die bei der Berechnung der Tabellen 6.4 und 6.5 auftreten. Sie stellen jeweils für die minimale Gitterweite $h_0 = \frac{1}{1024}$ die Entwicklung der Fehlernormen $||e_0^i||$ bei zunehmender Konvektionsstärke dar.

6.3 Testbeispiel T^β: Konvektion in Diagonalrichtung

6.3.1 Hierarchische Zerlegungen des Ansatzraums

Für diagonal gerichtete Konvektion erhält man mit hierachisch basierten Multiskalen-V-Zyklen im wesentlichen analoge Ergebnisse wie im achsenorientierten Fall. Wir sehen in Tabelle 6.7, daß die matrixabhängige hierarchische Wahl der Ansatzraum-Zerlegung und der

$h_0^{-1} \setminus b$	0.0	10^1	10^2	10^3	10^4	10^5	10^6	10^8
8	0.17	0.10	0.08	0.09	0.09	0.10	0.10	0.10
16	0.20	0.14	0.13	0.17	0.17	0.17	0.17	0.17
32	0.24	0.17	0.19	0.22	0.26	0.26	0.25	0.25
64	0.25	0.21	0.21	0.27	0.28	0.28	0.28	0.28
128	0.26	0.22	0.21	0.29	0.31	0.31	0.31	0.31
256	0.27	0.23	0.22	0.30	0.37	0.38	0.39	0.39
512	0.27	0.24	0.23	0.33	0.48	0.51	0.51	0.51
1024	0.28	0.25	0.23	0.36	0.76	0.90	0.92	0.92

$P^k_{k+1,\mathcal{V}}$: matrixabhängig $Q^k_{k+1,\mathcal{V}}$: Prewavelet Filter
$P^k_{k+1,\mathcal{S}}$: bilinear $Q^k_{k+1,\mathcal{S}}$: hierarchischer Filter [1]

TABELLE 6.10: *Beispiel T^β: Gewöhnlicher Multiskalen-V(1,1)-Zyklus, $\varrho_{it,1}$.*

$h_0^{-1} \setminus b$	0.0	10^1	10^2	10^3	10^4	10^5	10^6	10^8
8	0.13	0.07	0.09	0.14	0.15	0.15	0.16	0.15
16	0.15	0.09	0.13	0.22	0.24	0.25	0.24	0.24
32	0.16	0.10	0.15	0.27	0.29	0.29	0.29	0.29
64	0.17	0.12	0.17	0.29	0.32	0.32	0.32	0.32
128	0.17	0.13	0.18	0.32	0.36	0.36	0.37	0.37
256	0.17	0.14	0.19	0.35	0.42	0.42	0.42	0.43
512	0.17	0.14	0.20	0.38	0.50	0.53	0.53	0.53
1024	0.17	0.14	0.20	0.40	0.62	0.67	0.68	0.68

$P^k_{k+1,\mathcal{V}}$: matrixabhängig $Q^k_{k+1,\mathcal{V}}$: Pre-Prewavelet Filter
$P^k_{k+1,\mathcal{S}}$: bilinear $Q^k_{k+1,\mathcal{S}}$: hierarchischer Filter [1]

TABELLE 6.11: *Beispiel T^β: Gewöhnlicher Multiskalen-V(1,1)-Zyklus, $\varrho_{it,1}$.*

gewöhnliche Multiskalen-V(1,1)-Zyklus wieder zu keinem robusten Verfahren führen. Der Vergleich mit Tabelle 6.8, die die Ergebnisse im Fall exakter Detailkorrekturen enthält, zeigt, daß auch hier der Verlust an Robustheit nicht durch die unvollständigen Zerlegungen sondern durch einen prinzipiellen Defekt des Multiskalen-Glätters im Fall hierarchischer Zerlegungen der Ansatz- und Testseite zustande kommt. Verbesserte Multiskalen-V(1,1)-Zyklen liefern wiederum kaum Verbesserungen der Reduktionsraten im Vergleich zu den gewöhnlichen Verfahren.

Ein gewöhnlicher Multiskalen-W(1,1)-Zyklus ergibt im schwachkonvektiven Bereich wiederum ein optimales Verfahren, für große Konvektionsstärken erweist er sich jedoch ebenfalls als nicht besonders robust, wie direkt Tabelle 6.9 zu entnehmen ist. Im Unterschied zum Beispiel mit achsenorientierter Konvektion konvergiert das Verfahren aber in allen Fällen. Für große Konvektionsstärken scheint die Konvergenzrate des Multiskalen-W(1,1)-Zyklus durch 0.7 nach oben hin beschränkt zu sein.

6.3.2 Wavelet-artige Zerlegungen des Ansatzraums

Wir betrachten nun die Ergebnisse, die sich durch Wavelet-artige Zerlegungen des Ansatzraums zusammen mit denselben matrixabhängigen Prolongationen $P^k_{k+1,\mathcal{V}}$ wie im vorherigen Abschnitt ergeben.

Benutzt man die festen eindimensionalen Moleküle $[1 \quad -6 \quad \underline{10} \, -6 \quad 1]/10$ sowie $[-1 \quad \underline{2} \, -1]/2$ und ihre zugehörigen Randmodifikationen zusammen mit matrixabhängig bestimmten eindimensionalen Skalierungsfiltern zur Konstruktion der Matrizen $Q^k_{k+1,\mathcal{V}}$, so erkennt man im Vergleich der Tabellen 6.10 und 6.11 mit Tabelle 6.7 wiederum deutlich die Stabilisierung hinsichtlich der Abhängigkeit von der Problemgröße. Robustheit ist aus beiden Tabellen je-

$h_0^{-1} \setminus b$	0.0	10^1	10^2	10^3	10^4	10^5	10^6	10^8
8	0.10	0.11	0.15	0.12	0.14	0.14	0.14	0.14
16	0.17	0.18	0.26	0.26	0.26	0.26	0.26	0.26
32	0.22	0.27	0.37	0.37	0.40	0.39	0.40	0.39
64	0.24	0.28	0.49	0.53	0.55	0.56	0.56	0.52
128	0.25	0.29	0.51	0.68	0.75	0.75	0.76	0.76
256	0.26	0.29	0.54	0.88	0.99	†	†	†
512	0.27	0.30	0.58	0.98	†	†	†	†
1024	0.27	0.30	0.59	0.99	†	†	†	†

$P^k_{k+1,\mathcal{V}}$: matrixabhängig $Q^k_{k+1,\mathcal{V}}$: L-Spline Prewavelet Filter
$P^k_{k+1,\mathcal{S}}$: bilinear $Q^k_{k+1,\mathcal{S}}$: hierarchischer Filter [1]

TABELLE 6.12: *Beispiel T^β: Gewöhnlicher Multiskalen-V(1,1)-Zyklus, $\varrho_{it,1}$.*

$h_0^{-1} \setminus b$	0.0	10^1	10^2	10^3	10^4	10^5	10^6	10^8
8	0.13	0.07	0.10	0.15	0.16	0.16	0.16	0.16
16	0.15	0.09	0.13	0.23	0.24	0.25	0.24	0.24
32	0.16	0.10	0.16	0.27	0.29	0.28	0.29	0.28
64	0.17	0.13	0.18	0.29	0.31	0.31	0.31	0.32
128	0.18	0.14	0.20	0.32	0.36	0.36	0.37	0.36
256	0.18	0.15	0.21	0.36	0.41	0.41	0.41	0.41
512	0.18	0.15	0.21	0.39	0.45	0.46	0.46	0.46
1024	0.19	0.15	0.22	0.41	0.48	0.49	0.50	0.50

$P^k_{k+1,\mathcal{V}}$: matrixabhängig $Q^k_{k+1,\mathcal{V}}$: Pre-Prewavelet Filter
$P^k_{k+1,\mathcal{S}}$: bilinear $Q^k_{k+1,\mathcal{S}}$: hierarchischer Filter [1]

TABELLE 6.13: *Beispiel T^β: Gewöhnlicher Multiskalen-V(1,1)-Zyklus, $\varrho_{it,1}$, exakte Glättung.*

$h_0^{-1} \setminus b$	0.0	10^1	10^2	10^3	10^4	10^5	10^6	10^8
8	0.13	0.06	0.08	0.10	0.10	0.10	0.10	0.10
16	0.14	0.07	0.10	0.13	0.13	0.13	0.13	0.13
32	0.15	0.09	0.12	0.15	0.16	0.16	0.16	0.16
64	0.15	0.10	0.12	0.17	0.18	0.18	0.18	0.18
128	0.15	0.11	0.13	0.19	0.20	0.20	0.20	0.20
256	0.15	0.11	0.13	0.20	0.24	0.25	0.25	0.25
512	0.15	0.12	0.14	0.21	0.32	0.33	0.33	0.33
1024	0.15	0.12	0.14	0.24	0.41	0.48	0.49	0.49

$P^k_{k+1,\mathcal{V}}$: matrixabhängig $Q^k_{k+1,\mathcal{V}}$: Pre-Prewavelet Filter
$P^k_{k+1,\mathcal{S}}$: bilinear $Q^k_{k+1,\mathcal{S}}$: hierarchischer Filter [1]

TABELLE 6.14: *Beispiel T^β: Verbesserter Multiskalen-V(1,1)-Zyklus, $\varrho_{it,1}$.*

doch nicht herauszulesen. Das gilt insbesondere für Tabelle 6.10, in deren letzter Zeile die durchschnittlichen Konvergenzraten sogar auf 0.92 steigen.

Die Verwendung problemangepaßter L-Spline Prewavelet Filter wirkt sich ungünstig aus. Die damit gebildeten Zyklen sind nicht robust und divergieren sogar für große Konvektionsstärken und kleine Gitterweiten, wie man in Tabelle 6.12 erkennt. Die Zerlegungstiefe beträgt hierbei wiederum nur $lt = -\log h_0 - 2$.

Auch für Wavelet-artige Zerlegungen benutzen wir wie schon im Fall hierarchischer Zerlegungen einen iterativen Löser (diesmal ein "least squares"-Verfahren), um möglichst genaue Detailkorrekturen zu berechnen. Dies geschieht ebenfalls mit der Absicht, die Auswirkungen der unvollständigen Zerlegung der Blöcke A_k bei zunehmender Konvektion von dem Effekt der matrixabhängigen Bestimmung der Grobgitteroperatoren zu trennen. Die dazu dargestellte

$h_0^{-1} \setminus b$	0.0	10^1	10^2	10^3	10^4	10^5	10^6	10^8
8	0.07	0.06	0.08	0.13	0.15	0.15	0.15	0.15
16	0.08	0.07	0.11	0.23	0.24	0.25	0.24	0.24
32	0.08	0.08	0.11	0.25	0.29	0.29	0.30	0.28
64	0.08	0.08	0.09	0.24	0.30	0.31	0.30	0.31
128	0.08	0.08	0.08	0.20	0.30	0.31	0.31	0.31
256	0.08	0.08	0.08	0.14	0.28	0.31	0.31	0.31
512	0.08	0.08	0.08	0.10	0.26	0.30	0.30	0.30
1024	0.08	0.08	0.08	0.09	0.20	0.28	0.29	0.29

$P^k_{k+1,\mathcal{V}}$: matrixabhängig $Q^k_{k+1,\mathcal{V}}$: Pre-Prewavelet Filter
$P^k_{k+1,\mathcal{S}}$: bilinear $Q^k_{k+1,\mathcal{S}}$: hierarchischer Filter [1]

TABELLE 6.15: *Beispiel T^β: Gewöhnlicher Multiskalen-W(1,1)-Zyklus, $\varrho_{it,1}$.*

Tabelle 6.13 für die Pre-Prewavelet-basierten Zerlegungen des Ansatzraums belegt wiederum die prinzipielle Schwäche der Multiskalen-Glätter. Man erwartet, daß die Verfahren mit exakten Detailkorrekturen zwangsläufig besser konvergieren als die ILU-Varianten. Dem ist jedoch nicht immer so, was sich sogar im konvektionsfreien Fall beobachten läßt. Man vergleiche dazu die jeweils ersten Spalten der Tabelle 6.13 mit den entsprechenden Spalten von Tabelle 6.11. Ein möglicher Grund hierfür ist, daß die während des Vergröberungsprozesses berechneten Grobgitteroperatoren nicht mehr genau zu den Feingitterproblemen passen. Sie liefern lediglich gemittelte Gleichungen und nicht die durch die Schur-Komplement Operatoren definierten reduzierten Gleichungen, die feinskalige Effekte berücksichtigen [31]. Detailkorrekturen, die mittels unvollständiger Zerlegungen der Blöcke A_k bestimmt werden, verzeihen dies unter Umständen eher als es exakte Korrekturen tun.

Anders als im Fall hierarchischer Zerlegungen erreichen wir bei Wavelet-artigen Zerlegungen der Ansatzseite mit einem verbesserten Multiskalen-V(1,1)-Zyklus eine merkliche Verbesserung der Fehlerreduktionsraten durch die Schur-Komplement Approximation der verallgemeinerten hierarchischen Residuen. Dies zeigt Tabelle 6.14. Man stellt dennoch ein merkliches Ansteigen der Raten fest. Auch durch die verbesserten Multiskalen-V(1,1)-Zyklen erhält man also keine robusten Löser. Wie wir bei der Aufwandsdiskussion zu Anfang des Kapitels gesehen haben, sind die Kosten für die verbesserten Varianten sogar höher als die für gewöhnliche W-Zyklen. Wir versuchen deshalb den Verlust an Robustheit durch ein Erhöhen der Zyklustiefe μ zu beheben.

Für die von uns favorisierten Pre-Prewavelet-Zerlegungen der Ansaztseite findet man in Tabelle 6.15 die Ergebnisse, die man mit einem gewöhnlichen Multiskalen-W(1,1)-Zyklus erhält, der die Korrekturen der Detailanteile wie üblich über eine unvollständige Faktorisierung der Blöcke A_k berechnet. Wir sehen, daß der mehrfache Aufruf der Grobgitterkorrekturen und die damit zusätzlich geleistete Multiskalen-Glättungsarbeit diesmal zu einem anscheinend robusten Verfahren führt mit durch 0.31 beschränkten Konvergenzraten führt. Zur genaueren Analyse betrachten wir nun die Verfahren für große Konvektionsstärke im Detail.

Abbildung 6.2 zeigt das Verhalten der Fehlernormen, die bei der Berechnung der Tabellen 6.11, 6.13, 6.14 und 6.15 auftreten. Die einzelnen Bilder stellen für den jeweils festen Störungsparameter $b = 10^6$ die Entwicklung der Fehlernormen zu unterschiedlichen Gitterweiten dar. Interessant ist im oberen linken Bild, das die Ergebnisse im Fall gewöhnlicher Multiskalen-V(1,1)-Zyklen wiedergibt, das Ansteigen der Fehlernormen für Gitterweiten $h_0 \geq \frac{1}{512}$ während der ersten Iterationsschritte. Die verbesserten Multiskalen-V(1,1)-Zyklen scheinen diese Phase sogar noch etwas zu verlängern (siehe das obere rechte Bild von Abbildung

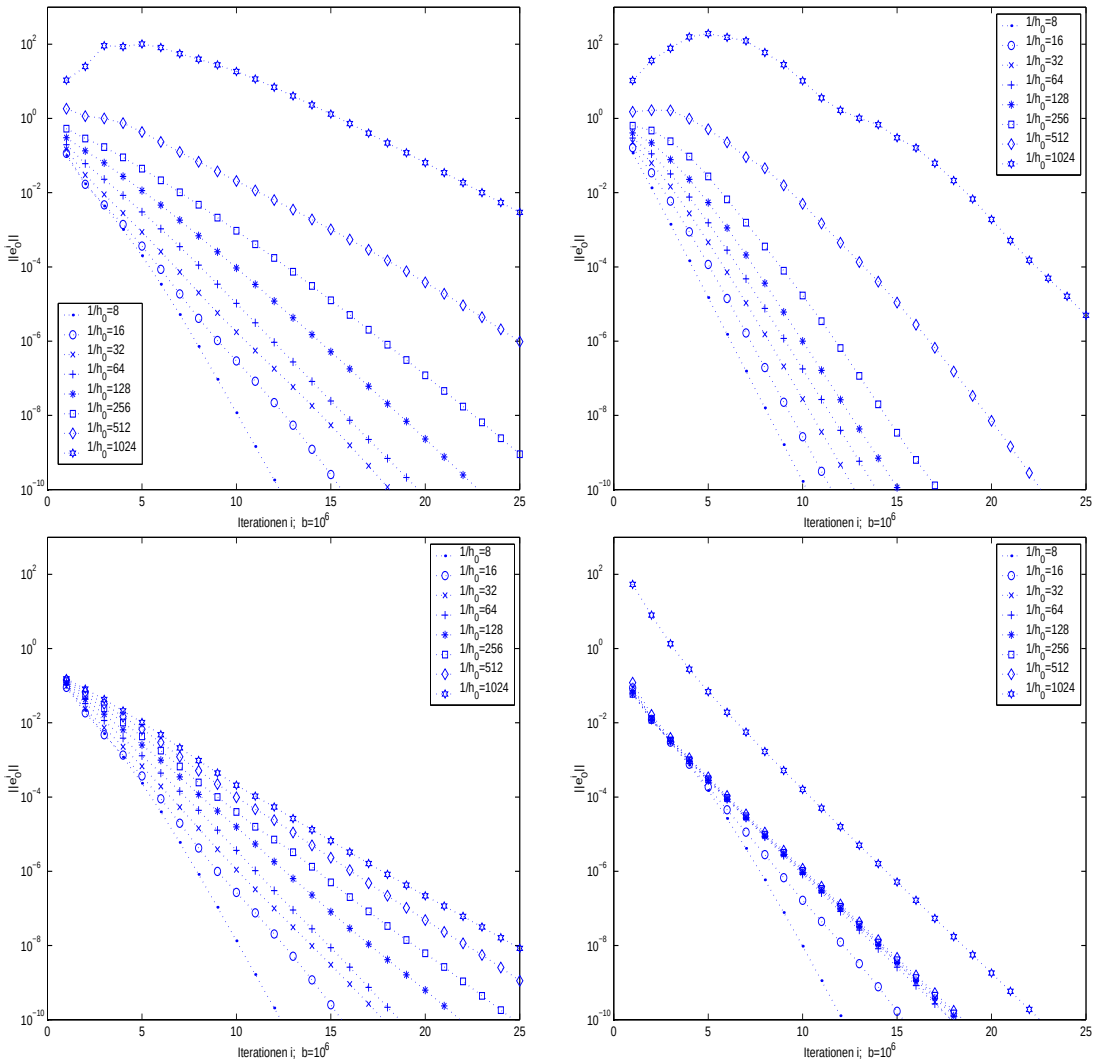


ABBILDUNG 6.2: *Beispiel T^β: Entwicklung der Fehlernormen für unterschiedliche Gitterweiten, $b = 10^6$, klassische hierarchische Zerlegung des Testraums, Pre-Prewavelet-basierte Zerlegung des Ansatzraums; gewöhnlicher Multiskalen-V(1,1)-Zyklus (oben links), verbesserter Multiskalen-V(1,1)-Zyklus (oben rechts), gewöhnlicher Multiskalen-V(1,1)-Zyklus mit exakter Multiskalen-Glättung (unten links), gewöhnlicher Multiskalen-W(1,1)-Zyklus (unten rechts).*

6.2). Dieses Verhalten deutet darauf hin, daß die Multiskalen-Glättung zu Beginn der Iteration gewissen Schwierigkeiten unterliegt. Im Fall der exakten Multiskalen-Glättung, deren zugehörige Fehlerentwicklung im unteren linken Bild von Abbildung 6.2 dargestellt ist, ist die Konvergenz hingegen monoton. Für das Ansteigen der Fehlernormen während der ersten Iterationen im Fall approximativer Detailkorrekturen machen wir dementsprechend Instabilitäten verantwortlich, die im Zusammenhang mit den unvollständigen Faktorisierungen der Blöcke A_k bei sehr großen Konvektionsstärken und kleinen Gitterweiten auftreten. Nach unserer Erfahrung treten solche Instabilitäten der ILU-Zerlegung insbesondere im Fall Prewavelet-basierter Zerlegungen auf, bei denen eindimensionale Detailfilter der Länge fünf verwendet werden. Dies erklärt die schlechten Werte und den Sprung der Konvergenzraten

$\omega \backslash\, b$	10^3	10^4	10^5	10^6	10^8	10^3	10^4	10^5	10^6	10^8
	$h_0 = 1/256$									
0	0.18	0.27	0.31	0.32	0.32	0.17	0.26	0.31	0.33	0.33
$\pi/8$	0.21	0.34	0.33	0.34	0.34	0.12	0.20	0.21	0.22	0.21
$\pi/4$	0.30	0.36	0.37	0.37	0.37	0.16	0.17	0.19	0.19	0.19
$3\pi/8$	0.20	0.34	0.35	0.35	0.35	0.20	0.17	0.24	0.25	0.25
$\pi/2$	0.17	0.27	0.31	0.32	0.32	0.17	0.26	0.31	0.33	0.33
$5\pi/8$	0.19	0.29	0.32	0.32	0.31	0.09	0.15	0.17	0.22	0.21
$6\pi/8$	0.31	0.36	0.37	0.38	0.38	0.15	0.16	0.17	0.17	0.17
$7\pi/8$	0.22	0.33	0.33	0.33	0.34	0.16	0.21	0.21	0.22	0.22
π	0.18	0.27	0.31	0.32	0.32	0.17	0.26	0.31	0.33	0.33
	$h_0 = 1/512$									
0	0.18	0.27	0.33	0.35	0.35	0.17	0.26	0.32	0.35	0.36
$\pi/8$	0.32	0.50	0.63	0.56	0.58	0.14	0.50	0.79	0.80	0.79
$\pi/4$	0.32	0.43	0.45	0.45	0.45	0.16	0.19	0.20	0.20	0.20
$3\pi/8$	0.30	0.39	0.63	0.69	0.65	0.18	0.37	0.38	0.37	0.37
$\pi/2$	0.18	0.27	0.33	0.35	0.35	0.17	0.26	0.32	0.35	0.36
$5\pi/8$	0.29	0.46	0.76	0.79	0.81	0.14	0.31	0.31	0.32	0.32
$6\pi/8$	0.33	0.42	0.44	0.45	0.45	0.17	0.26	0.25	0.25	0.25
$7\pi/8$	0.32	0.42	0.72	0.69	0.62	0.15	0.41	0.39	0.41	0.40
π	0.18	0.28	0.33	0.35	0.35	0.18	0.27	0.33	0.35	0.36
	$h_0 = 1/1024$									
0	0.18	0.27	0.33	0.36	0.38	0.17	0.26	0.33	0.36	0.38
$\pi/8$	0.31	0.59	0.62	0.63	0.63	0.18	†	†	†	†
$\pi/4$	0.34	0.46	0.52	0.55	0.55	0.16	0.27	0.28	0.28	0.28
$3\pi/8$	0.30	0.77	0.94	†	0.97	0.15	0.46	†	0.88	0.87
$\pi/2$	0.18	0.27	0.33	0.36	0.38	0.17	0.26	0.33	0.36	0.38
$5\pi/8$	0.28	0.88	†	†	†	0.13	0.42	0.62	0.71	0.72
$6\pi/8$	0.38	0.50	0.51	0.51	0.51	0.19	0.20	0.23	0.24	0.24
$7\pi/8$	0.31	0.79	0.95	0.94	0.95	0.15	0.61	†	†	†
π	0.18	0.28	0.34	0.36	0.38	0.18	0.27	0.33	0.37	0.38

$P^k_{k+1,\mathcal{V}}$: matrixabhängig $\quad Q^k_{k+1,\mathcal{V}}$: Pre-Prewavelet Filter
$P^k_{k+1,\mathcal{S}}$: bilinear $\quad Q^k_{k+1,\mathcal{S}}$: hierarchischer Filter [1]

TABELLE 6.16: *Beispiel T^γ: Gewöhnlicher (erste fünf Spalten) und verbesserter Multiskalen-V(1,1)-Zyklus (letzte fünf Spalten), $\varrho_{it,it-4}$.*

für $h_0 = \frac{1}{1024}$ in der letzten Zeile von Tabelle 6.10. Auch die Multiskalen-W(1,1)-Zyklen bleiben von dem beobachteten Phänomen nicht ganz verschont, wie das untere rechte Bild von Abbildung 6.2 zeigt. Für $h_0 = \frac{1}{1024}$ springt die Fehlernorm nach dem ersten Zyklus von 1 auf etwa 54, um dann fast parallel zu den entsprechenden Werten für $h_0 = \frac{1}{512}$ monoton zu fallen. Man beobachtet dies im gesamten hochkonvektiven Bereich.

6.4 Testbeispiel T^γ: Winkelabhängige Konvektion

Um die Abhängigkeit der Multiskalen-Verfahren bei konstantem Konvektionsfeld vom Winkel der Konvektionsrichtung zu prüfen, untersuchen wir nun die Klasse der Testbeispiele mit Operator T^γ. Die nachfolgenden Tabellen zeigen jeweils für die drei kleinsten betrachteten Gitterweiten $h_0 = \frac{1}{256}$, $h_0 = \frac{1}{512}$ und $h_0 = \frac{1}{1024}$ die geometrischen Mittel $\varrho_{it,it-4}$ der Fehlerreduktionsraten in Abhängigkeit des Störungsparameters b und des Winkels ω, den die Konvektionsrichtung relativ zur x-Achse einnimmt. Es wird hierbei nur über die letzten fünf Iterationen gemittelt, um frei von den bereits im Fall der Diagonalkonvektion beobachteten Anfangseffekten zu sein und damit einen Eindruck vom asymptotischen Verhalten der Verfahren zu geben.

$\omega \setminus b$	0.0	10^1	10^2	10^3	10^4	10^5	10^6	10^8
	$h_0 = 1/256$							
0	0.09	0.09	0.09	0.09	0.09	0.04	0.01	0.01
$\pi/8$	0.09	0.09	0.09	0.13	0.24	0.26	0.26	0.26
$\pi/4$	0.09	0.09	0.09	0.15	0.30	0.33	0.33	0.33
$3\pi/8$	0.09	0.09	0.09	0.13	0.23	0.25	0.25	0.25
$\pi/2$	0.09	0.09	0.09	0.08	0.08	0.03	0.01	0.01
$5\pi/8$	0.09	0.09	0.09	0.13	0.23	0.25	0.25	0.25
$6\pi/8$	0.09	0.09	0.09	0.15	0.30	0.33	0.33	0.34
$7\pi/8$	0.09	0.09	0.09	0.13	0.23	0.25	0.25	0.25
π	0.09	0.09	0.09	0.09	0.08	0.03	0.01	0.01
	$h_0 = 1/512$							
0	0.09	0.09	0.09	0.09	0.09	0.07	0.02	0.01
$\pi/8$	0.09	0.09	0.09	0.11	0.19	0.22	0.23	0.23
$\pi/4$	0.09	0.09	0.09	0.11	0.28	0.33	0.33	0.33
$3\pi/8$	0.09	0.09	0.09	0.10	0.21	0.24	0.25	0.25
$\pi/2$	0.09	0.09	0.09	0.08	0.09	0.06	0.02	0.01
$5\pi/8$	0.09	0.09	0.09	0.10	0.21	0.22	0.22	0.22
$6\pi/8$	0.09	0.09	0.09	0.11	0.28	0.33	0.34	0.34
$7\pi/8$	0.09	0.09	0.09	0.10	0.20	0.18	0.19	0.19
π	0.09	0.09	0.09	0.09	0.08	0.06	0.02	0.02
	$h_0 = 1/1024$							
0	0.09	0.09	0.09	0.09	0.09	0.11	0.05	0.01
$\pi/8$	0.09	0.09	0.09	0.09	0.18	0.19	0.22	0.21
$\pi/4$	0.09	0.09	0.09	0.10	0.23	0.31	0.32	0.32
$3\pi/8$	0.09	0.09	0.09	0.09	0.18	0.23	0.24	0.24
$\pi/2$	0.09	0.09	0.09	0.08	0.09	0.11	0.05	0.01
$5\pi/8$	0.09	0.09	0.09	0.09	0.19	0.24	0.25	0.25
$6\pi/8$	0.09	0.09	0.09	0.09	0.24	0.32	0.33	0.34
$7\pi/8$	0.09	0.09	0.09	0.09	0.18	0.19	0.17	0.18
π	0.09	0.09	0.09	0.09	0.09	0.10	0.04	0.01

$P^k_{k+1,\mathcal{V}}$: matrixabhängig $\quad Q^k_{k+1,\mathcal{V}}$: Pre-Prewavelet Filter
$P^k_{k+1,\mathcal{S}}$: bilinear $\quad Q^k_{k+1,\mathcal{S}}$: hierarchischer Filter [1]

TABELLE 6.17: *Beispiel T^γ: Gewöhnlicher Multiskalen-W(1,1)-Zyklus, $\varrho_{it,it-4}$.*

6.4.1 Hierarchische Zerlegungen des Ansatzraums

Man erhält für sämtliche betrachteten Winkel mit unseren Multiskalen-Zyklen im Fall einer matrixabhängigen hierarchischen Wahl der Ansatzraum-Zerlegungen kombiniert mit klassischen hierarchischen Transformationen auf der Testseite ähnlich schlechte Resultate wie bereits für die gezeigten Beispiele achsenorientierter und diagonalgerichteter Konvektion. Wir verzichten deshalb auf die explizite Wiedergabe von gemittelten Reduktionsraten.

6.4.2 Wavelet-artige Zerlegungen des Ansatzraums

Im Fall Wavelet-artiger Zerlegungen des Ansatzraums betrachten wir zunächst einen gewöhnlichen Multiskalen-V(1,1)-Zyklus und verwenden den Pre-Prewavelet Filter$[-1 \quad \underline{2} \quad -1]/2$ und seine zugehörigen Randmodifikationen zusammen mit matrixabhängig bestimmten eindimensionalen Skalierungsfiltern zur Konstruktion der Matrizen $Q^k_{k+1,\mathcal{V}}$. Es handelt sich dabei also um die von uns bereits beim Beispiel diagonalgerichteter Konvektion bevorzugte Zerlegung. Im schwachkonvektiven Bereich $b \leq 10^2$ sind die durchschnittlichen Reduktionsraten $\varrho_{it,it-4}$ gleichmäßig für alle Winkel durch 0.2 nach oben hin beschränkt. Die ersten fünf Spalten von Tabelle 6.16 zeigen die gemittelten Reduktionsraten im Fall großer Konvektionsstärke ($b = 10^3, ..., 10^6, 10^8$). Man erkennt anhand der ersten Spalte von Tabelle 6.16 für den Wert

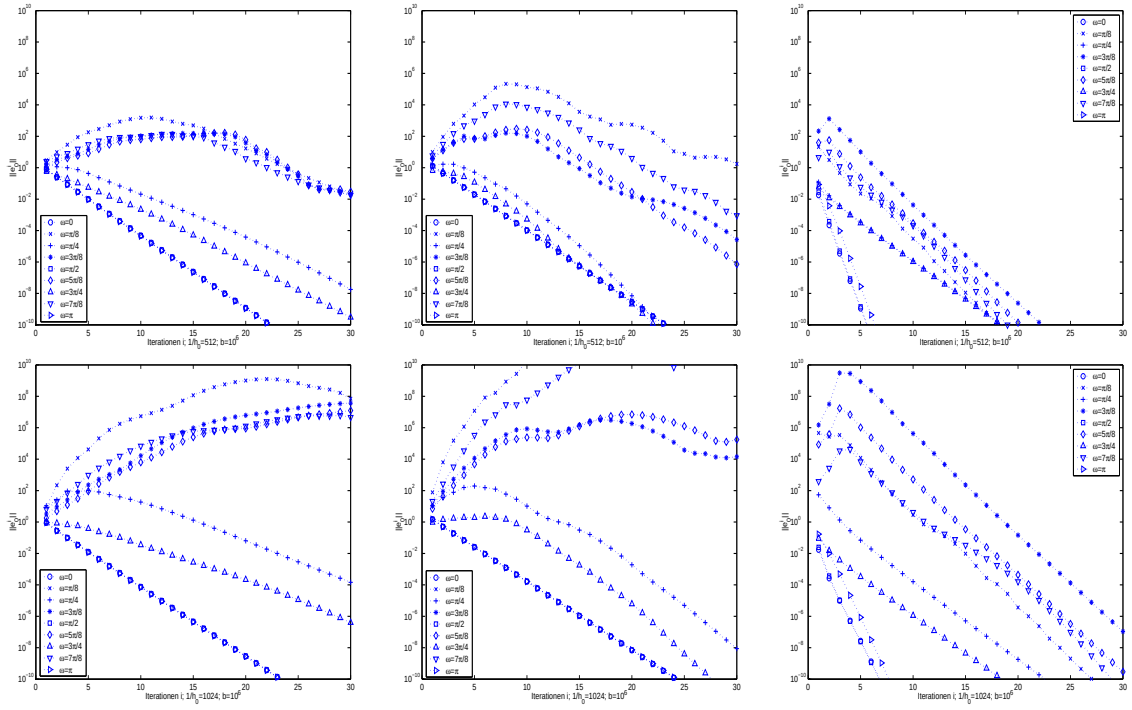


ABBILDUNG 6.3: *Beispiel T^{γ}, Entwicklung der Fehlernormen für unterschiedliche Winkel und $b = 10^6$ zu den Maschenweiten $h_0 = \frac{1}{512}$ (obere Bilder) und $h_0 = \frac{1}{1024}$ (untere Bilder), jeweils klassische hierarchische Zerlegung des Testraums, Pre-Prewavelet-basierte Zerlegung des Ansatzraums zusammen mit matrixabhängiger Prolongation; gewöhnliche Multiskalen-V(1,1)-Zyklen (links), verbesserte Multiskalen-V(1,1)-Zyklen (mitte), gewöhnliche Multiskalen-W(1,1)-Zyklen (rechts).*

$b = 10^3$ noch deutlich die Stabilisierung hinsichtlich der Abhängigkeit von der Problemgröße. Man beobachtet allerdings auch deutliche Konvergenzprobleme bei der minimalen Gitterweite $h_0 = \frac{1}{1024}$ und nichtdiagonaler oder nichtachsenorientierter Konvektion, das heißt für die Winkel $\omega = \frac{\pi}{4}$, $\omega = \frac{3\pi}{8}$, $\omega = \frac{5\pi}{8}$ und $\omega = \frac{7\pi}{8}$. Dies belegen auch die für den Störungsparameter $b = 10^6$ und die Gitterweiten $h_0 = \frac{1}{512}$ und $h_0 = \frac{1}{1024}$ in den linken Bildern von Abbildung 6.3 gezeichneten Entwicklungen der zugehörigen Fehlernormen.

Im schwachkonvektiven Bereich $b \leq 10^2$ sind für den entsprechenden verbesserten Multiskalen-V(1,1)-Zyklus die durchschnittlichen Reduktionsraten $\varrho_{it,it-4}$ gleichmäßig durch 0.15 nach oben beschränkt. Wir verzichten wiederum auf die explizite Wiedergabe der zugehörigen Raten und zeigen in den letzten fünf Spalten von Tabelle 6.16 die gemittelten Reduktionsraten im Fall großer Konvektionsstärke ($b = 10^3, ..., 10^6, 10^8$). Der verbesserte Multiskalen-V(1,1)-Zyklus zeigt zunächst bis auf den Winkel $\omega = \frac{\pi}{8}$ bei der Gitterweite $h_0 = \frac{1}{512}$ recht gute Resultate in Bezug auf Robustheit. Für $h_0 = \frac{1}{1024}$ ergeben sich jedoch für sämtliche nichtdiagonalen und nichtachsenorientierten Fälle ebenfalls große Schwierigkeiten. Um einen genaueren Eindruck im hochkonvektiven Bereich zu erhalten, zeichnen wir daher die Entwicklungen der Fehlernormen für den Störungsparameter $b = 10^6$ und die Gitterweiten $h_0 = \frac{1}{512}$ sowie $h_0 = \frac{1}{1024}$. Man sieht im Vergleich der mittleren Bildern von Abbildung 6.3 daß das Verfahren bei nichtdiagonaler oder nichtachsenorientierter Konvektion für $h_0 = \frac{1}{1024}$ tatsächlich divergiert.

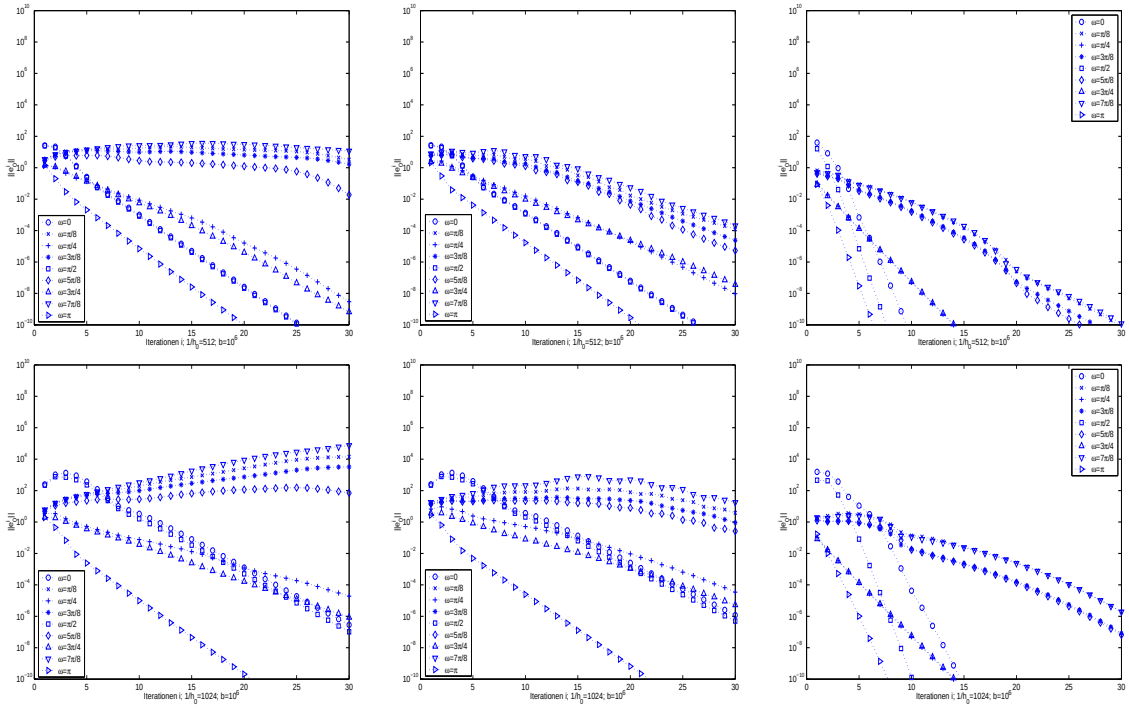


ABBILDUNG 6.4: *Beispiel T^γ, Entwicklung der Fehlernormen für unterschiedliche Winkel und $b = 10^6$ zu den Maschenweiten $h_0 = \frac{1}{512}$ (obere Bilder) und $h_0 = \frac{1}{1024}$ (untere Bilder), jeweils klassische hierarchische Zerlegung des Testraums, Pre-Prewavelet-basierte Zerlegung des Ansatzraums zusammen mit matrixabhängiger Prolongation nach de Zeeuw; gewöhnliche Multiskalen-V(1,1)-Zyklen (links), verbesserte Multiskalen-V(1,1)-Zyklen (mitte), gewöhnliche Multiskalen-W(1,1)-Zyklen (rechts).*

Der verbesserte Multiskalen-V(1,1)-Zyklus wird für fast alle betrachteten Winkel vom gewöhnlichen Multiskalen-W(1,1)-Zyklus geschlagen, der zu einem robusten Verfahren führt, wie man anhand von Tabelle 6.17 erkennt. Das verdeutlichen auch die rechten Graphen von Abbildung 6.3. Die achsenorientierten Fällen stechen hier durch ihr extrem gutes Verhalten etwas heraus. Man erkennt in Abbildung 6.3 allerdings auch, daß für die nichtachsenorientierten Fälle die Fehlernormen während der ersten paar Iterationen zunächst stark ansteigen. Dies liegt wieder an Instabilitäten, die im Zusammenhang mit den unvollständigen Faktorisierungen der Blöcke A_k bei sehr großen Konvektionsstärken auftreten. Sie erscheinen besonders deutlich im Fall nichtdiagonaler und nichtachsenorientierter Konvektion.

Es wurden von uns sämtliche Versuche auch mit solchen Zerlegungen des Ansatzraums durchgeführt, die die matrixabhängige Prolongation nach de Zeeuw benutzen [133]. Man erhält damit qualitativ ähnliche Ergebnisse wie mit den bisher verwendeten matrixabhängigen Prolongationen. Trotz der ausgefeilteren Technik sind die absoluten Raten, die man mittels der Prolongationen de Zeeuws erhält, nicht notwendig besser. Wir verzichten hier auf die Angabe weiterer Tabellen und zeigen lediglich in Abbildung 6.4 die den Kurven aus Abbildung 6.3 entsprechenden Graphen.

6.5 Testbeispiel T^δ: Kreiskonvektion

Aufgrund der einschränkenden Mehrgitter-Interpretation von Multiskalen-Zyklen erscheint es aussichtslos, wirbelbehaftete Strömungen und die daraus abgeleiteten Konvektions-Diffusions Probleme mit den bisherigen Tensorprodukt-Ansätzen erfolgreich angehen zu können.

Im Fall geometrischer Vergröberung benutzen robuste Mehrgitter-Verfahren meist speziell konstruierte Glättungsiterationen, die im Grenzfall entartender Konvektion auf jedem Level zu einem direkten Löser entarten [62]. Man erreicht dies etwa durch Anordnungsverfahren, die die Unbekannten so ordnen, daß sie dem Verlauf der Strömung folgen. Ein gewöhnlicher Gauß-Seidel-Glätter erfüllt dann die eben genannte Robustheitsbedingung [17, 79, 94]. Neben der Notwendigkeit stabiler Grobgitteroperatoren spielt der Glättungsaspekt mit zunehmender Kompliziertheit der Strömungen eine zunehmend wichtige Rolle. Es ist klar, daß die zu Multiskalen-Zyklen gehörenden Glätter das oben genannte Robustheitskriterium nicht erfüllen. Dies ist ein prinzipieller Nachteil unseres Zugangs. So führten sämtliche Versuche, Testbeispiel T^δ auf der Basis geometrisch definierter Multiskalen-Zerlegungen mit matrixabhängigen Multiskalen-Transformationen zu behandeln, zu nichtzufriedenstellenden Ergebnissen. Es sei aber angemerkt, daß auch Anordnungs-basierte geometrische Mehrgitter-Verfahren hierbei auf ernste Schwierigkeiten stoßen. Will man die Freiheitsgrade so ordnen, daß sie dem Strömungsverlauf folgen, so treten sogenannte *Zyklen* von Unbekannten auf, die „aufgeschnitten“ werden müssen. Auch *künstlichen Zyklen*, die alleine aufgrund der Anordnungsalgorithmen entstehen und nicht dem physikalischen Verlauf der Strömung entsprechen, können zu Problemen führen. In drei Raumdimensionen ist die Suche geeigneter Anordnungstechniken Gegenstand aktueller Forschung und bislang außerhalb für die Praxis relevanter Reichweite [17, 76, 79, 94]. Da das algebraische Mehrgitter-Konzept bewußt auf speziell definierte Glättungsprozeduren verzichtet, stellt sich die Frage, ob im Fall einer Kreisströmung mit Hilfe AMG-basierter Zerlegungen und Komponenten ein robuster Multiskalen-Löser in unserem Sinne gefunden werden kann.

Aufgrund der nichtgeometrischen Auswahl der Grobgitterpunkte von AMG sind Verfahren, die eine Zyklustiefe μ größer als Eins besitzen, oft nicht mehr mit linearem Aufwand durchführbar. Wir betrachten daher im Zusammenhang mit algebraischen Multiskalen-Zerlegungen ausschließlich Multiskalen-V(1,1)-Zyklen. In allen Experimenten verwenden wir für die Transformationen $P^k_{k+1,\mathcal{V}}$ und $P^k_{k+1,\mathcal{S}}$ jeweils die gleichen mittels AMG$(\alpha, \beta, \gamma_P)$ berechneten Prolongationen. Die Komplementräume auf der Testseite werden stets hierarchisch gewählt, so daß man insgesamt eine AMG-hierarchische Basis Transformation für die Testseite erhält. Die Detailkorrekturen werden standardmäßig über unvollständige Faktorisierungen der Blöcke A_k berechnet, die sich nach Transformation ergeben.

6.5.1 Hierarchische Zerlegungen des Ansatzraums

Tabelle 6.18 zeigt die Ergebnisse, die man erhält, wenn die Transformation der Testseite ebenfalls AMG-hierarchisch gewählt wird. Die AMG-Prolongationen werden dabei insbesondere nicht abgeschnitten. Betrachtet man die erste Spalte der Tabelle, die das Verhalten im konvektionsfreien Fall beschreibt, so erkennt man ein für hierarchische Zerlegungen typisches Ansteigen der gemittelten Fehlerreduktionsraten. Sie liegen jedoch etwas niedriger als beim zugehörigen klassischen hierarchische Basis Verfahren (siehe z.B. die erste Spalte von Tabelle 6.1). Ein Grund hierfür ist, daß wir als AMG-Prolongation eine besonders „kräftige“ Variante gewählt haben, die sogenannte *Standardprolongation* [112, 117]. Ein weiterer Grund

$h_0^{-1} \setminus b$	0.0	10^1	10^2	10^3	10^4	10^5	10^6	10^8
8	0.02	0.03	0.14	0.06	0.02	0.01	0.01	0.01
16	0.08	0.09	0.18	0.23	0.16	0.14	0.17	0.16
32	0.17	0.17	0.26	0.30	0.28	0.22	0.24	0.25
64	0.21	0.30	0.46	0.52	0.37	0.33	0.29	0.28
128	0.35	0.33	0.43	0.46	0.46	0.32	0.40	0.37
256	0.37	0.42	0.51	0.46	0.52	0.43	0.43	0.49
512	0.45	0.41	0.53	0.51	0.59	0.51	0.46	0.60
1024	0.48	0.52	0.57	0.61	0.59	0.63	0.61	0.66

$P^k_{k+1,\mathcal{V}}$: AMG$(0.1, 0.35, 0.0)$ $Q^k_{k+1,\mathcal{V}}$: hierarchischer Filter [1]
$P^k_{k+1,\mathcal{S}}$: $P^k_{k+1,\mathcal{V}}$ $Q^k_{k+1,\mathcal{S}}$: hierarchischer Filter [1]

TABELLE 6.18: *Beispiel T^δ: Gewöhnlicher Multiskalen-V(1,1)-Zyklus, $\varrho_{it,1}$.*

$h_0^{-1} \setminus b$	0.0	10^1	10^2	10^3	10^4	10^5	10^6	10^8
8	0.03	0.04	0.06	0.03	0.01	0.01	0.01	0.01
16	0.07	0.07	0.08	0.07	0.10	0.05	0.05	0.05
32	0.10	0.09	0.10	0.12	0.09	0.10	0.10	0.10
64	0.11	0.09	0.10	0.18	0.11	0.10	0.10	0.11
128	0.10	0.10	0.12	0.16	0.14	0.11	0.13	0.12
256	0.13	0.11	0.13	0.15	0.16	0.14	0.14	0.12
512	0.12	0.13	0.12	0.14	0.18	0.18	0.14	0.14
1024	0.13	0.12	0.11	0.25	0.32	0.25	0.17	0.22

$P^k_{k+1,\mathcal{V}}$: AMG$(0.1, 0.35, 0.0)$ $Q^k_{k+1,\mathcal{V}}$: $[(T_{k-1,(i,j)})_*]_{0.125}$
$P^k_{k+1,\mathcal{S}}$: $P^k_{k+1,\mathcal{V}}$ $Q^k_{k+1,\mathcal{S}}$: hierarchischer Filter [1]

TABELLE 6.19: *Beispiel T^δ: Gewöhnlicher Multiskalen-V(1,1)-Zyklus, $\varrho_{it,1}$.*

$h_0^{-1} \setminus b$	0.0	10^1	10^2	10^3	10^4	10^5	10^6	10^8
8	0.04	0.04	0.02	0.02	0.00	0.00	0.00	0.00
16	0.05	0.05	0.04	0.05	0.03	0.03	0.03	0.03
32	0.08	0.07	0.06	0.09	0.05	0.04	0.04	0.04
64	0.09	0.09	0.09	0.12	0.06	0.11	0.05	0.05
128	0.12	0.11	0.11	0.25	0.09	0.08	0.08	0.07
256	0.14	0.14	0.07	0.16	0.12	0.09	0.10	0.10
512	0.15	0.15	0.09	0.22	0.21	0.10	0.10	0.11
1024	0.16	0.17	0.11	0.18	0.38	0.14	0.13	0.12

TABELLE 6.20: *Beispiel T^δ: AMG$(0.1, 0.35, 0.0)$ basierter Mehrgitter-V(1,1)-Zyklus, $\varrho_{it,1}$, Standardprolongation, punktweise Gauß–Seidel Glättung, bei der zuerst über die Fein-ohne-Grobgitterfreiheitsgrade geglättet wird; abgebrochen wurde, wenn die $\mathcal{L}_2(\Omega)$-Norm der Residuen 10^{-10} unterschreitet.*

für das insgesamt gute Abschneiden des Verfahrens ist die relativ kleine Wahl von $\alpha = 0.1$. Man erkennt, daß das Lösungsverhalten nicht robust ist. Würde das Verfahren lediglich als Multiskalen-Vorkonditionierer für eine äußere Krylovraum-Iteration eingesetzt, so würde die Gitterweitenabhängigkeit des resultierenden Verfahrens natürlich abgemildert werden [98].

6.5.2 Wavelet-artige Zerlegungen des Ansatzraums

Betrachten wir nun in Tabelle 6.19 die entsprechenden Raten, die man mittels einer AMGlet-Zerlegung des Ansatzraums erhält. Der Vergleich der Tabellen 6.18 und 6.19 zeigt deutlich die Stabilisierung der Verfahren bezüglich der Abhängigkeit von der Problemgröße, die man durch die AMGlet-Zerlegung erreicht. Es werden zur Berechnung der Ergebnisse von Tabelle 6.19 $\gamma_Q = 0.125$ und ansonsten die gleichen Parameter wie im Fall von Tabelle 6.18 gewählt.

$h_0^{-1} \setminus b$	0.0	10^1	10^2	10^3	10^4	10^5	10^6	10^8
8	2.93	2.94	3.44	3.18	3.18	3.18	3.18	3.18
16	3.37	3.36	5.38	4.37	4.18	4.17	4.17	4.17
32	3.77	3.78	7.12	5.51	5.12	5.29	5.23	5.23
64	4.11	4.03	5.20	8.88	5.52	5.64	5.62	5.60
128	4.17	4.17	5.05	10.87	6.33	6.20	6.20	6.22
256	4.34	4.32	4.57	11.81	7.59	6.68	6.66	6.66
512	4.35	4.44	4.44	9.78	10.54	7.19	7.13	7.15
1024	4.38	4.41	4.33	6.58	12.15	7.47	7.60	7.83

$P^k_{k+1,\mathcal{V}}$: AMG$(0.1, 0.35, 0.0)$ $Q^k_{k+1,\mathcal{V}}$: hierarchischer Filter [1]
$P^k_{k+1,\mathcal{S}}$: $P^k_{k+1,\mathcal{V}}$ $Q^k_{k+1,\mathcal{S}}$: hierarchischer Filter [1]

TABELLE 6.21: *Beispiel T^δ: Nichtstandard-Operatorkomplexität C^o_{ns}.*

$h_0^{-1} \setminus b$	0.0	10^1	10^2	10^3	10^4	10^5	10^6	10^8
8	3.94	3.95	4.11	3.72	3.72	3.72	3.72	3.72
16	4.94	4.92	6.51	5.09	4.93	4.91	4.91	4.91
32	5.66	5.67	8.66	6.53	5.98	6.16	6.11	6.11
64	6.22	6.15	7.09	10.24	6.56	6.62	6.62	6.61
128	6.48	6.46	7.14	12.62	7.43	7.24	7.23	7.24
256	6.73	6.70	6.80	13.80	8.84	7.76	7.76	7.74
512	6.79	6.86	6.80	11.91	12.11	8.31	8.24	8.25
1024	6.84	6.87	6.76	8.77	14.04	8.65	8.73	8.94

$P^k_{k+1,\mathcal{V}}$: AMG$(0.1, 0.35, 0.0)$ $Q^k_{k+1,\mathcal{V}}$: $[(T_{k-1,(i,j)})_*]_{0.125}$
$P^k_{k+1,\mathcal{S}}$: $P^k_{k+1,\mathcal{V}}$ $Q^k_{k+1,\mathcal{S}}$: hierarchischer Filter [1]

TABELLE 6.22: *Beispiel T^δ: Nichtstandard-Operatorkomplexität C^o_{ns}.*

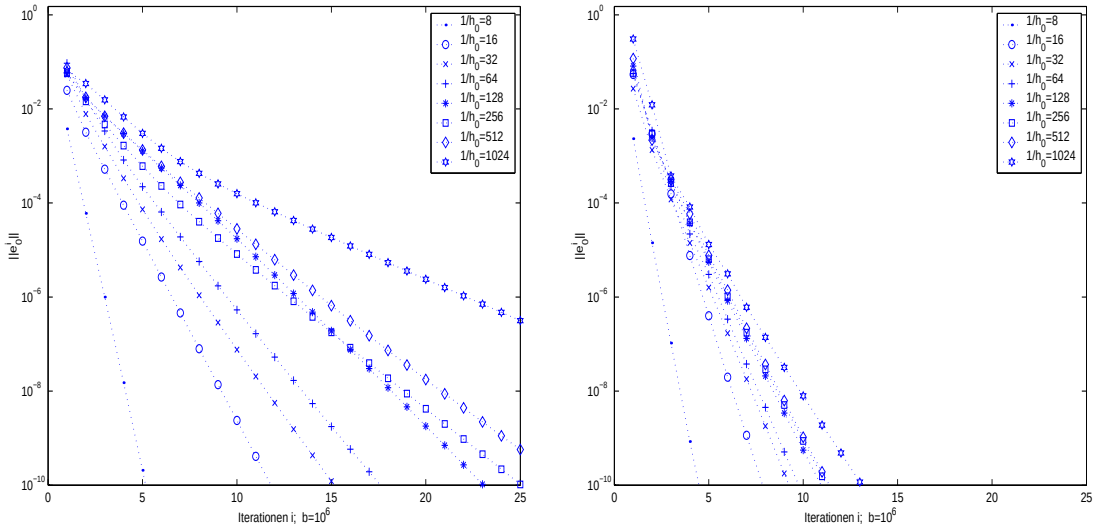


ABBILDUNG 6.5: *Beispiel T^δ: Entwicklung der Fehlernormen für gewöhnliche Multiskalen-V(1,1)-Zyklen, unterschiedliche Maschenweiten h_0 und feste Konvektionsstärke $b = 10^6$, AMG$(0.1, 0.35, 0.0)$ hierarchische Zerlegung (links) und AMGlet stabilisierte Zerlegung (rechts).*

Sowohl das Verwenden der Standardprolongation als auch der kleine Wert $\alpha = 0.1$ sind wiederum Grund für die hervorragenden Resultate, aus denen wir ein insgesamt robustes Konvergenzverhalten folgern. Für Gitterweiten $h_0 \leq \frac{1}{512}$ sind die gemittelten Fehlerreduktionsraten $\varrho_{it,1}$ gleichmäßig bezüglich des Störungsparameters b durch 0.2 nach oben hin beschränkt. Für $h_0 = \frac{1}{1024}$ beobachtet man ein leichtes Anwachsen der Werte mit einer maximalen Rate

$h_0^{-1} \setminus b$	0.0	10^1	10^2	10^3	10^4	10^5	10^6	10^8
8	1.88 (5)	1.94 (6)	2.20 (6)	2.22 (6)	2.22 (6)	2.22 (6)	2.22 (6)	2.22 (6)
16	1.73 (6)	1.74 (6)	2.13 (9)	2.25 (8)	2.26 (8)	2.26 (8)	2.26 (8)	2.26 (8)
32	1.70 (7)	1.71 (8)	2.02 (9)	2.22 (11)	2.39 (11)	2.36 (10)	2.36 (10)	2.36 (10)
64	1.69 (8)	1.69 (9)	1.81 (13)	2.22 (14)	2.25 (12)	2.31 (13)	2.32 (12)	2.32 (12)
128	1.68 (9)	1.68 (11)	1.74 (15)	2.17 (16)	2.30 (16)	2.34 (15)	2.37 (16)	2.37 (15)
256	1.67 (10)	1.67 (11)	1.69 (15)	2.07 (19)	2.26 (17)	2.32 (17)	2.32 (16)	2.34 (18)
512	1.67 (11)	1.67 (12)	1.67 (16)	1.90 (22)	2.20 (19)	2.33 (21)	2.33 (21)	2.32 (22)
1024	1.67 (13)	1.67 (13)	1.67 (17)	1.75 (23)	2.15 (23)	2.29 (23)	2.32 (23)	2.34 (24)

TABELLE 6.23: *Beispiel T^δ: Gitterkomplexität C^g und Anzahl lt der durch AMG $(0.1, 0.35, 0.0)$ entstehenden Skalen.*

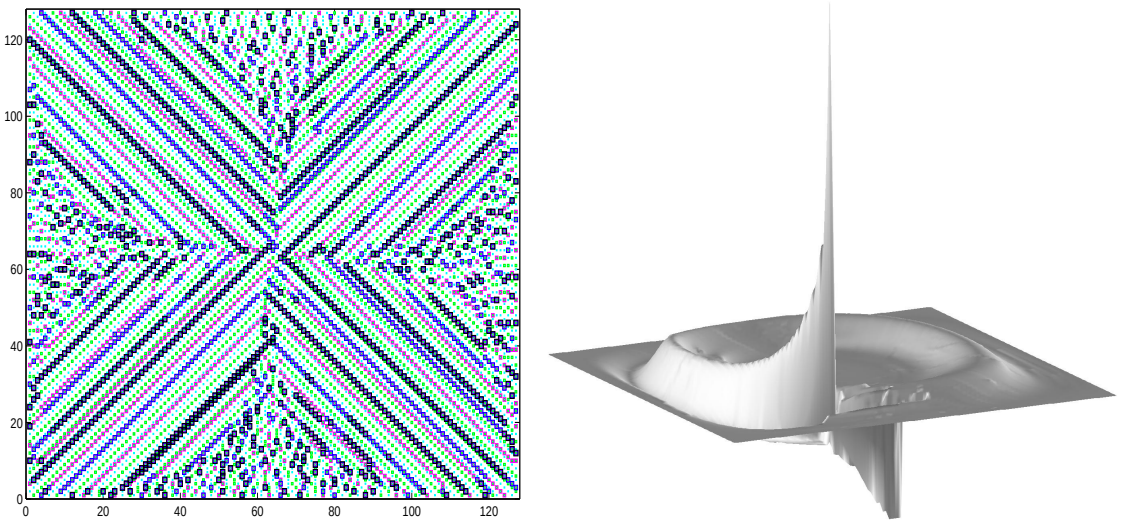


ABBILDUNG 6.6: *Beispiel T^δ: linkes Bild: Veranschaulichung der ersten fünf groben Gitter, die durch AMG$(0.1, 0.35, 0.0)$ ausgewählt werden (gröbere Gitter entsprechen dickeren und dunkleren Punkten), $h_0 = \frac{1}{128}$ und $b = 10^6$; rechtes Bild: zugehöriges AMGlet, das durch die Stabilisierung der AMG$(0.1, 0.35, 0.0)$ hierarchischen Basis Zerlegung der Ansatzseite erzeugt wird.*

von 0.32 im Fall $b = 10^4$. Es handelt sich hierbei aber nicht um eine konzeptionelle Schwierigkeit des Zugangs mittels AMGlets. Ein leichtes Verschlechtern der Raten läßt sich nämlich aus Tabelle 6.20 auch für das entsprechende algebraische Mehrgitter-Verfahren ablesen, das eine punktweise Gauß–Seidel Glättung benutzt. Man erhält hier für $b = 10^4$ sogar die etwas schlechtere Rate $\varrho_{it,1} = 0.38$. Der stabilisierende Einfluß der AMGlet-Zerlegung wird ebenfalls durch die in Abbildung 6.5 gezeichneten Fehlernormen deutlich. Auf der linken Seite sehen wir die Fehlernormen, die sich durch rein AMG-hierarchische Verfahren zu unterschiedlichen Gitterweiten und fester Konvektionsstärke $b = 10^6$ ergeben. Rechts sind die entsprechenden Kurven des stabilisierten Verfahrens abgebildet.

Die Wahl der Basisfunktionen für die Komplementräume hat keinen Einfluß auf den AMG-Vergröberungsprozeß. Man erhält daher sowohl für die AMG-hierarische Zerlegung als auch für die Zerlegung mittels AMGlets die gleichen Gitterkomplexitäten als auch die gleiche Anzahl von Skalen. Beide Größen zeigen ein recht vernünftiges Verhalten, wie man anhand von Tabelle 6.23 erkennt. Vergleicht man die entsprechenden Nichtstandard-Operatorkomplexitäten in den Tabellen 6.21 und 6.22, so findet man, daß diese im schwachkonvektiven Bereich für das Verfahren mit einer AMGlet-Zerlegung der Ansatzseite ungefähr 56% höher

liegen als die entsprechenden Komplexitäten im Fall AMG-hierarchischer Zerlegungen beider Seiten. Für $b > 10^3$ ist der Unterschied viel kleiner und liegt bei circa 16%. Vergleichen wir dies mit den entsprechenden gemittelten Fehlererduktionsraten aus den Tabellen 6.18 und 6.19, so schließen wir, daß der AMGlet-stabilisierte Löser dem AMG-hierarchischen Löser trotz der hohen Kosten vorzuziehen ist.

Abbildung 6.6 zeigt auf der linken Seite die Grobgitterhierarchie, die durch fünf Vergröberungsschritte mittels AMG$(0.1, 0.35, 0.0)$ bei der Gitterweite $h_0 = \frac{1}{128}$ und Konvektionsstärke $b = 10^6$ erzeugt wird. Auf dem rechten Bild sehen wir ein zugehöriges AMGlet. Es ist deutlich zu erkennen, daß die Basisfunktion bestrebt ist, dem Verlauf der Strömung zu folgen. Man beobachtet ein für Wavelets charakteristisches „Durchschwingen“. Die Funktion nimmt positive *und* negative Werte an.

Kapitel 7

Schlußbemerkungen

Die vorliegende Arbeit kann als eine praktisch orientierte Antwort auf die folgende Frage angesehen werden:

> *Inwieweit ist es möglich,* robuste *Multiskalen-Verfahren für Konvektions-Diffusions Probleme zu entwickeln ausgehend von* direkten *Unterraumzerlegungen der zugrundeliegenden Funktionenräume?*

Man kann sie als eine Konkretisierung der in [120] gestellten wichtigen Aufgabe verstehen, die nach der Konstruktion effizienter Multiskalen-Löser für allgemeine elliptische Randwertprobleme auf der Basis direkter Unterraumzerlegungen fragt. Die in [120] vorgestellten Antworten sind nicht ohne weiteres erfolgreich auf das singulär gestörte Problem (1.1) anwendbar sondern führen lediglich im ungestörten Fall wie etwa beim Poissonproblem (1.1) mit $\vec{b} = \mathbf{0}$ zu Verfahren, die (nachweislich) gitterweitenunabhängige Konvergenzeigenschaften ähnlich wie klassische Mehrgitter-Methoden aufweisen.

Auf dem Weg hin zu robusten Multiskalen-Methoden für Konvektions-Diffusions Probleme begegnet man den folgenden drei Hauptschwierigkeiten, die schon zu Beginn in der Einleitung herausgestellt wurden. Es sind dies

- die Abhängigkeit des Verfahrens von der Feingitterweite h_0,
- die Abhängigkeit des Verfahrens von der Konvektion $\vec{b}$,
- die Stabilität der zur Konvergenzbeschleunigung eingesetzten Grobgitterprobleme.

Dementsprechend besteht die von uns gefundene Antwort auf die oben formulierte Frage aus drei Teilen. Als erstes ist die problemabhängige Wahl der Prolongationsoperatoren $P^k_{k+1,\mathcal{V}/\mathcal{S}}$ innerhalb unseres Petrov–Galerkin Multiskalen-Ansatzes unverzichtbar, um physikalisch sinnvolle Grobgitterprobleme zu erhalten. Wir verwendeten dazu entscheidend Techniken zur Konstruktion geeigneter Prolongationen, die bei robusten variationellen Mehrgitter-Verfahren eingesetzt werden. Hierbei konnten die von uns auf der Basis geometrischer Vergröberungen konstruierten problemangepaßten Prolongationen über ihre Interpretation mittels (Tensorprodukt) L-Splines in den Kontext verallgemeinerter Multiskalen-Analysen eingeordnet werden. Damit gelingt es, eine Brücke von der Welt der Mehrgitter- und Multilevel-Verfahren hin zu Wavelet-basierten Methoden zu schlagen. Dies zeigen nicht zuletzt auch unsere Untersuchungen der Standard- und Nichtstandardform von Operatoren sowie deren Einbettung in ein Multiskalen-Erzeugendensystem. Im Fall des herausfordernden Testbeispiels T^δ mit einem

nichtseparablen wirbelbehafteten Konvektionsfeld mußten wir sogar auf Prolongationstechniken zurückgreifen, die von algebraischen Mehrgitter-Verfahren herrühren. Es ist ebenfalls möglich, den Kontext variationeller Vergröberungen zu verlassen und stattdessen mit expliziten Diskretisierungen bezüglich gröberer Gitter zu arbeiten. Man erhält damit jedoch zumeist schlechtere numerische Ergebnisse, was von traditionellen Mehrgitter-Lösern her ebenfalls bekannt ist [62].

Betrachtet man hierarchische Zerlegungen der Feingitter Ansatz- und Testräume $\mathcal{V}_0$ und $\mathcal{S}_0$, so erscheint (für höherdimensionale Probleme) der Wechsel hin zu einer Wavelet-artigen Zerlegung auf einer der beiden Seiten notwendig für die Robustheit unserer Verfahren zu sein. Dies gilt sogar für das im wesentlichen eindimensionale Testbeispiel T^α. Zwar reichen hierarchische W-Zyklen aus, um für das Poissonproblem eine Stabilisierung eines hierarchischen Multiskalen-Lösers bezüglich seiner Gitterweitenabhängigkeit zu erreichen. Im Konvektions-Diffusions Fall erhält man dadurch jedoch keine Stabilisierung des Verfahrens in Bezug auf die Konvektionsstärke und damit keine Robustheit.
Wavelet-artige Multiskalen-Zerlegungen besitzen für das Poissonproblem und einen Galerkin-Ansatz optimale Stabilitätseigenschaften hinsichtlich der Abhängigkeit der Konvergenzeigenschaften resultierender Multiskalen-Löser von h_0. Diese Stabilitätseigenschaften manifestieren sich in einer Normäquivalenz zwischen der zugehörigen Energienorm und einer gewichteten Multiskalennorm, bei der die Konstanten unabhängig von h_0 sind. Dies führt zu optimalen additiven und multiplikativen Multiskalen-Verfahren. Unsere Experimente zeigen, daß für multiplikative Multiskalen-Verfahren basierend auf unterschiedlichen Zerlegungen der Ansatz- und Testseite, eine Wavelet-artige Zerlegung auf einer der beiden Seiten für solche optimalen Löser sogar ausreicht. Aus Sicht der Mehrgitter-Interpretation der Multiskalen-Zyklen ist die im Vergleich zu rein hierarchischen Glättern verbesserte Glättungsarbeit auf den einzelnen Skalen dafür verantwortlich. Die zugehörigen Multiskalen-Glätter erfüllen im Fall singulär gestörter Konvektions-Diffusions Probleme trotzdem nicht die von robusten Mehrgitter-Verfahren bekannte Forderung nach *robusten Glättern*, im Grenzfall sehr starker Konvektion zu exakten Lösern für die einzelnen Level zu entarten. Es handelt sich dabei um einen prinzipiellen Nachteil von Ansätzen, die auf direkten Unterraumzerlegungen beruhen. Dieser kann jedoch auch als Chance begriffen werden, da im Fall dreidimensionaler wirbelbehafteter Konvektions-Diffusions Gleichungen robuste Glättungsiterationen für gewöhnliche Mehrgitter-Zyklen nicht mehr ohne weiteres effizient konstruiert werden können [17, 76, 79, 94]. Unsere positiven Erfahrungen, die wir für das zweidimensionale Beispiel T^δ im Zusammenhang mit algebraischen Multiskalen-Zerlegungen gemacht haben, geben uns die Hoffnung, mit diesen Ansatz auch dreidimensionale wirbelbehaftete Probleme erfolgreich lösen zu können. Hierbei könnte es nützlich sein, effiziente Schur-Komplement Approximationen anstelle der Grobgitteroperatoren in den rekursiven AMG-Prozeß einfließen zu lassen, worauf bislang in unserem Ansatz verzichtet wurde.

Der dritte Grund für den Erfolg der von uns vorgeschlagenen Verfahren ist die Kombination der hierarchischen Zerlegung auf der Testseite mit einer problemangepaßten Vergröberung auf der Ansatzseite. Wir haben gezeigt, daß dies für eindimensionale Probleme äquivalent zu einer Entkopplung des transformierten linearen Gleichungssystems ist. Für höherdimensionale Probleme ist diese Entkopplung nur noch approximativ möglich. Sie scheint dennoch wesentlich für das Erhalten erfolgreicher Verfahren zu sein, wie weitere numerische Experimente zeigen.

Durch das von uns verwendete Petrov–Galerkin Multiskalen-Konzept ist es möglich, alle soeben aufgelisteten Ideen, die eng miteinander verzahnt sind, zu vereinigen. Besonders schön zeigt dies die schematische Abbildung in der Einleitung. Unser Ansatz geht damit weit über die uns bekannten Verfahren, die auf direkten Unterraumzerlegungen beruhen, hinaus, wie der ausführliche Vergleich etwa in Kapitel 4.2.3 mit den Methoden aus [5, 6, 48, 49, 96, 97] verdeutlicht. Die in [48, 49] als direkter Multiskalen-Löser vorgeschlagene Wavelet-basierte LU-Faktorisierung scheitert in Bezug auf singulär gestörte Konvektions-Diffusions Probleme, da sie keine problemangepaßten Multiskalen-Zerlegungen verwendet [107]. Viele andere Wavelet-Methoden zur Lösung von partiellen Differentialgleichungen leiden unter der gleichen Schwierigkeit [29, 30, 49, 92, 93, 100]. Dies trifft auch auf die Vorkonditionierung mittels AMLI zu, die zudem lediglich von hierarchischen Multiskalen-Zerlegungen ausgeht und daher eine W-Zyklus-artige Stabilisierung hinsichtlich h_0 durch innere Rekursionen benötigt. Die verallgemeinerte Hierarchische Basis Mehrgitter-Methode [14], die auf der Verwendung problemangepaßter Prolongationen $P^k_{k+1,\mathcal{V}/\mathcal{S}}$ zusammen mit zugehörigen hierarchischen Zerlegungen beruht, leidet spürbar unter der mehrfach herausgestellten prinzipiellen Schwäche hierarchischer Glätter, was die durchweg schlechten Ergebnisse in [79] auch belegen. Das gilt ebenfalls für die Arbeit [35], die einen adaptiven Mehrgitter-Löser auf der Basis verallgemeinerter hierarchischer Basis Transformationen vorstellt. Obwohl die Methode der Approximativen Zyklischen Reduktion [96] problemangepaßte hierarchische Multiskalen-Zerlegungen auf der Basis algebraischer Vergröberungen verwendet und damit gute Voraussetzungen für ein robustes Verfahren erfüllt, scheinen die durch sie gewonnenen Verfahren noch mit einer Gitterweitenabhängigkeit behaftet zu sein. Mit den von uns entwickelten AMGlet-Zerlegungen, die ebenfalls auf rein algebraischen Prinzipien beruhen, gelingt es, auf einfache Weise hier Abhilfe zu schaffen. Es ist darüberhinaus damit möglich, die für separable Probleme erfolgreichen Tensorprodukt-Ansätze zu verlassen, um schwierige nichtseparable Aufgaben in unter Umständen kompliziert berandeten Gebieten zu behandeln. Dies eröffnet den Übergang von Modellproblemen hin zu praxisnahen Fragestellungen.

Wegen der unterschiedlichen Wechselwirkungen der drei Hauptbausteine unseres Konzepts — *problemangepaßte Vergröberung*, *Wavelet-artige Testraumzerlegung* und *hierarchische Ansatzraumzerlegung* — sowie der prinzipiellen Schwäche zugehöriger Multiskalen-Glätter, ist es unserer Erfahrung nach ungleich schwieriger, robuste Multiskalen-Verfahren zu konstruieren, als etwa mittels spezieller Mehrgitter-Techniken. Durch unseren Ansatz ist es unseres Wissens allerdings erstmalig gelungen, robuste Multiskalen-Methoden für allgemeine Konvektions-Diffusions Probleme auf der Basis direkter Unterraumzerlegungen zu konzipieren und praktisch umzusetzen.

Anhang A

Iterative Verfahren für lineare Gleichungssysteme

Die in Kapitel 4 betrachteten Multiskalen-Verfahren können natürlich auch als Vorkonditionierer für andere Iterationsverfahren wie etwa Krylovraum-Methoden eingesetzt werden. Dies gestattet es, unter Umständen einen Multiskalen-Kontext mit Robustheitseigenschaften des äußeren iterativen Lösers geeignet zu verbinden, so daß man zu insgesamt robusteren Verfahren gelangt [22, 88, 98]. Wir haben davon in unseren numerischen Beispielen in Kapitel 6 abgesehen, da wir die Konvergenzeigenschaften der von uns konstruierten Multiskalen-Löser ohne den möglicherweise glättenden Charakter einer äußeren Iteration untersuchen wollten.

In diesem Anhang beschäftigen wir uns mit einer Konvergenztheorie für additiv Multiskalen-vorkonditionierte GMRES-artige Krylovraum-Verfahren. Wir geben dazu der Darstellung in [125] folgend zunächst eine systematische Herleitung moderner Krylovraum-Löser und ihrer Konvergenzeigenschaften. Ziel ist es, mit Hilfe von Konvergenzabschätzungen, die auf dem *Numerischen Wertebereich* der Steifigkeitsmatrix und ihrer Inversen basieren, die Optimalität additiv Multiskalen-vorkonditionierter GMRES-artiger Krylovraum-Löser in einem etwas allgemeineren Rahmen als in [108, 109] nachzuweisen. Ein Robustheitsbeweis ist mit diesem Zugang jedoch bislang noch nicht gelungen. Obwohl wir an einer Darstellung der Konvergenztheorie im Zusammenhang mit den von uns gewählten Petrov–Galerkin Multiskalen-Zerlegungen interessiert sind, zeigt sich, daß an einigen Stellen die Annahme Galerkin-artiger Multiskalen-Zerlegungen erforderlich ist. Diese werden klar herausgearbeitet.

Es wird zunächst eine sehr allgemeine Klasse von iterativen Verfahren zur Lösung linearer Gleichungssysteme eingeführt, *Orthogonalisierungsverfahren*, über die eine einfache Klassifikation der meisten zur Zeit bekannten iterativen Löser möglich ist [125]. Dies gilt sowohl für die modernen Krylovraum-Löser wie zum Beispiel GMRES, BICGSTAB und ihre zahlreichen Varianten als auch für die ältere konjugierte Gradienten und konjugierte Residuen Methode [104]. Es lassen sich aber auch klassische Iterationsverfahren wie das Jacobi- oder das Gauß–Seidel-Verfahren als Orthogonalisierungsverfahren ansehen. Orthogonalisierungsverfahren bestimmen ausgehend von den bisherigen Iterierten neue Iterierte durch die Forderung, daß die resultierenden Residuen in einem gewissen Sinne senkrecht auf einer Menge von Suchrichtungen stehen, wobei die möglichen Orthogonalisierungsbedingungen mit Hilfe sogenannter *Orthogonalisierungsmatrizen* Z_k schrittabhängig parametrisiert werden können. Nach ihrer Definition leiten wir eine allgemeine Konvergenzabschätzung für die entstehenden Verfahren her, die den Ausgangspunkt für genauere Konvergenzuntersuchungen im Zusammenhang mit

den spezielleren *Krylovraum-Methoden* bildet. Danach betrachten wir Orthogonalisierungsverfahren im Zusammenhang mit schrittabhängig vorkonditionierten Gleichungssystemen und zeigen, wie diese ausgehend vom ursprünglichen System mittels transformierter Suchrichtungen und Orthogonalisierungsbedingungen interpretiert werden können.

Im zweiten Teil dieses Anhangs leiten wir einige derzeit gebräuchliche Krylovraum-Methoden mit Hilfe speziell gewählter Orthogonalisierungsmatrizen anhand von Orthogonalisierungsverfahren her. Uns interessieren dabei weniger Aspekte ihrer Implementierungen als vielmehr die Frage nach den Eigenschaften der unterschiedlichen Verfahren und mögliche Konvergenzabschätzungen. Neben den in [125] gezeigten Konvergenzaussagen, die auf dem Spektrum der zum linearen Gleichungssystem gehörenden Systemmatrix beruhen, zeigen wir auch Abschätzungen auf der Basis des Numerischen Wertebereichs des diskreten Operators [72]. Letzteres tun wir insbesondere im Hinblick auf die bei der Diskretisierung von Konvektions-Diffusions Gleichungen entstehenden Matrizen. Diese sind im allgemeinen nichtnormal und die Konvergenz iterativer Löser ist für solche Matrizen bekanntlich nicht mehr alleine durch ihr Spektrum bestimmt [53, 115, 116].

Im dritten Teil untersuchen wir explizit das Verhältnis zwischen Krylovraum-Verfahren für das Standardsystem, das heißt für das Multiskalen-transformierte lineare Gleichungssystem, und dem additiv Multiskalen-vorkonditionierten Gleichungssystem. Es zeigt sich, daß man dadurch einander äquivalente Verfahren erhält. Dieser Abschnitt kann jedoch beim ersten Lesen übergangen werden.

Nach diesen Vorarbeiten untersuchen wir in einem abschließenden vierten Teil die Konvergenzeigenschaften additiv Multiskalen-vorkonditionierter Krylovraum-Verfahren für Konvektions-Diffusions Probleme. Auf der Basis verallgemeinerter Normäquivalenzen zeigen wir, daß die Konvergenzraten additiv Multiskalen-vorkonitionierter Krylovraum-Verfahren, die eine durch den Multiskalen-Vorkonditionierer induzierte Norm des Residuums minimieren, unabhängig von der Anzahl lt der auftretenden Skalen sind.

A.1 Orthogonalisierungsverfahren

A.1.1 Definition von Orthogonalisierungsverfahren

Wir definieren *Orthogonalisierungsverfahren* zur iterativen Lösung eines linearen Gleichungssystems

$$Ax = b. \tag{1.1}$$

Es wird vorausgesetzt, daß $A \in \mathbb{R}^{n\times n}$ eine reguläre reellwertige Matrix der Dimension n ist und daß x und b Vektoren des $\mathbb{R}^n$ sind. Im gesamten Anhang benutzen wir die Konvention, daß wenn $x_k \in \mathbb{R}^n$ eine Iterierte ist, $r_k = Ax_k - b$ das zugehörige Residuum bezeichnet.

Definition 4 (Orthogonalisierungsverfahren)
Es seien x_0 ein vorgegebener Startvektor und $r_0 = Ax_0 - b$ das zugehörige Startresiduum. Ein Orthogonalisierungsverfahren zur iterativen Lösung des linearen Gleichungssystems (1.1) besteht aus einer Rekursion

$$x_k \in \bar{x}_k + \langle q_{k-\sigma_k,k}, ..., q_{k-1,k}\rangle \quad (k \geq 1)$$

mit $\sigma_k \leq k$, Suchrichtungen $q_{k-j,k} \in \mathbb{R}^n$ und $\bar{x}_k \in \langle x_{k-\sigma_k}, ..., x_{k-1}\rangle$, so daß die neue

Approximation x_k den Orthogonalitätsrelationen

$$r_k^t Z_k q_{k-j,k} = 0 \quad (j = 1, ..., \sigma_k) \tag{1.2}$$

genügt. Die Matrizen $Z_k \in \mathbb{R}^{n\times n}$ werden dabei als nichtsingulär vorausgesetzt und können schrittabhängig gewählt werden. Der zweite Index k der Suchrichtungen $q_{k-j,k}$ zeigt an, daß diese ebenfalls schrittabhängig sein können. Das Verfahren heißt

- exakt*, falls im $k-$ten Iterationsschritt auch k Suchrichtungen zur Berechnung der neuen Iterierten verwendet werden, das heißt $\sigma_k = k$,*
- abgeschnitten*, falls höchstens σ_{max} Suchrichtungen zur Berechnung der neuen Iterierten verwendet werden, das heißt $\sigma_k = \min(k, \sigma_{max})$.*

Neben exakten und abgeschnittenen Orthogonalisierungsverfahren sind selbstverständlich weitere Varianten möglich. Orthogonalisierungsverfahren unterscheiden sich durch die Wahl der Größen $q_{k-j,k}$, Z_k, σ_k und $\bar{x}_k$. Man kann die Orthogonalitätsrelationen (1.2) für das Residuum r_k der neuen Iterierten x_k auch ausgehend von einer schwachen Formulierung für das Verschwinden des Residuums interpretieren. Durch die Bedingungen (1.2) wird gefordert, daß jeweils die Projektion des neuen Residuums r_k auf den Raum, der von den Vektoren $Z_k q_{k-j,k}$ $(j = 1, ..., \sigma_k)$ aufgespannt wird, gleich dem Nullvektor $\mathbf{0}$ ist. Es seien jetzt

$$\begin{aligned} \bar{r}_k &= A\bar{x}_k - b, \\ \bar{e}_k &= \bar{x}_k - x. \end{aligned}$$

Über die Darstellung der Iterierten im Schritt k mit Hilfe der Suchrichtungen und geeigneter reeller Koeffizienten $\gamma_{i,k}$

$$x_k = \sum_{i=1}^{\sigma_k} \gamma_{i,k} q_{k-i,k} + \bar{x}_k$$

erhalten wir als Darstellungen für das entsprechende Residuum und den Fehler

$$r_k = \sum_{i=1}^{\sigma_k} \gamma_{i,k} A q_{k-i,k} + \bar{r}_k, \tag{1.3}$$

$$e_k = \sum_{i=1}^{\sigma_k} \gamma_{i,k} q_{k-i,k} + \bar{e}_k. \tag{1.4}$$

Aus den Orthogonalitätsrelationen (1.2) ergibt sich damit das folgende lineare Gleichungssystem der Dimension σ_k zur Bestimmung der $\gamma_{i,k}$:

$$\sum_{i=1}^{\sigma_k} \gamma_{i,k} q_{k-i,k}^t A^t Z_k q_{k-j,k} = -\bar{r}_k^t Z_k q_{k-j,k} \quad (j = 1, ..., \sigma_k). \tag{1.5}$$

Die Lösung des Systems $Ax = b$ der Dimension n wird also ersetzt durch das Lösen kleinerer Systeme (1.5) der Dimension σ_k in jedem Iterationsschritt.

Wir zeigen als Beispiel eine Möglichkeit, wie klassische Iterationsverfahren der Form

$$x_k = x_{k-1} - M^{-1}(Ax_{k-1} - b)$$

mit $r_k = Ax_k - b = (\mathbf{1} - AM^{-1})r_{k-1}$ in die Klasse der Orthogonalisierungsverfahren eingeordnet werden können. Wählen wir $\bar{x}_k := x_{k-1}$, $\sigma_k := 1$, $q_{k-1,1} := M^{-1}r_{k-1}$ und mit Hilfe einer nichtsingulären, schiefsymmetrischen Matrix S als Orthogonalisierungsmatrizen $Z_k := S(M - A)$, dann erhalten wir

$$\begin{aligned} r_k^t Z_k q_{k-1,1} &= r_{k-1}^t(\mathbf{1} - AM^{-1})^t S(M - A)M^{-1}r_{k-1} \\ &= r_{k-1}^t(\mathbf{1} - AM^{-1})^t S(\mathbf{1} - AM^{-1})r_{k-1} = 0. \end{aligned}$$

Da die Matrix S bis auf die genannten Bedingungen frei gewählt werden kann, ist die Wahl der Orthogonalisierungsmatrizen Z_k für ein beabsichtigtes Verfahren nicht eindeutig festgelegt.

A.1.2 Konvergenzeigenschaften

Trotz der großen Allgemeinheit ihrer Definition läßt sich für Orthogonalisierungsverfahren relativ leicht eine grundlegende und aussagekräftige Konvergenzabschätzung herleiten. Sie bildet den Ausgangspunkt für spätere Konvergenzuntersuchungen im Zusammenhang mit den spezielleren Krylovraum-Methoden. Wir benötigen zunächst zwei vorbereitende Lemmata.

Lemma 2 *Für ein Orthogonalisierungsverfahren gilt mit beliebigen reellen Koeffizienten* η_i $(i = 1, ..., \sigma_k)$

$$r_k^t Z_k A^{-1} r_k = r_k^t Z_k A^{-1}(\sum_{i=1}^{\sigma_k} \eta_i A q_{k-i,k} + \bar{r}_k). \tag{1.6}$$

Beweis:
Mit der Darstellung (1.3) des Residuums im Schritt k erhalten wir

$$\begin{aligned} r_k^t Z_k A^{-1} r_k &= r_k^t Z_k A^{-1}(\sum_{i=1}^{\sigma_k} \gamma_{i,k} A q_{k-i,k} + \bar{r}_k) \\ &= r_k^t Z_k A^{-1}(\sum_{i=1}^{\sigma_k} \eta_i A q_{k-i,k} + \bar{r}_k), \end{aligned}$$

da aufgrund der Orthogonalitätsrelationen (1.2) $r_k^t Z_k q_{k-i,k} = 0$ für $i = 1, ..., \sigma_k$ gilt.

□

Durch Gleichung (1.6) wird als quadratische Form eine geometrische Figur definiert. Man erhält durch die Beziehung zwischen Fehlern und Residuen eine entsprechende Figur für die Fehler. Für eine geometrische Interpretation dieser Figuren und ihre Bedeutung für das Verhalten der Fehler und Residuen im Fall exakter Verfahren verweisen wir auf [124, 125]. Als Konsequenz der dort geführten Diskussion notieren wir lediglich: Um im Fall exakter Verfahren eine wirkliche Fehlerreduktion bezüglich der euklidischen Norm zu erreichen, sollten die Orthogonalisierungsmatrizen Z_k so gewählt sein, daß die Singulärwerte von $Z_k^{-1}A^{-t}$ möglichst nahe beieinander liegen. Insbesondere führt die Wahl $Z_k = A^{-t}$ zu Verfahren, die die Fehler bezüglich der euklidischen Norm minimieren. Hierbei erscheint die Berechnung der rechten Seite des linearen Gleichungssystems (1.5) zur Bestimmung der Koeffizienten $\gamma_{i,k}$ jedoch mindestens so aufwendig wie die Lösung des Ausgangsproblems. Diese Schwierigkeit kann umgangen werden, indem man spezielle Suchrichtungen der Form $\breve{q}_{k-i,k} = A^t q_{k-i,k}$ verwendet.

Lemma 3 *Es seien $M \in \mathbb{R}^{n\times n}$ eine symmetrisch positiv definite Matrix und $R \in \mathbb{R}^{n\times n}$ schiefsymmetrisch. Dann gilt, wenn $\rho(R)$ den Spektralradius von R und $\lambda_{min}(M)$ den kleinsten Eigenwert von M bezeichnen,*

$$||\mathbf{1} + M^{-1}R||_M^2 \leq 1 + \frac{\rho^2(R)}{\lambda_{min}^2(M)}. \tag{1.7}$$

Beweis:
Wir erhalten die Abschätzungen

$$\begin{aligned} ||\mathbf{1} + M^{-1}R||_M^2 &= \max_{x\neq 0} \frac{x^t(\mathbf{1} + M^{-1}R)^t M(\mathbf{1} + M^{-1}R)x}{x^t M x} \\ &= \max_{x\neq 0}\left\{1 + \frac{x^t R^t M^{-1} R x}{x^t M x}\right\} \\ &\leq \max_{x\neq 0}\left\{1 + \frac{||M^{-1}||\,||Rx||^2}{\lambda_{min}(M)||x||^2}\right\} \\ &\leq 1 + \frac{\rho^2(R)}{\lambda_{min}^2(M)}, \end{aligned}$$

da $||M^{-1}|| = \frac{1}{\lambda_{min}(M)}$ ist. Darüberhinaus ist R normal, so daß $\frac{||Rx||}{||x||} \leq ||R|| = \rho(R)$ gilt. □

Mit Hilfe der beiden soeben vorgestellten Lemmata kann man die folgende Konvergenzabschätzung für allgemeine Orthogonalisierungsverfahren zeigen.

Satz 16 (Konvergenzabschätzung für Orthogonalisierungsverfahren)
Es seien die $Z_k A^{-1}$ positiv reell, *das heißt ihr symmetrischer Anteil sei positiv definit. Dann gelten für die Residuen und Fehler, die durch das zugehörige Orthogonalisierungsverfahren entstehen, die Abschätzungen*

$$||r_k||_{Z_k A^{-1}} \leq \sqrt{1 + \frac{\rho^2(R_k)}{\lambda_{min}^2(M_k)}} \min_{\eta_1,\ldots,\eta_{\sigma_k}} ||\sum_{i=1}^{\sigma_k} \eta_i A q_{k-i,k} + \bar{r}_k||_{Z_k A^{-1}}, \tag{1.8}$$

$$||e_k||_{A^t Z_k} \leq \sqrt{1 + \frac{\rho^2(R_k)}{\lambda_{min}^2(M_k)}} \min_{\eta_1,\ldots,\eta_{\sigma_k}} ||\sum_{i=1}^{\sigma_k} \eta_i q_{k-i,k} + \bar{e}_k||_{A^t Z_k}, \tag{1.9}$$

wobei $R_k = (Z_k A^{-1})^{sk}$ und $M_k = (Z_k A^{-1})^{sy}$ den schiefsymmetrischen respektive symmetrischen Anteil bezeichnen. Darüberhinaus gelten, falls die $Z_k A^{-1}$ symmetrisch (und damit auch symmetrisch positiv definit) sind,

$$||r_k||_{Z_k A^{-1}} = \min_{\eta_1,\ldots,\eta_{\sigma_k}} ||\sum_{i=1}^{\sigma_k} \eta_i A q_{k-i,k} + \bar{r}_k||_{Z_k A^{-1}}, \tag{1.10}$$

$$||e_k||_{A^t Z_k} = \min_{\eta_1,\ldots,\eta_{\sigma_k}} ||\sum_{i=1}^{\sigma_k} \eta_i q_{k-i,k} + \bar{e}_k||_{A^t Z_k}. \tag{1.11}$$

Beweis:

Für beliebige η_i ist nach der Darstellung (1.6) aus Lemma 2

$$\begin{aligned} ||r_k||^2_{Z_kA^{-1}} &= r_k^t Z_k A^{-1}(\sum_{i=1}^{\sigma_k} \eta_i A q_{k-i,k} + \bar{r}_k) \\ &= r_k^t M_k(\mathbf{1} + M_k^{-1}R_k)(\sum_{i=1}^{\sigma_k} \eta_i A q_{k-i,k} + \bar{r}_k). \end{aligned}$$

Es sei jetzt $\hat{r}_k$ ein Vektor mit $||\hat{r}_k||_{Z_kA^{-1}} = \min\limits_{\eta_1,\ldots,\eta_{\sigma_k}} ||\sum_{i=1}^{\sigma_k} \eta_i A q_{k-i,k} + \bar{r}_k||_{Z_kA^{-1}}$. Dann folgt aus der Tatsache, daß M_k symmetrisch positiv definit ist, zusammen mit der Cauchy–Schwarzschen Ungleichung und der Abschätzung (1.7) gemäß Lemma 3

$$\begin{aligned} ||r_k||^2_{Z_kA^{-1}} &= r_k^t M_k(\mathbf{1} + M_k^{-1}R_k)\hat{r}_k \\ &\leq ||r_k||_{M_k}||(\mathbf{1} + M_k^{-1}R_k)\hat{r}_k||_{M_k} \\ &\leq ||r_k||_{M_k}||(\mathbf{1} + M_k^{-1}R_k)||_{M_k}||\hat{r}_k||_{M_k} \\ &\leq ||r_k||_{Z_kA^{-1}}\sqrt{1 + \frac{\rho^2(R_k)}{\lambda^2_{min}(M_k)}}||\hat{r}_k||_{Z_kA^{-1}}. \end{aligned}$$

Hieraus folgt Abschätzung (1.8). Zum Beweis der entsprechenden Identität im Fall, daß die Z_kA^{-1} symmetrisch sind, bemerken wir, daß der schiefsymmetrische Anteil R_k dann jeweils verschwindet und daher der Wurzelausdruck 1 ergibt („$\leq$“). Die Richtung „$\geq$“ ist trivial, da r_k in jedem Fall eine Darstellung der Form besitzt, über deren Norm das Minimum genommen wird. Die Aussagen für die Fehler e_k folgen aufgrund der Tatsache, daß $r_k = Ae_k$ ist.

□

Die Abschätzungen (1.8) und (1.9) sowie die Gleichungen (1.10) und (1.11) sind bei näherem Hinsehen aufgrund der Beziehung zwischen Fehlern und Residuen identisch. Sie beschreiben den gleichen Sachverhalt lediglich aus unterschiedlichen Perspektiven, einmal aus der Sicht der Residuen und zum anderen aus Sicht der Fehler. Die Ausdrücke $||\cdot||_{A^tZ_k}$ stellen dabei im allgemeinen keine Normen dar, sondern sind als Abkürzungen für $\sqrt{(\cdot, A^tZ_k\cdot)}$ zu lesen, wobei die Skalarprodukte in den betreffenden Fällen stets nichtnegativ sind. Sind die Suchrichtungen $q_{k-i,k}$ linear unabhängig und ist das Verfahren exakt, so wird (bei exakter Arithmetik) die exakte Lösung spätestens im Schritt n erreicht. Die Suchrichtungen spannen dann nämlich den gesamten Lösungsraum auf, so daß $\hat{e}_n = \mathbf{0}$ ist. Die Orthogonalisierungsbedingungen entsprechen nach dem zweiten Teil des Satzes einer Minimierung, falls die Z_kA^{-1} symmetrisch positiv definit sind. Da sich Minimierungseigenschaften stets über Orthogonalisierungsbedingungen ausdrücken lassen [125], bilden Minimierungsverfahren eine Teilklasse der allgemeinen Klasse der Orthogonalisierungsverfahren. Wegen $||r_k|| = ||Ae_k|| \leq ||A||\,||e_k||$ und $||e_k|| = ||A^{-1}r_k|| \leq ||A^{-1}||\,||r_k||$ gelten die Beziehungen

$$\frac{||r_k||}{||r_0||} \leq \kappa\frac{||e_k||}{||e_0||} \quad \text{und} \quad \frac{||e_k||}{||e_0||} \leq \kappa\frac{||r_k||}{||r_0||},$$

wobei $\kappa = ||A||\cdot||A^{-1}||$ die Kondition der Matrix darstellt. Für kleine Konditionen $\kappa \approx 1$ sind demnach Fehler und Residuen eng aneinander gekoppelt, so daß eine Reduktion der relativen Residuen eine solche der relativen Fehler mit sich bringt und umgekehrt. Terminierungskriterien basierend auf der Größe der relativen Residuen sind also angemessen. Im Fall von Konditionen $\kappa \gg 1$ bewirkt eine Reduktion der relativen Residuen jedoch nicht notwendig eine entsprechende Reduktion der relativen Fehler und umgekehrt.

A.1.3 Verfahren für vorkonditionierte Gleichungssysteme

Wir betrachten eine Folge schrittabhängig vorkonditionierter Gleichungssysteme

$$P_{L,k}AP_{R,k}y^k = P_{L,k}b, \tag{1.12}$$

wobei $P_{L,k}$ und $P_{R,k}$ jeweils nichtsinguläre Vorkonditionierungsmatrizen aus $\mathbb{R}^{n\times n}$ darstellen und die y^k die exakten Lösungen der so vorkonditionierten Gleichungssysteme sind. Ziel dieses Abschnitts ist die Formulierung von Konvergenzabschätzungen für Orthogonalisierungsverfahren bezüglich der vorkonditionierten Systeme (1.12) aus Sicht des Ausgangssystems. Wir interessieren uns also für die Frage, was Orthogonalisierungsverfahren bezüglich der vorkonditionierten Systeme „tatsächlich“ bedeuten. Mit Hilfe der Konvergenzabschätzung des letzten Abschnitts erhalten wir durch Einsetzen der Darstellungen für die resultierenden Iterierten, Residuen und Fehler aus Sicht des ursprünglichen Gleichungssystems direkt die gewünschten Aussagen. Diese erlauben auch eine Interpretation der Verfahren auf der Basis des ursprünglichen Gleichungssystems mittels transformierter Suchrichtungen und transformierter Orthogonalisierungsbedingungen.

Ein Orthogonalisierungsverfahren entlang der vorkonditionierten Gleichungssysteme (1.12) berechnet ausgehend von Suchrichtungen $q^P_{k-j,k} \in \mathbb{R}^n$ $(j = 1, ..., \sigma_k)$ und Orthogonalisierungsmatrizen Z^P_k Iterierte

$$y_k = \sum_{i=1}^{\sigma_k} \gamma_{i,k} q^P_{k-i,k} + \bar{y}_k$$

über die Orthogonalisierungsbedingungen

$$\begin{aligned} (r^P_k)^t Z^P_k q^P_{k-j,k} &= 0 \quad (j = 1, ..., \sigma_k), \quad \text{d.h.} \\ \sum_{i=1}^{\sigma_k} \gamma_{i,k} (q^P_{k-i,k})^t P^t_{R,k} A^t P^t_{L,k} Z^P_k q^P_{k-j,k} &= -(\bar{r}^P_k)^t Z^P_k q^P_{k-j,k} \quad (j = 1, ..., \sigma_k). \end{aligned} \tag{1.13}$$

Hierbei ist

$$\bar{r}^P_k = P_{L,k}AP_{R,k}y^k - P_{L,k}b.$$

Definieren wir ferner

$$\bar{e}^P_k = \bar{y}_k - y^k,$$

so erhalten wir für die Iterierten, Residuen und Fehler aus Sicht des ursprünglichen Systems

$$x_k = P_{R,k}y_k = \sum_{i=1}^{\sigma_k} \gamma_{i,k} P_{R,k} q^P_{k-i,k} + P_{R,k}\bar{y}_k, \tag{1.14}$$

$$r_k = P^{-1}_{L,k} r^P_k = \sum_{i=1}^{\sigma_k} \gamma_{i,k} AP_{R,k} q^P_{k-i,k} + P_{R,k}\bar{r}^P_k, \tag{1.15}$$

$$e_k = P_{R,k}e^P_k = \sum_{i=1}^{\sigma_k} \gamma_{i,k} P_{R,k} q^P_{k-i,k} + P_{R,k}\bar{e}^P_k. \tag{1.16}$$

Anhand der Konvergenzabschätzung des letzten Abschnitts können wir nun mit Hilfe dieser Darstellungen die nachfolgende Konvergenzabschätzung für Orthogonalisierungsverfahren bezüglich der vorkonditionierten Gleichungen (1.12) herleiten.

Satz 17 (Konvergenzabschätzung für vorkonditionierte Gleichungen)
Es seien die $Z_k^P(P_{L,k}AP_{R,k})^{-1}$ positiv reell. Dann gelten für die Residuen und Fehler aus Sicht des ursprünglichen Systems, die durch ein Orthogonalisierungsverfahren für die entsprechend vorkonditionierten Systeme entstehen, die Abschätzungen

$$||r_k||_{P_{L,k}^t Z_k^P P_{R,k}^{-1} A^{-1}} \leq \sqrt{1 + \frac{\rho^2(R_k^P)}{\lambda_{min}^2(M_k^P)}} \min_{\eta_1,\ldots,\eta_{\sigma_k}} ||\sum_{i=1}^{\sigma_k} \eta_i AP_{R,k} q_{k-i,k}^P + \bar{r}_k||_{P_{L,k}^t Z_k^P P_{R,k}^{-1} A^{-1}}, \quad (1.17)$$

$$||e_k||_{A^t P_{L,k}^t Z_k^P P_{R,k}^{-1}} \leq \sqrt{1 + \frac{\rho^2(R_k^P)}{\lambda_{min}^2(M_k^P)}} \min_{\eta_1,\ldots,\eta_{\sigma_k}} ||\sum_{i=1}^{\sigma_k} \eta_i P_{R,k} q_{k-i,k}^P + \bar{e}_k||_{A^t P_{L,k}^t Z_k^P P_{R,k}^{-1}}, \quad (1.18)$$

wobei $R_k^P := (Z_k^P(P_{L,k}AP_{R,k})^{-1})^{sk}$ und $M_k^P := (Z_k^P(P_{L,k}AP_{R,k})^{-1})^{sy}$ sind. Darüberhinaus gelten, falls die $Z_k^P(P_{L,k}AP_{R,k})^{-1}$ symmetrisch sind,

$$||r_k||_{P_{L,k}^t Z_k^P P_{R,k}^{-1} A^{-1}} = \min_{\eta_1,\ldots,\eta_{\sigma_k}} ||\sum_{i=1}^{\sigma_k} \eta_i AP_{R,k} q_{k-i,k}^P + \bar{r}_k||_{P_{L,k}^t Z_k^P P_{R,k}^{-1} A^{-1}}, \quad (1.19)$$

$$||e_k||_{A^t P_{L,k}^t Z_k^P P_{R,k}^{-1}} = \min_{\eta_1,\ldots,\eta_{\sigma_k}} ||\sum_{i=1}^{\sigma_k} \eta_i P_{R,k} q_{k-i,k}^P + \bar{e}_k||_{A^t P_{L,k}^t Z_k^P P_{R,k}^{-1}}. \quad (1.20)$$

Beweis:
Der Beweis folgt direkt aus Satz 16 durch Einsetzen der Darstellungen (1.15) und (1.16) für die Residuen und Fehler aus Sicht des ursprünglichen Systems.

□

Vorkonditionierte Orthogonalisierungsverfahren entsprechen daher Orthogonalisierungsverfahren für das ursprüngliche System mit gemäß

$$P_{R,k} q_{k-i,k}^P \longrightarrow q_{k-i,k}, \quad (1.21)$$

$$P_{L,k}^t Z_k^P P_{R,k}^{-1} \longrightarrow Z_k \quad (1.22)$$

transformierten Suchrichtungen und Orthogonalisierungsmatrizen.

A.2 Krylovraum-Verfahren

A.2.1 Allgemeine Eigenschaften

Zu den gebräuchlichsten Orthogonalisierungsverfahren zählen die sogenannten *Krylovraum-Verfahren*. Ein Grund hierfür ist, daß man für ihre Implementierung im wesentlichen nur Matrix-Vektor-Multiplikationen benötigt. Sie eignen sich daher insbesondere zur Lösung von linearen Gleichungssystemen mit schwachbesetzten Systemmatrizen. Krylovraum-Verfahren sind Orthogonalisierungsverfahren mit

$$q_{k-i,k} \in \mathcal{K}_{k-i}(B,z) := \langle z, Bz, ..., B^{k-i}z \rangle,$$
$$\bar{x}_k = x_{k-\sigma_k},$$

wobei $B \in \mathbb{R}^{n\times n}$ eine gegebene Matrix und $z \in \mathbb{R}^n$ ein gegebener Vektor sind. Unsere Definition des Krylovraums $\mathcal{K}_k(B,z)$ stimmt nicht ganz mit der in der Literatur üblichen Definition $\bar{\mathcal{K}}_k(B,z) := \langle z, Bz, ..., B^{k-1}z \rangle$ überein [104]. Für *Konjugierte Krylovraum-Verfahren*

(Conjugate Krylov Subspace methods, CKS-Verfahren) werden speziell $B = A$ und $z = r_0$ gewählt. Wir fassen im folgenden Lemma kurz einige einfache Eigenschaften der zugehörigen Krylovräume und Verfahren zusammen.

Lemma 4 (Elementare Eigenschaften von CKS-Verfahren)
Es sei $p_{min}(A, r_0)$ das Minimalpolynom zum Vektor r_0, das heißt $p_{min}(A, r_0)$ sei ein Polynom minimalen Grades μ, so daß $p_{min}(A, r_0) = \sum_{k=0}^{\mu} \alpha_k A^k r_0 = 0$ mit $\alpha_k \neq 0$ für mindestens ein $k \in \{0, ..., \mu\}$ gilt. Dann gelten die folgenden Aussagen:

1. *dim $\mathcal{K}_m(A, r_0) = m + 1 \Longleftrightarrow$ Grad $(p_{min}(A, r_0)) > m$.*

2. *$\mathcal{K}_{\mu-1}(A, r_0)$ ist invariant bezüglich A und $\mathcal{K}_{\mu-1}(A, r_0) = \mathcal{K}_{m-1}(A, r_0)$ für alle $m \geq \mu$.*

3. *Die Lösung $x = x_0 - e_0$ des linearen Gleichungssystems (1.1) liegt in $x_0 + \mathcal{K}_{\mu-1}(A, r_0)$.*

Beweis:
Die beiden ersten Aussagen sind offensichtlich. Wir beweisen daher lediglich die dritte Behauptung. Es sei $j \geq 0$ der minimale Index, so daß $\alpha_j \neq 0$. Dann gilt

$$\begin{aligned} A^j r_0 &= -\sum_{k=j+1}^{\mu} \frac{\alpha_k}{\alpha_j} A^k r_0 \\ \Longleftrightarrow \qquad r_0 &= \underbrace{A\,(-\sum_{k=j+1}^{\mu} \frac{\alpha_k}{\alpha_j} A^{k-j-1} r_0)}_{=e_0 \in \mathcal{K}_{\mu-1}(A, r_0)}. \end{aligned}$$

□

Da der Grad μ des Minimalpolynoms stets kleiner als die Dimension n des linearen Gleichungssystems ist, ist die Lösung des Gleichungssystems in Krylovräumen überhaupt möglich. Die naheliegendste Basis $\{r_0, Ar_0, ..., A^k r_0\}$ des Krylovraums $\mathcal{K}_k(A, r_0)$ ist aus numerischer Sicht für die Implementierung der Verfahren ungeeignet. Die Vektoren $A^j r_0$ zeigen für wachsendes j unter Umständen immer stärker in Richtung des dominierenden Eigenvektors von A (Stichwort „Powermethode“, siehe [51]) und die resultierende Basis kann im Verfahren numerische Instabilitäten hervorrufen. Deshalb verwenden die meisten Krylovraum-Verfahren nach Möglichkeit orthogonale oder biorthogonale Basen, die über Arnoldi- und (Bi-)Lanczos-Prozesse konstruiert werden [104]. Sie führen im Idealfall zu numerisch stabilen Verfahren mit kurzen Rekursionen. Letzteres bedeutet, daß zur Bestimmung der Iterierten im Schritt k nicht die gesamte Menge der den Krylovraum $\mathcal{K}_{k-1}(A, r_0)$ aufspannenden Basisvektoren gespeichert zu werden braucht, sondern nur einige wenige zuletzt berechnete Vektoren. Die Orthogonalitätsforderungen lassen sich dabei über verallgemeinerte Projektionen des ursprünglichen Systems in die Krylovräume $\mathcal{K}_{k-1}(A, r_0)$ interpretieren. In Fällen mit indefiniten oder nichtsymmetrischen Systemmatrizen können die Prozeduren zum Auffinden orthogonaler oder biorthogonaler Basen jedoch versagen (“serious breakdown”) und zur Divergenz des Krylovraum-Verfahrens führen.

A.2.2 Konjugierte Krylovraum-Verfahren

Für Konjugierte Krylovraum-Verfahren (CKS-Verfahren) liegen die Suchrichtungen $q_{k-i,k}$ im Krylovraum $\mathcal{K}_{k-i}(A, r_0)$. Wegen

$$\begin{aligned} x_k &\in x_0 + \mathcal{K}_{k-1}(A, r_0), \\ r_k &\in r_0 + A\mathcal{K}_{k-1}(A, r_0), \\ e_k &\in e_0 + \mathcal{K}_{k-1}(A, r_0) = e_0 + \mathcal{K}_{k-1}(A, Ae_0) \\ &= e_0 + A\mathcal{K}_{k-1}(A, e_0) \end{aligned}$$

folgen mit geeigneten Koeffizienten $\beta_{i,k}$ und $\eta_{i,k}$ die Darstellungen

$$x_k = \sum_{i=1}^{k} \beta_{i,k} A^{i-1} r_0 + x_0 = \sum_{i=1}^{k} \eta_{i,k} r_{i-1} + x_0, \tag{1.23}$$

$$r_k = \sum_{i=1}^{k} \beta_{i,k} A^{i} r_0 + r_0 = \sum_{i=1}^{k} \eta_{i,k} A r_{i-1} + r_0, \tag{1.24}$$

$$e_k = \sum_{i=1}^{k} \beta_{i,k} A^{i} e_0 + e_0 = \sum_{i=1}^{k} \eta_{i,k} A e_{i-1} + e_0. \tag{1.25}$$

Für exakte CKS-Verfahren betrachten wir die beiden naheliegenden Zuordnungen

$$\begin{aligned} q_{k-i,k} &= r_{k-i}, \\ q_{k-i,k} &= A^{k-i} r_0 \end{aligned}$$

und erhalten als zugehörige Orthogonalisierungsbedingungen die äquivalenten Forderungen

$$r_k^t Z_k r_{k-j} = 0, \tag{1.26}$$

$$r_k^t Z_k A^{k-j} r_0 = 0 \quad (j = 1, ..., (\sigma_k = k)). \tag{1.27}$$

Die Bedingungen (1.26) erklären die Bezeichnung „Konjugierte Krylovraum-Verfahren“, da die Suchrichtungen r_k paarweise Z_k-konjugiert sind. Eine allgemeine Konvergenzabschätzung für exakte Konjugierte Krylovraum-Verfahren erhält man unmittelbar aus Satz 16.

Satz 18 (Konvergenzabschätzung für exakte CKS-Verfahren)
Es seien die $Z_k A^{-1}$ positiv reell. Dann erhält man für die Residuen und Fehler, die durch exakte CKS-Verfahren entstehen, die Abschätzungen

$$||r_k||_{Z_k A^{-1}} \leq \sqrt{1 + \frac{\rho^2(R_k)}{\lambda_{min}^2(M_k)}} \min_{\beta_1,...,\beta_k} ||\sum_{i=1}^{k} \beta_i A^i r_0 + r_0||_{Z_k A^{-1}}, \tag{1.28}$$

$$||e_k||_{A^t Z_k} \leq \sqrt{1 + \frac{\rho^2(R_k)}{\lambda_{min}^2(M_k)}} \min_{\beta_1,...,\beta_k} ||\sum_{i=1}^{k} \beta_i A^i e_0 + e_0||_{A^t Z_k}, \tag{1.29}$$

wobei $R_k = (Z_k A^{-1})^{sk}$ und $M_k = (Z_k A^{-1})^{sy}$ sind. Sind die $Z_k A^{-1}$ symmetrisch, so gelten

$$||r_k||_{Z_k A^{-1}} = \min_{\beta_1,...,\beta_k} ||\sum_{i=1}^{k} \beta_i A^i r_0 + r_0||_{Z_k A^{-1}}, \tag{1.30}$$

$$||e_k||_{A^t Z_k} = \min_{\eta_1,...,\eta_{\sigma_k}} ||\sum_{i=1}^{k} \beta_i A^i e_0 + e_0||_{A^t Z_k}. \tag{1.31}$$

Beweis:
Die Behauptungen folgen mit Satz 16.

□

Wir untersuchen nun, wie sich hieraus verfeinerte Aussagen ergeben und zeigen zuerst Abschätzungen, die insbesondere im Falle normaler Matrizen A zu detaillierteren Konvergenzaussagen unter Kenntnis des Spektrums von A führen. Wir verwenden dazu eine Reihe von einfachen Lemmata. Für nichtnormale Matrizen ist die Konvergenz zugehöriger iterativer Verfahren bekanntlich nicht alleine durch das Spektrum der Matrizen bestimmt [41, 87, 115, 116]. Wir geben deshalb noch weitere Abschätzungen an, die auf dem *Numerischen Wertebereich* von A sowie von A^{-1} beruhen. Für noch detailliertere Konvergenzabschätzungen die auf dem *Pseudospektrum* und lokal verfeinerten Abschätzungen basieren verweisen wir auf [43].

Lemma 5 *Es sei $G \in \mathbb{R}^{n\times n}$ positiv reell und $S := (G^{sy})^{\frac{1}{2}}$. Dann gilt für alle $x \in \mathbb{R}^n$*

$$\frac{||x||}{||S^{-1}||} \leq ||x||_G \leq ||S||\,||x||. \tag{1.32}$$

Beweis:
Die behaupteten Ungleichungen folgen mit

$$||x||_G^2 = x^t G x = x^t \frac{1}{2}(G + G^t)x = x^t S^t S x = ||Sx||^2 \leq ||S||^2 ||x||^2,$$

$$||x|| = ||S^{-1}Sx|| \leq ||S^{-1}||\,||Sx|| = ||S^{-1}||\,||x||_G.$$

□

Lemma 6 *Es sei $A \in \mathbb{R}^{n\times n}$ eine positiv reelle Matrix. Dann erhält man die Abschätzung*

$$\min_\alpha ||\mathbf{1} + \alpha A||^2 \leq 1 - \frac{\lambda_{min}^2}{||A||^2}, \tag{1.33}$$

wobei λ_{min} den kleinsten Eigenwert des symmetrischen Anteils von A bezeichnet.

Beweis:
Der Ausdruck

$$\begin{aligned} ||\mathbf{1} + \alpha A||^2 &= \max_{x\neq 0} \frac{x^t(\mathbf{1} + \alpha A)^t(\mathbf{1} + \alpha A)x}{x^t x} \\ &= \max_{x\neq 0} \left(1 + 2\alpha \frac{x^t A x}{x^t x} + \alpha^2 \frac{x^t A^t A x}{x^t x}\right) \end{aligned}$$

nimmt sein Minimum an für $\alpha_0 = -\frac{x^t A x}{x^t A^t A x}$. Damit gilt

$$\begin{aligned} ||\mathbf{1} + \alpha_0 A||^2 &= \max_{x\neq 0} \left(1 - \frac{(x^t A x)^2}{(x^t x)^2} \frac{x^t x}{x^t A^t A x}\right) \\ &\leq 1 - \frac{\lambda_{min}^2}{||A||^2}, \end{aligned}$$

da A^{sy} nach Voraussetzung positiv definit und $\frac{x^t A x}{x^t x} = \frac{x^t A^{sy} x}{x^t x}$ ist.

□

Lemma 7 *Es sei $A \in \mathbb{R}^{n\times n}$ eine positiv reelle Matrix. Dann gilt die Abschätzung*

$$\min_{\theta_1,\ldots,\theta_k} ||\sum_{i=1}^{k} \theta_i A^i + \mathbf{1}|| \leq \left(1 - \frac{\lambda_{min}^2}{||A||^2}\right)^{\frac{k}{2}}. \tag{1.34}$$

Beweis:
Mit Hilfe des vorigen Lemmas erhalten wir

$$\begin{aligned} \min_{\theta_1,\ldots,\theta_k} ||\sum_{i=1}^{k} \theta_i A^i + \mathbf{1}|| &\leq \min_{\alpha} ||(\mathbf{1} + \alpha A)^k|| \\ &\leq \min_{\alpha} ||\mathbf{1} + \alpha A||^k \\ &\leq \left(1 - \frac{\lambda_{min}^2}{||A||^2}\right)^{\frac{k}{2}}. \end{aligned}$$

□

Lemma 8 *Es sei $J \in \mathbb{R}^{n\times n}$ die Matrix der Jordanschen Normalform zur Matrix $A \in \mathbb{R}^{n\times n}$, so daß $A = C^{-1}JC$ für eine reguläre Matrix $C \in \mathbb{R}^{n\times n}$ ist. Dann gilt*

$$\min_{\theta_1,\ldots,\theta_k} ||\sum_{i=1}^{k} \theta_i A^i + \mathbf{1}|| \leq \kappa(C) \min_{\theta_1,\ldots,\theta_k} ||\sum_{i=1}^{k} \theta_i J^i + \mathbf{1}||. \tag{1.35}$$

Beweis:
Es gilt

$$\begin{aligned} \min_{\theta_1,\ldots,\theta_k} ||\sum_{i=1}^{k} \theta_i A^i + \mathbf{1}|| &= \min_{\theta_1,\ldots,\theta_k} ||\sum_{i=1}^{k} \theta_i C^{-1} J^i C + C^{-1}C|| \\ &\leq ||C^{-1}||\,||C|| \min_{\theta_1,\ldots,\theta_k} ||\sum_{i=1}^{k} \theta_i J^i + \mathbf{1}||. \end{aligned}$$

□

Lemma 9 *Es sei $A \in \mathbb{R}^{n\times n}$ diagonalisierbar, das heißt es existiere eine Diagonalmatrix $D \in \mathbb{R}^{n\times n}$, so daß $A = C^{-1}DC$ für eine reguläre Matrix $C \in \mathbb{R}^{n\times n}$ ist. Dann gilt*

$$\min_{\theta_1,\ldots,\theta_k} ||\sum_{i=1}^{k} \theta_i A^i + \mathbf{1}|| \leq \kappa(C) \min_{\theta_1,\ldots,\theta_k} \max_{\lambda\in\sigma(A)} |\sum_{i=1}^{k} \theta_i \lambda^i + 1|. \tag{1.36}$$

Ist A darüberhinaus normal, dann gilt die obige Abschätzung sogar mit $\kappa(C) = 1$, da die Transformation C orthogonal gewählt werden kann.

Beweis:
Die Behauptung folgt wie Lemma 8.

□

Satz 19 (Konvergenzabschätzungen anhand des Spektrums)
Es seien A und die $Z_k A^{-1}$ positiv reell. Dann erhält man für die Residuen und Fehler aus exakten Konjugierten Krylovraum-Verfahren die Abschätzungen

$$||r_k|| \leq \sqrt{1+\frac{\rho^2(R_k)}{\lambda^2_{min}(M_k)}}\kappa(S_k)\left(1-\frac{\lambda^2_{min}}{||A||^2}\right)^{\frac{k}{2}}||r_0||, \tag{1.37}$$

$$||e_k|| \leq \sqrt{1+\frac{\rho^2(R_k)}{\lambda^2_{min}(M_k)}}\kappa(T_k)\left(1-\frac{\lambda^2_{min}}{||A||^2}\right)^{\frac{k}{2}}||e_0||, \tag{1.38}$$

wobei $R_k = (Z_k A^{-1})^{sk}$, $M_k = (Z_k A^{-1})^{sy}$, $S_k = ((Z_k A^{-1})^{sy})^{\frac{1}{2}}$ und $T_k = ((A^t Z_k)^{sy})^{\frac{1}{2}}$ sind. Darüberhinaus gelten

$$||r_k|| \leq \sqrt{1+\frac{\rho^2(R_k)}{\lambda^2_{min}(M_k)}}\kappa(S_k)\min_{\theta_1,\ldots,\theta_k}||\sum_{i=1}^{k}\theta_i A^i + \mathbf{1}||\,||r_0||, \tag{1.39}$$

$$||e_k|| \leq \sqrt{1+\frac{\rho^2(R_k)}{\lambda^2_{min}(M_k)}}\kappa(T_k)\min_{\theta_1,\ldots,\theta_k}||\sum_{i=1}^{k}\theta_i A^i + \mathbf{1}||\,||e_0||. \tag{1.40}$$

Es sei J die Matrix der Jordanschen Normalform zur Matrix A, so daß $A = C^{-1}JC$ für eine reguläre Matrix C ist. Es folgt

$$||r_k|| \leq \sqrt{1+\frac{\rho^2(R_k)}{\lambda^2_{min}(M_k)}}\kappa(S_k)\kappa(C)\min_{\theta_1,\ldots,\theta_k}||\sum_{i=1}^{k}\theta_i J^i + \mathbf{1}||\,||r_0||, \tag{1.41}$$

$$||e_k|| \leq \sqrt{1+\frac{\rho^2(R_k)}{\lambda^2_{min}(M_k)}}\kappa(T_k)\kappa(C)\min_{\theta_1,\ldots,\theta_k}||\sum_{i=1}^{k}\theta_i J^i + \mathbf{1}||\,||e_0||. \tag{1.42}$$

Ist A diagonalisierbar, das heißt existiert eine Diagonalmatrix D, so daß $A = C^{-1}DC$ für eine reguläre Matrix C ist, dann erhält man

$$||r_k|| \leq \sqrt{1+\frac{\rho^2(R_k)}{\lambda^2_{min}(M_k)}}\kappa(S_k)\kappa(C)\min_{\theta_1,\ldots,\theta_k}\max_{\lambda\in\sigma(A)}|\sum_{i=1}^{k}\theta_i\lambda^i + 1|\,||r_0||, \tag{1.43}$$

$$||e_k|| \leq \sqrt{1+\frac{\rho^2(R_k)}{\lambda^2_{min}(M_k)}}\kappa(T_k)\kappa(C)\min_{\theta_1,\ldots,\theta_k}\max_{\lambda\in\sigma(A)}|\sum_{i=1}^{k}\theta_i\lambda^i + 1|\,||e_0||. \tag{1.44}$$

Ist A darüberhinaus sogar normal, so gelten die letzten beiden Abschätzung mit $\kappa(C) = 1$.

Beweis:
Der Beweis ergibt sich aus Satz 18 durch Anwenden der oben stehenden Lemmata. Die Ungleichungen (1.39) und (1.40) erhält man mit Hilfe von Lemma 5 aus den Abschätzungen (1.28) und (1.29). Zusammen mit Lemma 7 folgen dann (1.37) und (1.38), zusammen mit Lemma 8 und Lemma 9 jeweils (1.41) und (1.42) sowie (1.43) und (1.44).

□

Bekanntlich ist der *Numerische Wertebereich* von A bezüglich eines Skalaprodukts $(\cdot,\cdot)_B$, das durch eine symmetrisch positiv definite Matrix B induziert wird, definiert als [72]

$$W_B(A) := \{\frac{x^*BAx}{x^*Bx} : \mathbf{0} \neq x \in \mathbb{C}^n\}. \tag{1.45}$$

Der folgende Satz gibt eine Konvergenzabschätzung für exakte CKS-Verfahren auf der Basis der Numerischen Wertebereiche von A und A^{-1} bezüglich der Skalarprodukte $(\cdot,\cdot)_{M_k}$, wobei die $M_k = (Z_k A^{-1})^{sy}$ sind.

Satz 20 (Konvergenzabschätzungen anhand des Numerischen Wertebereichs) *Es seien die $Z_k A^{-1}$ positiv reell Wir definieren die Größen*

$$\tau_k := \inf_{\substack{w\in\mathbb{R}^n \\ w\neq 0}} \frac{(w,Aw)_{Z_kA^{-1}}}{(w,w)_{Z_kA^{-1}}} \quad , \qquad \bar{\tau}_k := \inf_{\substack{w\in\mathbb{R}^n \\ w\neq 0}} \frac{(w,A^{-1}w)_{Z_kA^{-1}}}{(w,w)_{Z_kA^{-1}}}.$$

Dann gelten für die Residuen und Fehler, die durch exakte Konjugierte Krylovraum-Verfahren entstehen, die Abschätzungen

$$||r_k||_{Z_kA^{-1}} \leq \sqrt{1+\frac{\rho^2(R_k)}{\lambda^2_{min}(M_k)}}(1-\tau_k\cdot\bar{\tau}_k)^{\frac{k}{2}}||r_0||_{Z_kA^{-1}}, \tag{1.46}$$

$$||e_k||_{A^tZ_k} \leq \sqrt{1+\frac{\rho^2(R_k)}{\lambda^2_{min}(M_k)}}(1-\tau_k\cdot\bar{\tau}_k)^{\frac{k}{2}}||e_0||_{A^tZ_k}, \tag{1.47}$$

wobei $R_k = (Z_kA^{-1})^{sk}$ und $M_k = (Z_kA^{-1})^{sy}$ sind. Sind die Z_kA^{-1} symmetrisch, so gelten die entsprechenden Abschätzungen ohne die Wurzelausdrücke.

Beweis:
Wir gehen von Satz 18 aus und schätzen die Minimumsausdrücke auf geschickte Art und Weise ab. Wir tun dies zunächst für das Residuum und betrachten dazu die folgende Umformulierung der Minimumseigenschaft. Es sei $\mathcal{P}_k$ der Raum der Polynome vom Grad k, für die $p_k(0) = 1$ ist. Dann gilt

$$\min_{\beta_1,\ldots,\beta_k} ||\sum_{i=1}^{k}\beta_iA^ir_0 + r_0||_{Z_kA^{-1}} = \min_{p_k\in\mathcal{P}_k} ||p_k(A)r_0||_{Z_kA^{-1}}.$$

Wir konstruieren nun rekursiv eine spezielle Folge $\{\tilde{p}_j\}_{0\leq j\leq k}$ solcher Polynome durch

$$\begin{aligned}\tilde{p}_0(z) &:= 1,\\ \tilde{p}_j(z) &:= \left(1-\frac{(A\tilde{r}^{j-1},\tilde{r}^{j-1})_{Z_kA^{-1}}}{||A\tilde{r}^{j-1}||^2_{Z_kA^{-1}}}z\right)\tilde{p}_{j-1}(z).\end{aligned}$$

mit $\tilde{r}^{j-1} = \tilde{p}_{j-1}(A)r_0$ für $j \geq 1$. Damit erhalten wir

$$\begin{aligned}||\tilde{r}^j||^2_{Z_kA^{-1}} = ||\tilde{p}_j(A)r_0||^2_{Z_kA^{-1}} &= ||\tilde{r}^{j-1} - \frac{(A\tilde{r}^{j-1},\tilde{r}^{j-1})_{Z_kA^{-1}}}{||A\tilde{r}^{j-1}||^2_{Z_kA^{-1}}}A\tilde{r}^{j-1}||^2_{Z_kA^{-1}}\\ &= \left(1-\frac{(\tilde{r}^{j-1},A\tilde{r}^{j-1})_{Z_kA^{-1}}}{||\tilde{r}^{j-1}||^2_{Z_kA^{-1}}}\frac{(A\tilde{r}^{j-1},\tilde{r}^{j-1})_{Z_kA^{-1}}}{||A\tilde{r}^{j-1}||^2_{Z_kA^{-1}}}\right)||\tilde{r}^{j-1}||^2_{Z_kA^{-1}}\\ &\leq (1-\tau_k\cdot\bar{\tau}_k)||\tilde{r}^{j-1}||^2_{Z_kA^{-1}}.\end{aligned}$$

Zusammen mit Satz 18 folgt hieraus die behauptete Abschätzung (1.46) für die Residuen. Abschätzung (1.47) für die Fehler erhält man dann wegen $r_k = Ae_k$.

□

Die Größen τ_k und $\bar{\tau}_k$ sind gerade die Infima der Realteile der Numerischen Wertebereiche von A beziehungsweise von A^{-1} bezüglich der Skalarprodukte $(\cdot\,,\cdot)_{M_k}$, wobei die $M_k = (Z_k A^{-1})^{sy}$ sind [109]:

$$\begin{aligned} \tau_k &= \inf\{\mathrm{Re}(z) \,:\, z \in W_{M_k}(A)\}, \\ \bar{\tau}_k &= \inf\{\mathrm{Re}(z) \,:\, z \in W_{M_k}(A^{-1})\}. \end{aligned}$$

Die entsprechenden Aussagen des obigen Satzes für vorkonditionierte Systeme lauten:

Satz 21 (Konvergenzabschätzungen im vorkonditionierten Fall)
Es seien die $Z_k(P_{L,k}AP_{R,k})^{-1}$ positiv reell sowie

$$\tau_k^P := \inf_{\substack{w\in\mathbb{R}^n \\ w\neq 0}} \frac{(w, P_{L,k}AP_{R,k}w)_{Z_k(P_{L,k}AP_{R,k})^{-1}}}{(w,w)_{Z_k(P_{L,k}AP_{R,k})^{-1}}},\ \bar{\tau}_k^P := \inf_{\substack{w\in\mathbb{R}^n \\ w\neq 0}} \frac{(w, P_{R,k}^{-1}A^{-1}P_{L,k}^{-1}w)_{Z_k(P_{L,k}AP_{R,k})^{-1}}}{(w,w)_{Z_k(P_{L,k}AP_{R,k})^{-1}}}.$$

Dann gelten für die Residuen und Fehler, die durch ein vorkonditioniertes exaktes CKS-Verfahren entstehen, aus Sicht des ursprünglichen Systems die Abschätzungen

$$||r_k||_{P_{L,k}^t Z_k P_{R,k}^{-1} A^{-1}} \leq \sqrt{1 + \frac{\rho^2(R_k^P)}{\lambda_{min}^2(M_k^P)}}(1 - \tau_k^P \cdot \bar{\tau}_k^P)^{\frac{k}{2}} ||r_0||_{P_{L,k}^t Z_k P_{R,k}^{-1} A^{-1}}, \tag{1.48}$$

$$||e_k||_{A^t P_{L,k}^t Z_k P_{R,k}^{-1}} \leq \sqrt{1 + \frac{\rho^2(R_k^P)}{\lambda_{min}^2(M_k^P)}}(1 - \tau_k^P \cdot \bar{\tau}_k^P)^{\frac{k}{2}} ||e_0||_{A^t P_{L,k}^t Z_k P_{R,k}^{-1}}, \tag{1.49}$$

wobei $R_k^P = (Z_k(P_{L,k}AP_{R,k})^{-1})^{sk}$ und $M_k^P = (Z_k(P_{L,k}AP_{R,k})^{-1})^{sy}$ sind. Sind darüberhinaus die $Z_k(P_{L,k}AP_{R,k})^{-1}$ symmetrisch, so gelten die entsprechenden Abschätzungen ohne die Wurzelausdrücke.

Beweis:
Der Beweis ergibt sich aus Satz 20 durch Einsetzen der Darstellungen der Residuen und Fehler aus Sicht des unvorkonditionierten Gleichungssystems.

□

A.2.3 Verallgemeinerte Krylovraum-Verfahren

Verallgemeinerte Krylovraum-Verfahren sind CKS-Verfahren, deren Orthogonalisierungsvorschriften nicht mehr vom Iterationsschritt abhängen, für die also $Z_k = Z = const$ ist. Wir stellen im folgenden einige Verfahren und die sie charakterisierenden Eigenschaften vor, was auf der Basis der im vorherigen Abschnitt gezeigten Konvergenzabschätzungen einfach durch Wahl geeigneter Orthogonalisierungsmatrizen geschieht. Für Details ihrer Implementierungen verweisen wir auf [104, 125].

Verfahren, die die verallgemeinerte Energienorm des Fehlers „minimieren“

Wählt man für das zu lösende lineare Gleichungssystem (1.1) ein Orthogonalisierungsverfahren mit $Z = \mathbf{1}$, so minimiert das exakte Verfahren gemäß Satz (19) die verallgemeinerte Energienorm des Fehlers im Sinne

$$||e_k||_A \leq \sqrt{1 + \frac{\rho^2(R)}{\lambda_{min}^2(M)}} \min_{\beta_1,\ldots,\beta_k} ||\sum_{i=1}^{k} \beta_i A^i e_0 + e_0||_A.$$

Ist A symmetrisch positiv definit, also insbesondere normal, so erhalten wir darüberhinaus für die Fehler und Residuen die Abschätzungen

$$\begin{aligned} ||e_k|| &\leq \kappa(S) \min_{\theta_1,\ldots,\theta_k} \max_{\lambda\in\sigma(A)} |\sum_{i=1}^{k} \theta_i \lambda^i + 1| \, ||e_0||, \\ ||r_k|| &\leq \kappa(S) \min_{\theta_1,\ldots,\theta_k} \max_{\lambda\in\sigma(A)} |\sum_{i=1}^{k} \theta_i \lambda^i + 1| \, ||r_0||, \end{aligned}$$

wobei $S = A^{\frac{1}{2}}$ ist. Durch Substitution des darin auftretenden Polynoms in λ durch ein normiertes Tschebyscheff-Polynom folgt die bekannte Abschätzung für das klassische CG-Verfahren

$$||e_k|| \leq 2\kappa(S) \left(\frac{\sqrt{\kappa(A)} - 1}{\sqrt{\kappa(A)} + 1} \right)^k ||e_0||.$$

Ist A symmetrisch, so sind als Verfahren, die die verallgemeinerte Energienorm des Fehlers im obigen Sinne minimieren, FOM und SYMMLQ zu nennen. Beide beruhen auf den im symmetrischen Fall beim Arnoldi-Prozeß entstehenden tridiagonalen Hessenbergmatrizen, wobei SYMMLQ eine LQ-Zerlegung hiervon berechnet. Sie können daher mit kurzen Rekursionen implementiert werden. Ist A sogar symmetrisch positiv definit, so führt die LU-Faktorisierung der tridiagonalen Hessenbergmatrizen zusammen mit FOM zum CG-Verfahren. Im Fall positiv reeller Matrizen bietet ORTHORES eine Verallgemeinerung des CG-Verfahrens für nichtsymmetrische Matrizen dieses Typs. Das exakte Verfahren kann außer für symmetrische Matrizen nicht mit kurzer Rekursion implementiert werden. Für symmetrisch positiv definite Matrizen ist ORTHORES mit $\sigma_k = 1$ äquivalent zum Verfahren des steilsten Abstiegs und mit $\sigma_k = \min(k, 2)$ wiederum äquivalent zum CG-Verfahren.

Verfahren, die die euklidische Norm des Residuums minimieren

Wählt man zur Lösung des linearen Gleichungssystem (1.1) ein Orthogonalisierungsverfahren mit $Z = A$, so ist $ZA^{-1} = \mathbf{1}$, und wir erhalten nach Satz 19 und Lemma 5 für die Residuen und Fehler

$$\begin{aligned} ||r_k|| &= \min_{\beta_1,\ldots,\beta_k} ||\sum_{i=1}^{k} \beta_i A^i r_0 + r_0||, \\ ||e_k||_{A^tA} &= \min_{\beta_1,\ldots,\beta_{\sigma_k}} ||\sum_{i=1}^{k} \beta_i A^i e_0 + e_0||_{A^tA} \\ &\leq \kappa(A) \min_{\beta_1,\ldots,\beta_{\sigma_k}} ||\sum_{i=1}^{k} \beta_i A^i e_0 + e_0||. \end{aligned}$$

Es wird also in jedem Schritt die euklidische Norm des Residuums minimiert. Zu den Verfahren, die dies leisten, zählen ORTHOMIN für symmetrisch positiv definite Matrizen, MINRES für symmetrische Matrizen sowie GMRES für allgemeine Matrizen. Kurze Rekursionen sind hierbei in der Regel nicht möglich.

Man gelangt auch mittels Orthogonalisierungsverfahren für die *Normalengleichung erster Art*

$$A^tAx = A^tb \tag{1.51}$$

zu Verfahren, die die euklidische Norm der Residuen minimieren. Zwischen den Residuen r_k^N eines Iterationsverfahrens für (1.51) und den Residuen aus Sicht des ursprünglichen Systems besteht die Beziehung

$$r_k^N = A^t A x_k - A^t b = A^t r_k. \tag{1.52}$$

Betrachtet man ein Orthogonalisierungsverfahren für das Normalensystem (1.51) mit $Z = \mathbf{1}$ – dies entspricht einem gewöhnlichen CG-Verfahren für die Normalengleichung erster Art –, so folgt aus Satz 18

$$||r_k^N||_{A^{-1}A^{-t}} = \min_{\beta_1,\ldots,\beta_k} ||\sum_{i=1}^{k} \beta_i (A^t A)^i r_0^N + r_0^N||_{A^{-1}A^{-t}}.$$

Wir erhalten dann für die Residuen im ursprünglichen System aufgrund von (1.52)

$$||r_k|| = \min_{\beta_1,\ldots,\beta_k} ||\sum_{i=1}^{k} \beta_i (AA^t)^i r_0 + r_0||.$$

Das Verfahren minimiert also die euklidische Norm des entsprechenden Residuums im ursprünglichen System. Für die Fehler erhält man die Abschätzung

$$||e_k|| \leq \kappa(A) \min_{\beta_1,\ldots,\beta_k} ||\sum_{i=1}^{k} \beta_i (A^t A)^i e_0 + e_0||.$$

Eine Implementierung im Rahmen des ursprünglichen Systems mit kurzer Rekursion ist CGNR.

Verfahren, die die euklidische Norm des Fehlers minimieren

Betrachten wir zum linearen Gleichungssystem (1.1) die *Normalengleichung zweiter Art*

$$AA^t y = b, \tag{1.56}$$

so ergibt sich die Lösung x des ursprünglichen Problems durch

$$x = A^t y.$$

Man erhält daher aus den Fehlern eines Iterationsverfahrens für (1.56) als Fehler im ursprünglichen System,

$$e_k = x_k - x = A^t y_k - A^t y = A^t e_k^N. \tag{1.58}$$

Wählt man jetzt $Z = \mathbf{1}$ und betrachtet damit ein Orthogonalisierungsverfahren für das Normalensystem (1.56) – dies entspricht also einem gewöhnlichen CG-Verfahren für die Normalengleichung zweiter Art –, so folgt wiederum aus Satz 18

$$||e_k^N||_{AA^t} = \min_{\beta_1,\ldots,\beta_k} ||\sum_{i=1}^{k} \beta_i (AA^t)^i e_0^N + e_0^N||_{AA^t}.$$

Wir erhalten für die Fehler im ursprünglichen System dann wegen (1.58)

$$||e_k|| = \min_{\beta_1,\ldots,\beta_k} ||\sum_{i=1}^{k} \beta_i (A^t A)^i e_0 + e_0||.$$

Das Verfahren minimiert also die euklidische Norm des entsprechenden Fehlers im ursprünglichen System und ist mit einer kurzen Rekursion implementierbar. Es wird mit CGNE bezeichnet. Die Implementierung im Rahmen des ursprünglichen Systems nennt man auch *Craig's Methode*. Für die Residuen erhält man die Abschätzung

$$||r_k|| \leq \kappa(A) \min_{\beta_1,...,\beta_k} ||\sum_{i=1}^{k} \beta_i (AA^t)^i r_0 + r_0||,$$

die wir der Vollständigkeit halber noch angeben wollen. Es konvergiert CGNE zwar sicher, im allgemeinen jedoch sehr langsam, etwa wenn die Eigenwerte von AA^t stärker in der komplexen Ebene verstreut sind als die von A. Da in die relevante CG-Abschätzung im Falle von CGNE der im Vergleich zur ursprünglichen Konditionszahl quadrierte Wert eingeht, ist die Konvergenz von CGNE insbesondere langsam für Matrizen, die (fast) symmetrisch und schlecht konditioniert sind. Für allgemeine Matrizen ist dies jedoch nicht notwendig der Fall.

Bi-Lanczos-basierende Verfahren

Wir zeigen nun noch wie Verfahren, die in ihrer klassischen Herleitung auf einem Bi-Lanczos-Prozeß zur Biorthogonalisierung des Krylovraums beruhen und daher mit kurzen Rekursionen implementiert werden können, ebenfalls über Orthogonalisierungsverfahren darstellbar sind. Wir besprechen zunächst das BICG-Verfahren, das eine Grundlage für die meisten Bi-Lanczos-basierenden Verfahren bildet. Wir wählen eine Formulierung ausgehend von dem „verdoppelten" System

$$\begin{pmatrix} A & \mathbf{0} \\ \mathbf{0} & A^t \end{pmatrix} \begin{pmatrix} x \\ \bar{x} \end{pmatrix} = \begin{pmatrix} b \\ \bar{b} \end{pmatrix} \quad \Longleftrightarrow: \quad \hat{A}\hat{x} = \hat{b} \tag{1.62}$$

mit einem beliebigen Vektor $\bar{b}$ von gleicher Dimension wie b. Man erhält dann für die Residuen und Fehler eines Krylovraum-Verfahrens bezüglich (1.62)

$$\hat{r}_k = \begin{pmatrix} r_k \\ \bar{r}_k \end{pmatrix} = \begin{pmatrix} Ax_k - b \\ A^t\bar{x}_k - \bar{b} \end{pmatrix} = \begin{pmatrix} p_k(A)r_0 \\ p_k(A^t)\bar{r}_0 \end{pmatrix} = p_k(\hat{A})\hat{r}_0 \in \mathcal{K}_k(\hat{A}, \hat{r}_0), \quad \hat{e}_k = \hat{A}^{-1}\hat{r}_k,$$

wobei p_k ein Polynom vom Grad k mit $p_k(0) = 1$ ist. Das BICG-Verfahren kann nun als Verallgemeinertes CG-Verfahren über dem verdoppelten System mit der Orthogonalisierungsmatrix $\hat{Z} = \begin{pmatrix} \mathbf{0} & \mathbf{1} \\ \mathbf{1} & \mathbf{0} \end{pmatrix}$ angesehen werden. Wir erhalten damit die Orthogonalitätsrelationen

$$\hat{r}_k^T \hat{Z} \hat{r}_{k-i} = 0, \quad \text{d.h.} \quad r_k^T \bar{r}_{k-i} + (\bar{r}_k)^T r_{k-i} = 0 \quad (i = 1, ..., k).$$

Es ist $\hat{Z}\hat{A}^{-1}$ zwar symmetrisch jedoch indefinit, so daß die Konvergenzabschätzungen aus Satz 18 nicht mehr zur Verfügung stehen. Man kann aber zeigen, daß die folgenden beiden schwachen Minimierungsbedingungen erfüllt sind:

$$\begin{aligned} |(\bar{r}_k)^T A^{-1} r_k| &= \min_{p_k \in \mathcal{P}_k} |(\bar{r}_k)^T A^{-1} p_k(A) r_0|, \\ |(\bar{e}_k)^T A e_k| &= \min_{p_k \in \mathcal{P}_k} |(\bar{e}_k)^T A p_k(A) e_0|. \end{aligned}$$

Das BICG-Verfahren liefert im allgemeinen stark oszillierende Residuen und die schwache Minimierungsbedingung für die Iterationsfehler garantiert nicht, daß diese im Verlauf der Iteration gegen Null streben, da aus $(\bar{e}_k)^T A e_k = 0$ nicht notwendig $e_k = \mathbf{0}$ folgt.

Im BICG-Algorithmus werden einzelne Biorthogonalitätsbedingungen der Gestalt $r_k^T \bar{r}_{k-i} = 0$ für $i = 1, ..., k$ verwendet. Diese lassen sich mit Hilfe der Residuenpolynome p_k auch in der Form

$$(r_k, \bar{r}_{k-i}) = (p_k(A)r_0, p_{k-i}(A^t)\bar{r}_0) = (p_{k-i}(A)p_k(A)r_0, \bar{r}_0) = 0$$

schreiben. Berechnet man mittels solcher A^t-freien Skalarprodukte

$$r_k^{\mathrm{CGS}} := p_k^2(A)r_0, \quad e_k^{\mathrm{CGS}} := p_k^2(A)e_0,$$

so gelangt man zum CGS-Verfahren. Falls die Folge $\{p_k(A)\}_k$ der zum BICG-Verfahren gehörenden Matrixpolynome gegen Null konvergiert, so erhält man über die quadrierte Polynomfolge $\{p_k^2(A)\}_k$ eine schnellere Konvergenz des CGS-Verfahrens mit bis zu doppelter Konvergenzgeschwindigkeit im Vergleich zu BICG. Das Konvergenzverhalten von CGS kann aber durchaus noch irreführender sein als beim BICG-Verfahren.

Die Kernidee des BICGSTAB-Verfahrens beruht darauf, die quadrierten Polynome p_k^2 zum CGS-Verfahren durch das Produkt zweier im allgemeinen unterschiedlicher Polynome q_k und p_k zu ersetzen, um so eine glattere Konvergenz und eine stabilere Iteration zu erreichen. Man erhält für die Residuen und Fehler damit

$$r_k^{\mathrm{BICGSTAB}} = q_k p_k(A)r_0, \quad e_k^{\mathrm{BICGSTAB}} = q_k p_k(A)e_0.$$

Die Polynome q_k werden rekursiv definiert durch $q_0(z) := 1$, $q_{j+1}(z) := (\alpha_{j+1}z + 1)q_j(z)$. Der zugehörige Parameter α_{j+1} wird dabei durch das eindimensionale Minimierungsproblem

$$||r_{j+1}^{\mathrm{BICGSTAB}}|| \stackrel{!}{=} \min_{\alpha_{j+1}} ||(\alpha_{j+1}A + \mathbf{1})q_j(A)p_{j+1}(A)r_0||$$

bestimmt. Sinnvolle Polynome q_k lassen sich ebenfalls über l-dimensionale Minimierungsproblem zu gewinnen, was zu den BICGSTAB(l)-Varianten führt.

A.2.4 Bemerkungen zur Vorkonditionierung

Die Vorkonditionierung linearer Gleichungssysteme kann auch als Transformation der Gleichungssysteme in geeignetere Koordinaten verstanden werden. Sie muß sich dabei nicht nur an dem vorliegenden Problem sondern auch an den Eigenschaften des vorzukonditionierenden Verfahrens orientieren. Es ist außerdem die Frage zu stellen, was ein vorkonditioniertes und damit (hoffentlich) schneller konvergierendes Verfahren für die Reduktion des wirklichen Fehlers, das heißt für die Reduktion des Fehlers aus Sicht des ursprünglichen Koordinatensystems bedeutet. Wir wollen diese Problematik am Beispiel des CG- und GMRES-Verfahrens verdeutlichen.

Verfahren, die $||e_k||_A$ minimieren

Es sei A symmetrisch positiv definit und M eine ebenfalls symmetrisch positiv definite Matrix, die als Vorkonditionierer für A dient. Dann erhalten wir als Fehlerabschätzungen für das CG-Verfahren im ursprünglichen und mittels M transformierten System

$$\begin{aligned} ||e_k||_A &\leq 2\left(\frac{\sqrt{\kappa(A)}-1}{\sqrt{\kappa(A)}+1}\right)^k ||e_0||_A, \\ ||e_k^P||_{M^{\frac{1}{2}}AM^{\frac{1}{2}}} &\leq 2\left(\frac{\sqrt{\kappa(M^{\frac{1}{2}}AM^{\frac{1}{2}})}-1}{\sqrt{\kappa(M^{\frac{1}{2}}AM^{\frac{1}{2}})}+1}\right)^k ||e_0^P||_{M^{\frac{1}{2}}AM^{\frac{1}{2}}}, \end{aligned} \tag{1.63}$$

wobei $e_j^P = M^{-\frac{1}{2}} e_j$ ist. Um eine schnellere Konvergenz im transformierten System zu erhalten, sollte die Transformation M daher so gewählt werden, daß gilt:

$$\kappa(M^{\frac{1}{2}} A M^{\frac{1}{2}}) < \kappa(A).$$

Wir fragen nun, was das vorkonditionierte Verfahren für die Reduktion des wirklichen Fehlers bedeutet. Eine zentrale Eigenschaft des vorkonditionierten CG-Verfahrens ist, daß trotz der Transformation des Systems durch den Vorkonditionierer immer noch die ursprünglichen Fehler in der ursprünglichen Energienorm minimiert werden. Wir erhalten aus (1.63) die Abschätzung

$$||e_k||_A \leq 2 \left(\frac{\sqrt{\kappa(M^{\frac{1}{2}} A M^{\frac{1}{2}})} - 1}{\sqrt{\kappa(M^{\frac{1}{2}} A M^{\frac{1}{2}})} + 1} \right)^k ||e_0||_A.$$

Mißt man den aus dem vorkonditionierten Verfahren stammenden ursprünglichen Fehler daher in einer problemunabhängigen Norm wie der euklidischen, so kommt die Kondition $\kappa(A^{\frac{1}{2}})$, die zum ursprünglichen Problem und nicht zum vorkonditionierten Problems gehört, ins Spiel. Es ist also vorstellbar, daß trotz guter Vorkonditionierung und damit schneller Konvergenz, ein PCG-Verfahren für schlecht konditionierte Probleme keine verläßliche Reduktion des tatsächlichen Fehlers mit sich bringt.

Verfahren, die $||r_k||$ minimieren

Die einfache Abschätzung der Residuen für das GMRES-Verfahren lautet

$$||r_k|| \leq \min_{p_k \in \mathcal{P}_k} ||p_k(A) r_0||. \tag{1.64}$$

Hieraus können wir wie folgt eine gröbere Abschätzung herleiten, die zur Definition eines geeigneten Vorkonditionierers für GMRES-Verfahren hilfreich ist. Es sei $||\mathbf{1} - A|| = \rho < 1$ und $\tilde{p}_k(z) := (1-z)^k \in \mathcal{P}_k$. Dann gilt aufgrund von Abschätzung (1.64)

$$||r_k|| \leq ||\tilde{p}_k(A) r_0|| \leq \rho^k ||r_0||. \tag{1.65}$$

Ein GMRES-Verfahren angewandt auf das durch eine nichtsinguläre Matrix M von links transformierte System $MAx = Mb$ minimiert die euklidische Norm der transformierten Residuen $r_k^P = M r_k$. Um eine schnellere Konvergenz im transformierten System zu erhalten, sollte gemäß Abschätzung (1.65) daher die Transformation M so gewählt werden, daß gilt:

$$||\mathbf{1} - MA|| < \rho.$$

Für das links-vorkonditionierte GMRES-Verfahren erhalten wir als Kopplung der transformierten Residuen und ursprünglichen Fehler

$$\frac{||e_k||}{||e_0||} \leq \kappa(MA) \frac{||r_k^P||}{||r_0^P||},$$

da aufgrund der Linkstransformation transformierte und ursprüngliche Fehler übereinstimmen. Falls $\kappa(MA) < \kappa(A)$ ist, erwarten wir, daß die transformierten Residuen besser zur Steuerung der Iteration geeignet sind als die ursprünglichen Residuen.

A.3 Konjugierte Krylovraum-Verfahren im Standardsystem

Die Ausführungen in Kapitel 4 haben gezeigt, daß multiplikative und additive Multiskalen-Verfahren als klassische Iterationsverfahren (Gauß–Seidel-, Jacobi-Verfahren) bezüglich des Standardsystems interpretiert und mit Hilfe der Nichtstandardform effizient implementiert werden können. In diesem Abschnitt betrachten wir exakte Konjugierte Krylovraum-Verfahren für das linksvorkonditionierte Standardsystem

$$\begin{aligned} \tilde{M}_{0,s}^{-1}(W_{lt,\mathcal{S}}^0)^t T_0 W_{lt,\mathcal{V}}^0 \tilde{u}_{0,s} &= \tilde{M}_{0,s}^{-1}(W_{lt,\mathcal{S}}^0)^t f_0 \\ \iff \qquad \tilde{M}_{0,s}^{-1}\tilde{T}_{0,s}\tilde{u}_{0,s} &= \tilde{M}_{0,s}^{-1}\tilde{f}_{0,s}. \end{aligned} \tag{1.66}$$

Hierbei bezeichnet $\tilde{M}_{0,s}^{-1}$ eine approximative Inverse der Standardform $\tilde{T}_{0,s}$. Wir untersuchen zunächst ausführlich, wie die darüber gewonnenen CKS-Verfahren äquivalent aus Sicht des ursprünglichen Gleichungssystems mittels transformierter Suchrichtungen $q_{k-i,k}$ und transformierter Orthogonalisierungsmatrizen Z_k formuliert werden können. Dabei sehen wir, daß die so definierten Verfahren auch als CKS-Verfahren für die von links durch

$$M_{0,s}^{-1} := W_{lt,\mathcal{V}}^0 \tilde{M}_{0,s}^{-1}(W_{lt,\mathcal{S}}^0)^t$$

Multiskalen-vorkonditionierte ursprüngliche Gleichung aufgefaßt werden kann. Unter den Voraussetzungen von Satz 8, ist eine äquivalente Realisierung des Multiskalen-Vorkonditionierers $M_{0,s}^{-1}$ mit Hilfe des Nichtstandardsystems möglich. Damit erhalten wir dann eine meist effizientere Implementierung der beabsichtigten CKS-Verfahren als direkt über das linksvorkonditionierte Standardsystem. Der Übersichtlichkeit halber verzichten wir im folgenden auf die Levelbezeichnungen und kennzeichnen Objekte, die mittels des Standardsystems definiert sind, lediglich durch einen unteren Index $_s$. Die folgende Diskussion kann selbstverständlich auch für das rechts- oder symmetrisch vorkonditionierte Standardsystem geführt werden.

Wir betrachten Krylovraum-Verfahren für das linksvorkonditionierte standard transformierte Gleichungssystem (1.66) mit

$$\begin{aligned} B &= \tilde{M}_s^{-1}\tilde{T}_s, \\ z &= \tilde{r}_s^{0,P} := \tilde{M}_s^{-1}\tilde{r}_s^0 = \tilde{M}_s^{-1}(W_{s,\mathcal{S}})^t r^0, \end{aligned}$$

so daß die Suchrichtungen $\tilde{q}_{s,k-i,k}$ im Krylovraum $\mathcal{K}_{k-i}(\tilde{M}_s^{-1}\tilde{T}_s, \tilde{r}_s^{0,P})$ liegen. Wegen

$$\tilde{r}_s^{k,P} \in \tilde{r}_s^{0,P} + (\tilde{M}_s^{-1}\tilde{T}_s)\mathcal{K}_{k-1}(\tilde{M}_s^{-1}\tilde{T}_s, \tilde{r}_s^{0,P})$$

existiert eine Darstellung

$$\tilde{r}_s^{k,P} = \sum_{j=1}^{k} \beta_{j,k}(\tilde{M}_s^{-1}\tilde{T}_s)^j \tilde{r}_s^{0,P} + \tilde{r}_s^{0,P}.$$

Betrachten wir die Suchrichtungen

$$\tilde{q}_{s,k-i,k} := (\tilde{M}_s^{-1}\tilde{T}_s)^{k-i}\tilde{r}_s^{0,P} \quad (i = 1, ..., (\sigma_k = k)),$$

so erhalten wir mit Hilfe nichtsingulärer Orthogonalisierungsmatrizen $\tilde{Z}_k$ die Orthogonalisierungsbedingungen

$$(\tilde{r}_s^{k,P})^t \tilde{Z}_k (\tilde{M}_s^{-1}\tilde{T}_s)^{k-i}\tilde{r}_s^{0,P} = 0 \quad (i = 1, ..., (\sigma_k = k)).$$

Sie schreiben sich für $i = 1, ..., (\sigma_k = k)$ auch ausführlicher als

$$\sum_{j=1}^{k} \beta_{j,k} (\tilde{r}_s^{0,P})^t [(\tilde{M}_s^{-1}\tilde{T}_s)^j]^t \tilde{Z}_k (\tilde{M}_s^{-1}\tilde{T}_s)^{k-i} \tilde{r}_s^{0,P} = -(\tilde{r}_s^{0,P})^t \tilde{Z}_k (\tilde{M}_s^{-1}\tilde{T}_s)^{k-i} \tilde{r}_s^{0,P}.$$

Für Gleichung i dieses Systems erhält man dann

$$\begin{aligned}
\sum_{j=1}^{k} \beta_{j,k} (r^0)^t W_{s,\mathcal{S}} \tilde{M}_s^{-t} \cdot & \quad (W_{s,\mathcal{V}})^t T^t W_{s,\mathcal{S}} \tilde{M}_s^{-t} \\
\cdot & \quad \vdots \qquad (j \text{ mal}) \\
\cdot & \quad (W_{s,\mathcal{V}})^t T^t W_{s,\mathcal{S}} \tilde{M}_s^{-t} \\
\cdot & \quad \tilde{Z}_k \\
\cdot & \quad \tilde{M}_s^{-1} (W_{s,\mathcal{S}})^t T W_{s,\mathcal{V}} \\
\cdot & \quad \vdots \qquad (k-i \text{ mal}) \\
\cdot & \quad \tilde{M}_s^{-1} (W_{s,\mathcal{S}})^t T W_{s,\mathcal{V}} \quad \cdot \tilde{M}_s^{-1} (W_{s,\mathcal{S}})^t r^0 \\
= -(r^0)^t W_{s,\mathcal{S}} \tilde{M}_s^{-t} \cdot & \quad \tilde{Z}_k \\
\cdot & \quad \tilde{M}_s^{-1} (W_{s,\mathcal{S}})^t T W_{s,\mathcal{V}} \\
\cdot & \quad \vdots \qquad (k-i \text{ mal}) \\
\cdot & \quad \tilde{M}_s^{-1} (W_{s,\mathcal{S}})^t T W_{s,\mathcal{V}} \quad \cdot \tilde{M}_s^{-1} (W_{s,\mathcal{S}})^t r^0.
\end{aligned} \tag{1.67}$$

Im Abschnitt über Orthogonalisierungsverfahren bezüglich vorkonditionierter Gleichungssysteme konnten wir die Verfahren alleine anhand der dafür geltenden Abschätzungen auch aus Sicht des ursprünglichen Gleichungssystems mit Hilfe transformierter Suchrichtungen und transformierter Orthogonalisierungsbedingungen interpretieren. Wir wollen dies nun für CKS-Verfahren bezüglich des linksvorkonditionierten Standardsystems untersuchen und betrachten dazu gemäß (1.21) die transformierten Suchrichtungen

$$\begin{aligned}
q_{k-i,k} &= W_{s,\mathcal{V}} \tilde{q}_{s,k-i,k} \\
&= W_{s,\mathcal{V}} [\tilde{M}_s^{-1} (W_{s,\mathcal{S}})^t T W_{s,\mathcal{V}}]^{k-i} \tilde{M}_s^{-1} (W_{s,\mathcal{S}})^t r^0 \\
&= [W_{s,\mathcal{V}} \tilde{M}_s^{-1} (W_{s,\mathcal{S}})^t T]^{k-i} W_{s,\mathcal{V}} \tilde{M}_s^{-1} (W_{s,\mathcal{S}})^t r^0
\end{aligned}$$

und gemäß (1.22) die transformierten Orthogonalisierungsmatrizen

$$Z_k = W_{s,\mathcal{S}} \tilde{M}_s^{-t} \tilde{Z}_k (W_{s,\mathcal{V}})^{-1}.$$

Durch Rücktransformation eines Residuenvektors

$$\tilde{r}_s^{k,P} = \sum_{j=1}^{k} \gamma_{j,k} (\tilde{M}_s^{-1}\tilde{T}_s)^j \tilde{r}_s^{0,P} + \tilde{r}_s^{0,P}$$

bezüglich des vorkonditionierten Standardsystems mittels $W_{s,\mathcal{V}}$ erhalten wir

$$W_{s,\mathcal{V}} \tilde{M}_s^{-1} (W_{s,\mathcal{S}})^t r^k = \sum_{j}^{k} \gamma_{j,k} [W_{s,\mathcal{V}} \tilde{M}_s^{-1} (W_{s,\mathcal{S}})^t T]^j W_{s,\mathcal{V}} \tilde{M}_s^{-1} (W_{s,\mathcal{S}})^t r^0 + W_{s,\mathcal{V}} \tilde{M}_s^{-1} (W_{s,\mathcal{S}})^t r^0$$

und es folgt die Darstellung

$$\begin{aligned} r^k &= (W_{s,\mathcal{S}})^{-t}\tilde{M}_s(W_{s,\mathcal{V}})^{-1} \\ &\quad \cdot[\sum_{j=1}^{k}\gamma_{j,k}[W_{s,\mathcal{V}}\tilde{M}_s^{-1}(W_{s,\mathcal{S}})^tT]^jW_{s,\mathcal{V}}\tilde{M}_s^{-1}(W_{s,\mathcal{S}})^tr^0 + W_{s,\mathcal{V}}\tilde{M}_s^{-1}(W_{s,\mathcal{S}})^tr^0] \end{aligned}$$

aus Sicht des ursprünglichen Systems. Fordert man nun die Orthogonalisierungsbedingungen

$$(r^k)^tZ_kq_{k-i,k} = 0 \quad (i = 1, ..., (\sigma_k = k)),$$

so erhalten wir für Gleichung i

$$\begin{aligned} \sum_{j=1}^{k}\gamma_{j,k}(r^0)^tW_{s,\mathcal{S}}\tilde{M}_s^{-t}(W_{s,\mathcal{V}})^t\cdot &\quad T^tW_{s,\mathcal{S}}\tilde{M}_s^{-t}(W_{s,\mathcal{V}})^t \\ \cdot &\quad \vdots \quad (j \text{ mal}) \\ \cdot &\quad T^tW_{s,\mathcal{S}}\tilde{M}_s^{-t}(W_{s,\mathcal{V}})^t \\ \cdot &\quad (W_{s,\mathcal{V}})^{-t}\tilde{M}_s^t(W_{s,\mathcal{S}})^{-1} \\ \cdot &\quad W_{s,\mathcal{S}}\tilde{M}_s^{-t}\tilde{Z}_k(W_{s,\mathcal{V}})^{-1} \\ \cdot &\quad W_{s,\mathcal{V}}\tilde{M}_s^{-1}(W_{s,\mathcal{S}})^tT \\ \cdot &\quad \vdots \quad (k-i \text{ mal}) \\ \cdot &\quad W_{s,\mathcal{V}}\tilde{M}_s^{-1}(W_{s,\mathcal{S}})^tT \quad \cdot W_{s,\mathcal{V}}\tilde{M}_s^{-1}(W_{s,\mathcal{S}})^tr^0 \\ = -(r^0)^tW_{s,\mathcal{S}}\tilde{M}_s^{-t}(W_{s,\mathcal{V}})^t &\quad \cdot(W_{s,\mathcal{V}})^{-t}\tilde{M}_s^t(W_{s,\mathcal{S}})^{-1} \\ \cdot &\quad W_{s,\mathcal{S}}\tilde{M}_s^{-t}\tilde{Z}_k(W_{s,\mathcal{V}})^{-1} \\ \cdot &\quad W_{s,\mathcal{V}}\tilde{M}_s^{-1}(W_{s,\mathcal{S}})^tT \\ \cdot &\quad \vdots \quad (k-i \text{ mal}) \\ \cdot &\quad W_{s,\mathcal{V}}\tilde{M}_s^{-1}(W_{s,\mathcal{S}})^tT \quad \cdot W_{s,\mathcal{V}}\tilde{M}_s^{-1}(W_{s,\mathcal{S}})^tr^0. \end{aligned} \tag{1.68}$$

Der direkte Vergleich von (1.68) mit (1.67) zeigt, daß man beide Male das gleiche lineare Gleichungssystem zur Bestimmung der Koeffizienten $\beta_{j,k}$ und $\gamma_{j,k}$ erhält. Es gilt damit in jedem Iterationsschritt $\beta_{j,k} = \gamma_{j,k}$ für $j = 1, ..., (\sigma_k = k)$. Man beweist durch Induktion, daß damit auch die Iterierten $\tilde{x}_s^{k,P}$ im vorkonditionierten Standardsystem nach Rücktransformation mittels $W_{s,\mathcal{V}}$ jeweils mit den Iterierten x^k, die man über das CKS-Verfahren aus Sicht des ursprünglichen Systems mit den entsprechend transformierten Suchrichtungen und Orthogonalisierungsmatrizen erhält, übereinstimmen.

Das eben beschriebene Vorgehen kann auch als CKS-Verfahren für die von links mittels

$$M_s^{-1} := W_{s,\mathcal{V}}\tilde{M}_s^{-1}(W_{s,\mathcal{S}})^t$$

Multiskalen-vorkonditionierte ursprüngliche Gleichung interpretiert werden. Man betrachte dazu ein Krylovraum-Verfahren für das so linksvorkonditionierte Gleichungssystem mit

$$\begin{aligned} B &= M_s^{-1}T, \\ z &= r^{0,P} := M_s^{-1}r^0, \end{aligned}$$

so daß die Suchrichtungen $q_{k-i,k}$ im Krylovraum $\mathcal{K}_{k-i}(M_s^{-1}T, r^{0,P})$ liegen, und wähle als Orthogonalisierungsmatrizen

$$Z_k = W_{s,\mathcal{S}}\tilde{M}_s^{-t}\tilde{Z}_k(W_{s,\mathcal{V}})^{-1}.$$

Fordert man nunmehr die Z_k-Orthogonalität der Multiskalen-vorkonditionierten Residuen aus Sicht des ursprünglichen Systems $r^k = M_s r^{k,P}$ bezüglich des Krylovraums $\mathcal{K}_{k-i}(M_s^{-1}T, r^{0,P})$, so erhält man wieder die Bestimmungsgleichungen (1.68) und die gleichen Iterierten x^k. Will man hingegen Orthogonalitätsbedingungen direkt für die Multiskalen-vorkonditionierten Residuen $r^{k,P}$ aufstellen, so muß man wegen

$$\begin{aligned}(r^k)^t Z_k &= [(r^{k,P})^t (W_{s,\mathcal{V}})^{-t} \tilde{M}_s^t (W_{s,\mathcal{S}})^{-1}][W_{s,\mathcal{S}} \tilde{M}_s^{-t} \tilde{Z}_k (W_{s,\mathcal{V}})^{-1}] \\ &= (r^{k,P})^t (W_{s,\mathcal{V}})^{-t} \tilde{Z}_k (W_{s,\mathcal{V}})^{-1}\end{aligned}$$

dann

$$Z_k^M := (W_{s,\mathcal{V}})^{-t} \tilde{Z}_k (W_{s,\mathcal{V}})^{-1}.$$

als Orthogonalisierungsmatrizen wählen. Wir erhalten somit

Satz 22 (Äquivalenz von (vorkonditionierten) CKS-Verfahren im Standardsystem zu Multiskalen-vorkonditionierten CKS-Verfahren)
Ist $\tilde{M}_s^{-1}$ eine approximative Inverse der Standardform $\tilde{T}_s$ und betrachtet man CKS-Verfahren im damit linksvorkonditionierten Standardsystem, so sind diese äquivalent zu den entsprechenden CKS-Verfahren für das mittels $M_s^{-1} := W_{s,\mathcal{V}} \tilde{M}_s^{-1} (W_{s,\mathcal{S}})^t$ Multiskalen-linksvorkonditionierte ursprüngliche Gleichungssystem, wenn man für letztere als Orthogonalisierungsmatrizen $Z_k^M := (W_{s,\mathcal{V}})^{-t} \tilde{Z}_k (W_{s,\mathcal{V}})^{-1}$ wählt.

Beweis:
Der Beweis folgt aus der unmittelbar vorangehenden Diskussion.

□

A.4 Additiv Multiskalen-vorkonditionierte Verallgemeinerte Krylovraum-Verfahren

A.4.1 Eine spezielle Normäquivalenz

Im letzten Abschnitt haben wir explizit die Äquivalenz von (vorkonditionierten) Konjugierten Krylovraum-Verfahren (CKS-Verfahren) im Standardsystem zu Multiskalen-vorkonditionierten CKS-Verfahren für das ursprüngliche lineare Gleichungssystem gezeigt. Anhand der durch die Sätze 20 und 21 im vorletzten Abschnitt bereitgestellten Konvergenzaussagen untersuchen wir jetzt additiv Multiskalen-vorkonditionierte Verallgemeinerte Krylovraum-Verfahren für das Konvektions-Diffusions Problem

$$Tu = -\triangle u + \vec{b} \cdot \nabla u + cu = f \quad \text{in } \Omega =]0,1[^2 \quad \text{und } u = 0 \quad \text{auf } \partial\Omega \tag{1.69}$$

mit hinreichend glatten Koeffizientenfunktionen und rechter Seite. Wir starten von der gewöhnlichen schwachen Formulierung mittels des Sobolevraums $\mathcal{H}_0^1(\Omega)$ und zerlegen die zugehörige Bilinearform $a(v,w) := \int_\Omega \nabla v \cdot \nabla w + (\vec{b} \cdot \nabla v) w + cvw \, dx$ gemäß

$$a^{sy}(v,w) := \frac{1}{2}\Big[a(v,w) + a(w,v)\Big] \quad \text{und} \quad a^{sk}(v,w) := \frac{1}{2}\Big[a(v,w) - a(w,v)\Big]$$

in ihren symmetrischen und schiefsymmetrischen Anteil. Partielle Integration ergibt

$$a^{sy}(v,w) = \int_\Omega \nabla v \cdot \nabla w + \Big[c - \frac{1}{2}(\nabla \cdot \vec{b})\Big] vw\, dx, \tag{1.71}$$

$$a^{sk}(v,w) = \int_\Omega (\vec{b} \cdot \nabla v) w + \frac{1}{2}\Big(\nabla \cdot \vec{b}\Big)\, vw dx. \tag{1.72}$$

Um Schwierigkeiten, die durch einen indefiniten symmetrischen Anteil entstehen, zu vermeiden, nehmen wir an, daß $\nabla \cdot \vec{b} = 0$ und $c \geq 0$ auf Ω sind. Dies gilt in vielen praktisch relevanten Fällen und kann sogar ohne Beschränkung der Allgemeinheit angenommen werden [86]. Ferner erfülle die wegen $a^{sy}(v,v) = a(v,v)$ über den symmetrischen Anteil der Bilinearform definierbare *verallgemeinerte Energienorm* $||v||_a := a(v,v)$ die elementare Normäquivalenz

$$||v||_a \approx ||v||_1^2 \quad \forall v \in \mathcal{H}_0^1(\Omega). \tag{1.73}$$

Für den schiefsymmetrischen Anteil gelte die Abschätzung

$$|a^{sk}(v,w)| \leq \gamma ||v||_1 ||w||_0 \quad \forall v, w \in \mathcal{H}_0^1(\Omega). \tag{1.74}$$

Beides folgt unter geeigneten Voraussetzungen an c und $\vec{b}$ sofort aus den Darstellungen (1.71) und (1.72). Die Konstante γ ist hierbei offensichtlich von der Konvektion $\vec{b}$ abhängig.

Zur Diskretisierung mittels einer entspechenden diskreten schwachen Formulierung betrachten wir ein Galerkin-Verfahren, bei dem also Ansatz- und Testraum $\mathcal{V}_0$ und $\mathcal{S}_0$ gleich sind, und gelangen zu einem linearen Gleichungssystem der Form

$$T_0 u_0 = f_0. \tag{1.75}$$

Die Voraussetzung $\mathcal{S}_0 = \mathcal{V}_0$ ist notwendig, um überhaupt sinnvoll über den symmetrischen Anteil der Bilinearform eine problemabhängige Norm definieren zu können. Für die Lösung von (1.75) mit Hilfe eines additiven Multiskalen-Vorkonditionierers nehmen wir weiter an, daß die Multiskalen-Zerlegungen der Ansatz- und Testseite gleich sind. Obwohl ein solcher Vorkonditionierer ebenfalls im Falle unterschiedlicher Zerlegungen der Ansatz- und Testseite definiert werden kann (siehe Kapitel 4.3), ist die Annahme einer Galerkin-artigen Multiskalen-Zerlegung für unsere Analyse notwendig. Wir werden auf mögliche Erweiterungen auf den allgemeineren Petrov–Galerkin-Fall und dabei auftretende Schwierigkeiten jedoch hinweisen. Deshalb unterscheiden wir im folgenden auch die entsprechenden Transformationsmatrizen für die Ansatz- und Testseite.

Approximieren wir die Blöcke A_k^{-1} der Nichtstandardform $\tilde{T}_{0,ns}$ der Steifigkeitsmatrix T_0 durch Diagonalmatrizen $\breve{A}_k^{-1} = h_k^2 \cdot \mathbf{1}$ für $k = 1, ..., lt$ und T_{lt}^{-1} durch $h_{lt}^2 \cdot \mathbf{1}$, wobei h_k die charakteristische Maschenweite zur Skala k bedeutet, so ergeben sich als additiver Multiskalen-Vorkonditionierer $M_{0,s}^{-1}$ und seine Inverse $M_{0,s}$ (siehe (4.39))

$$M_{0,s}^{-1} = \sum_{k=1}^{lt} Q_{k,\mathcal{V}}^0 h_k^2 (Q_{k,\mathcal{S}}^0)^t + P_{lt,\mathcal{V}}^0 h_{lt}^2 (P_{lt,\mathcal{S}}^0)^t, \tag{1.76}$$

$$M_{0,s} = \sum_{k=1}^{lt} (Q_{k,\mathcal{S}}^0)^{-t} h_k^{-2} Q_{0,\mathcal{V}}^k + (P_{lt,\mathcal{S}}^0)^{-t} h_{lt}^{-2} P_{0,\mathcal{V}}^{lt}. \tag{1.77}$$

Damit definieren wir die beiden Skalarprodukte und Multiskalen-Normen

$$||r_0||_{M_{0,s}^{-1}} := (M_{0,s}^{-1} r_0, r_0) = h_{lt}^2 ||(P_{lt,\mathcal{S}}^0)^t r_0||^2 + \sum_{k=1}^{lt} h_k^2 ||(Q_{k,\mathcal{S}}^0)^t r_0||^2, \tag{1.78}$$

$$||z_0||_{M_{0,s}} := (M_{0,s} z_0, z_0) = h_{lt}^{-2} ||P_{0,\mathcal{V}}^{lt} z_0||^2 + \sum_{k=1}^{lt} h_k^{-2} ||Q_{0,\mathcal{V}}^k z_0||^2. \tag{1.79}$$

Wir setzen ferner voraus, daß eine Normäquivalenz zwischen der verallgemeinerten diskreten Energienorm und der durch $M_{0,s}$ induzierten Multiskalen-Norm besteht:

$$ca(u_0, u_0) \leq h_{lt}^{-2} ||P_{0,\mathcal{V}}^{lt} u_0||^2 + \sum_{k=1}^{lt} h_k^{-2} ||Q_{0,\mathcal{V}}^k u_0||^2 \leq Ca(u_0, u_0). \tag{1.80}$$

Die Konstanten seien unabhängig von der Maschenweite h_0 des feinen Gitters, der Anzahl lt der bei der Zerlegung betrachteten Skalen. Wegen $\nabla \cdot \vec{b} = 0$ auf Ω sind die Konstanten ebenfalls unabhängig von der Konvektion $\vec{b}$. Für Beweise einer solchen Normäquivalenz im Fall Wavelet-artiger Multiskalen-Zerlegungen verweisen wir auf [20, 33, 90, 127, 131]. Für zweidimensionale Probleme, die durch eine uniforme Verfeinerung entstanden sind, hängt im Fall hierarchischer Multiskalen-Zerlegungen jedoch die Konstante C der oberen Abschätzung von (1.80) typischerweise quadratisch von der Anzahl der betrachteten Level lt ab. Dies führt zu einer logarithmischen Maschenweitenabhängigkeit der Konvergenzraten zugehöriger hierarchischer Multiskalen-Löser [129, 131].

Der folgende auf [108] zurückgehende Satz zeigt, daß die verallgemeinerte Energienorm des Fehlers einer Approximation an die exakte Lösung des diskreten Problems (1.75) äquivalent zur durch den additiven Multiskalen-Vorkonditionierer (1.76) induzierten Norm des zugehörigen Residuums ist. Ist e_0 ein Iterationsfehler und $r_0 = T_0 e_0$ das entsprechende Residuum, so wird die behauptete Normäquivalenz

$$||r_0||_{M_{0,s}^{-1}} = (M_{0,s}^{-1} r_0, r_0) = (M_{0,s}^{-1} T_0 e_0, T_0 e_0) \approx (T_0 e_0, e_0) = a(e_0, e_0)$$

verständlich, wenn man bedenkt, daß $M_{0,s}^{-1}$ eine approximative Inverse zu T_0 darstellt.

Satz 23 (Verallgemeinete Energienorm des Fehlers ≈ MS-Norm des Residuums) *Unter den in diesem Abschnitt gemachten Voraussetzungen sei $e_0 \in \mathcal{V}_0$ der Fehler einer Approximation an die exakte Lösung des diskreten Problems (1.75) und $r_0 = T_0 e_0$ das zugehörige Residuum. Dann gelten die Abschätzungen*

$$\frac{c}{1+\gamma^2} \Big(h_{lt}^2 ||(P_{lt,\mathcal{S}}^0)^t r_0||^2 + \sum_{k=1}^{lt} h_k^2 ||(Q_{k,\mathcal{S}}^0)^t r_0||^2 \Big) \leq a(e_0, e_0), \tag{1.81}$$

$$a(e_0, e_0) \leq C \Big(h_{lt}^2 ||(P_{lt,\mathcal{S}}^0)^t r_0||^2 + \sum_{k=1}^{lt} h_k^2 ||(Q_{k,\mathcal{S}}^0)^t r_0||^2 \Big), \tag{1.82}$$

wobei die Konstanten c und C unabhängig von der Feingittermaschenweite h_0, der Anzahl lt der betrachteten Skalen sowie unabhängig von γ sind.

Beweis:
Wir unterscheiden im folgenden Beweis trotz der Annahme gleicher Zerlegungen der Ansatz- und Testseite die entsprechenden Transformationsmatrizen, um die Schwierigkeiten zu verdeutlichen, die beim Versuch einer Übertragung auf den allgemeineren Petrov–Galerkin-Fall auftreten. Die kritischen Stellen sind durch $(*)$ und $(**)$ markiert.
Wir zeigen zuerst die zweite Ungleichung. Aufgrund der Cauchy–Schwarzschen Ungleichung und der rechten Seite der Normäquivalenz (1.80) gilt

$$\begin{aligned}
a(e_0,e_0) &= (T_0e_0,e_0) = (r_0,e_0)\\
&= ((W^0_{lt,\mathcal{S}})^{-t}(W^0_{lt,\mathcal{S}})^t r_0, e_0) = ((W^0_{lt,\mathcal{S}})^t r_0, (W^0_{lt,\mathcal{S}})^{-1}e_0)\\
&= ((W^0_{lt,\mathcal{S}})^t r_0, W^{lt}_{0,\mathcal{S}}e_0) \overset{(*)}{=} ((W^0_{lt,\mathcal{S}})^t r_0, W^{lt}_{0,\mathcal{V}}e_0)\\
&= (h_{lt}(P^0_{lt,\mathcal{S}})^t r_0, h_{lt}^{-1}P^{lt}_{0,\mathcal{V}}e_0) + \sum_{k=1}^{lt}(h_k(Q^0_{k,\mathcal{S}})^t r_0, h_k^{-1}Q^k_{0,\mathcal{V}}e_0)\\
&\le \Big(h_{lt}^2||(P^0_{lt,\mathcal{S}})^t r_0||^2 + \sum_{k=1}^{lt} h_k^2||(Q^0_{k,\mathcal{S}})^t r_0||^2\Big)^{\frac12}\cdot\Big(h_{lt}^{-2}||P^{lt}_{0,\mathcal{V}}e_0||^2 + \sum_{k=1}^{lt} h_k^{-2}||Q^k_{0,\mathcal{V}}e_0||^2\Big)^{\frac12}\\
&\le C\Big(h_{lt}^2||(P^0_{lt,\mathcal{S}})^t r_0||^2 + \sum_{k=1}^{lt} h_k^2||(Q^0_{k,\mathcal{S}})^t r_0||^2\Big)^{\frac12}\cdot a(e_0,e_0)^{\frac12}.
\end{aligned}$$

Hieraus folgt (1.82). Die erste Ungleichung erhält man wegen

$$\begin{aligned}
h_{lt}^2||(P^0_{lt,\mathcal{S}})^t r_0||^2 + \sum_{k=1}^{lt} h_k^2||(Q^0_{k,\mathcal{S}})^t r_0||^2 &\overset{(**)}{=} (\underbrace{M_{0,s}^{-1}r_0}_{=:z_0}, r_0)\\
&= a(e_0,z_0) = a^{sy}(e_0,z_0) + a^{sk}(e_0,z_0)\\
&\overset{i)}{\le} a^{sy}(e_0,e_0)^{\frac12}a^{sy}(z_0,z_0)^{\frac12} + \gamma||e_0||_1||z_0||_1\\
&\overset{ii)}{\le} a^{sy}(e_0,e_0)^{\frac12}a^{sy}(z_0,z_0)^{\frac12} + C\gamma a(e_0,e_0)^{\frac12}a(z_0,z_0)^{\frac12}\\
&\overset{iii)}{\le} C(1+C\gamma)a(e_0,e_0)^{\frac12}\cdot\Big(h_{lt}^2||(P^0_{lt,\mathcal{S}})^t r_0||^2 + \sum_{k=1}^{lt} h_k^2||(Q^0_{k,\mathcal{S}})^t r_0||^2\Big)^{\frac12}.
\end{aligned}$$

Abschätzung i) ergibt sich hierbei durch Anwenden der Cauchy–Schwarzschen Ungleichung und Ausnutzen der Abschätzung (1.74) des schiefsymmetrischen Anteils. Abschätzung ii) gilt aufgrund der Normäquivalenz (1.73). Die zugehörigen Konstanten hängen unter den geforderten Voraussetzungen bis auf γ nicht von der Stärke des unsymmetrischen Anteils ab. Abschätzung iii) sieht man wie folgt: Wegen $z_0 = P^0_{lt,\mathcal{V}}h_{lt}^2(P^0_{lt,\mathcal{S}})^t r_0 + \sum_{k=1}^{lt} Q^0_{k,\mathcal{V}}h_k^2(Q^0_{k,\mathcal{S}})^t r_0$ gilt $P^{lt}_{0,\mathcal{V}}z_0 = h_{lt}^2(P^0_{lt,\mathcal{S}})^t r_0$ sowie $Q^k_{0,\mathcal{V}}z_0 = h_k^2(Q^0_{k,\mathcal{S}})^t r_0$ für alle $k = 1,...,lt$. Aufgrund der Normäquivalenz (1.80) erhalten wir dann

$$\begin{aligned}
a(z_0,z_0) &\le C\Big(h_{lt}^{-2}||P^{lt}_{0,\mathcal{V}}z_0||^2 + \sum_{k=1}^{lt} h_k^{-2}||Q^k_{0,\mathcal{V}}z_0||^2\Big)\\
&= C\Big(h_{lt}^2||(P^0_{lt,\mathcal{S}})^t r_0||^2 + \sum_{k=1}^{lt} h_k^2||(Q^0_{k,\mathcal{S}})^t r_0||^2\Big).
\end{aligned}$$

□

A.4.2 Verfahren, die $||r_k||_{M_{0,s}^{-1}}$ minimieren

Wir zeigen jetzt, daß die Konvergenzraten additiv Multiskalen-vorkonditionierter Verallgemeinerter Krylovraum-Verfahren, die die durch den Multiskalen-Vorkonditionierer induzierte Norm des Residuums minimieren, unabhängig von der Anzahl lt der auftretenden Skalen sind.

Rechtsvorkonditionierung

Wir betrachten das rechts additiv Multiskalen-vorkonditionierte Gleichungssystem auf der feinsten Skala

$$T_0 M_{0,s}^{-1} v_0 = f_0$$

und wählen schrittunabhängige Orthogonalisierungsmatrizen $Z_i = Z$, so daß in jedem Iterationsschritt i

$$||r_0^i||_{P_{L,i}^t Z P_{R,i}^{-1} T_0^{-1}} = ||r_0^i||_{Z M_{0,s} T_0^{-1}} \stackrel{!}{=} ||r_0^i||_{M_{0,s}^{-1}},$$

gilt, also zum Beispiel $Z = M_{0,s}^{-1} T_0 M_{0,s}^{-1}$. Nach Satz 21 erhalten wir die Abschätzungen

$$||r_0^i||_{M_{0,s}^{-1}} \leq (1 - \tau^P \cdot \bar{\tau}^P)^{\frac{i}{2}} ||r_0^0||_{M_{0,s}^{-1}}, \tag{1.83}$$

$$||e_0^i||_{T_0^t M_{0,s}^{-1} T_0} \leq (1 - \tau^P \cdot \bar{\tau}^P)^{\frac{i}{2}} ||e_0^0||_{T_0^t M_{0,s}^{-1} T_0} \tag{1.84}$$

mit den Konstanten

$$\tau^P = \inf_{\substack{w_0 \in \mathbb{R}^{N_0} \\ w_0 \neq 0}} \frac{(w_0, T_0 M_{0,s}^{-1} w_0)_{M_{0,s}^{-1}}}{(w_0, w_0)_{M_{0,s}^{-1}}} \quad \text{und} \quad \bar{\tau}^P = \inf_{\substack{w_0 \in \mathbb{R}^{N_0} \\ w_0 \neq 0}} \frac{(w_0, M_{0,s} T_0^{-1} w_0)_{M_{0,s}^{-1}}}{(w_0, w_0)_{M_{0,s}^{-1}}}.$$

Zum Nachweis des Optimalität des zugehörigen Verfahrens, genügt es zu zeigen, daß

$$\tau^P > 0 \quad \text{und} \quad \bar{\tau}^P > 0$$

unabhängig von von der Feingitter-Maschenweite h_0 und der Anzahl lt der Skalen gelten.

1. Wir betrachten τ^P und erhalten für die Skalarprodukte im Zähler und Nenner

$$\begin{aligned}
(w_0, T_0 M_{0,s}^{-1} w_0)_{M_{0,s}^{-1}} &= (\underbrace{M_{0,s}^{-1} w_0}_{=:z_0}, T_0 M_{0,s}^{-1} w_0) = (T_0 z_0, z_0) \\
&= a(z_0, z_0), \\
(w_0, w_0)_{M_{0,s}^{-1}} &= (M_{0,s}^{-1} w_0, w_0) = (M_{0,s} z_0, z_0) \\
&= h_{lt}^{-2} ||P_{0,\mathcal{V}}^{lt} z_0||^2 + \sum_{k=1}^{lt} h_k^{-2} ||Q_{0,\mathcal{V}}^k z_0||^2.
\end{aligned}$$

   Wegen der Normäquivalenz (1.80) folgt $\tau^P > c > 0$ mit $c \neq c(h_0, lt, \gamma)$.

2. Wir betrachten $\bar{\tau}^P$ und erhalten für die Skalarprodukte im Zähler und Nenner aufgrund Satz 23

$$\begin{aligned}(w_0, M_{0,s}T_0^{-1}w_0)_{M_{0,s}^{-1}} &= (w_0, \underbrace{T_0^{-1}w_0}_{=:z_0}) = (T_0 z_0, z_0) = a(z_0, z_0)\\ &\geq \frac{c}{1+\gamma^2}(M_{0,s}^{-1}w_0, w_0),\\ (w_0, w_0)_{M_{0,s}^{-1}} &= (M_{0,s}^{-1}w_0, w_0).\end{aligned}$$

Damit gilt $\bar{\tau}^P > \frac{c}{1+\gamma^2} > 0$ mit $c \neq c(h_0, lt)$.

Linksvorkonditionierung

Wir betrachten nun das links additiv Multiskalen-vorkonditionierte Gleichungssystem

$$M_{0,s}^{-1}T_0 u_0 = M_{0,s}^{-1}f_0$$

und wählen konstante Orthogonalisierungsmatrizen $Z_i = Z$, so daß

$$||r_0^i||_{P_{L,i}^t Z P_{R,i}^{-1} T_0^{-1}} = ||r_0^i||_{M_{0,s}^{-1} Z T_0^{-1}} \stackrel{!}{=} ||r_0^i||_{M_{0,s}^{-1}}$$

gilt, etwa $Z = T_0$, da $M_{0,s}^{-1}$ symmetrisch ist. Nach Satz 21 erhalten wir für das entsprechende Verfahren die gleichen Abschätzungen (1.83) und (1.84) wie im Fall der Vorkonditionierung von rechts, wobei die darin auftretenden Konstanten jetzt aber wie folgt definiert sind:

$$\tau^P := \inf_{\substack{w_0 \in \mathbb{R}^{N_0} \\ w_0 \neq 0}} \frac{(w_0, M_{0,s}^{-1}T_0 w_0)_{M_{0,s}}}{(w_0, w_0)_{M_{0,s}}} \quad \text{und} \quad \bar{\tau}^P := \inf_{\substack{w_0 \in \mathbb{R}^{N_0} \\ w_0 \neq 0}} \frac{(w_0, T_0^{-1}M_{0,s}w_0)_{M_{0,s}}}{(w_0, w_0)_{M_{0,s}}}.$$

Wir zeigen, daß diese wiederum unabhängig von h_0 und lt von Null weg beschränkt sind.

1. Wir betrachten τ^P und erhalten für die Skalarprodukte im Zähler und Nenner

$$\begin{aligned}(w_0, M_{0,s}^{-1}T_0 w_0)_{M_{0,s}} = (T_0 w_0, w_0) &= a(w_0, w_0),\\ (w_0, w_0)_{M_{0,s}} &= (M_{0,s}w_0, w_0)\\ &= h_{lt}^{-2}||P_{0,\mathcal{V}}^{lt}w_0||^2 + \sum_{k=1}^{lt} h_k^{-2}||Q_{0,\mathcal{V}}^k w_0||^2.\end{aligned}$$

Aufgrund der Normäquivalenz (1.80) folgt $\tau^P > c > 0$ mit $c \neq c(h_0, lt, \gamma)$.

2. Wir betrachten $\bar{\tau}^P$ und erhalten für die Skalarprodukte im Zähler und Nenner wegen Satz 23

$$\begin{aligned}(w_0, T_0^{-1}M_{0,s}w_0)_{M_{0,s}} &= (\underbrace{M_{0,s}w_0}_{=:z_0}, T_0^{-1}M_{0,s}w_0) = (z_0, T_0^{-1}z_0)\\ &= a(e_0, e_0) \quad (z_0 =: T_0 e_0)\\ &\geq \frac{c}{1+\gamma^2}(M_{0,s}^{-1}z_0, z_0),\\ (w_0, w_0)_{M_{0,s}} &= (M_{0,s}w_0, w_0)\\ &= (M_{0,s}^{-1}z_0, z_0).\end{aligned}$$

Es gilt also $\bar{\tau}^P > \frac{c}{1+\gamma^2} > 0$ mit $c \neq c(h_0, lt)$.

Wir fassen die obigen Ergebnisse zusammen:

Es seien die Voraussetzungen (1.73) und (1.74) in Bezug auf die schwache Formulierung des Konvektions-Diffusions Problems (1.69) erfüllt. Ausgehend von einem Galerkin-artigen Multiskalen-Ansatz gelte ebenfalls (1.80) und die Konstanten in den Normäquivalenzen seien unabhängig von der Konvektion $\vec{b}$.

Für additiv Multiskalen-vorkonditionierte Verallgemeinerte Krylovraum-Verfahren, welche die durch den Vorkonditionierer induzierte Norm des Residuums in jedem Iterationsschritt minimieren, gelten dann die Abschätzungen (1.83) und (1.84) mit Konstanten, die nicht von der Maschenweite der Diskretisierung und der Tiefe der Multiskalen-Zerlegung abhängen.

In den Fällen von Links- und Rechtsvorkonditionierung sind die darin auftretenden Größen τ^P konvektionsunabhängig. Sie stellen die Infima — genommen über alle reellen Vektoren ungleich $\mathbf{0}$ — der verallgemeinerten Rayleigh-Quotienten der vorkonditionierten Operatoren bezüglich der Skalarprodukte $(\cdot\,,\cdot)_{M_{0,s}^{-1}}$ (Rechtsvorkonditionierung) und $(\cdot\,,\cdot)_{M_{0,s}}$ (Linksvorkonditionierung) dar. In beiden Fällen sind die Größen $\bar{\tau}^P$ jedoch konvektionsabhängig. Sie sind die entsprechenden Infima der verallgemeinerten Rayleigh-Quotienten in Bezug auf die Inversen der vorkonditionierten Operatoren.
Die Größen τ^P und $\bar{\tau}^P$ können ebenfalls als die Infima der Realteile der numerischen Wertebereiche der vorkonditionierten Operatoren und ihrer Inversen hinsichtlich der entsprechenden Skalarprodukte angesehen werden.

Literaturverzeichnis

[1] R. E. Alcouffe, A. Brandt, J. E. Dendy und J. W. Painter: *The Multi-Grid Method for the Diffusion Equation with Strongly Discontinuous Coefficients*. SIAM J. Sci. Comput., 2:430–454, 1981.

[2] D. N. Allen und R. V. Southwell: *Relaxation Methods Applied to Determine Motion, in Two Dimensions, of a Viscous Fluid Past a Fixed Cylinder*. J. Mech. Appl. Math., 8:129–145, 1955.

[3] O. Axelsson: *An Algebraic Framework for Hierarchical Basis Functions Multilevel Methods or the Search for 'Optimal' Preconditioners*. In: D. R. Kincaid und L. J. Hayes (Hrsg.): *Iterative Methods for Large Linear Systems*. Academic Press, Boston, 1990.

[4] O. Axelsson: *Iterative Solution Methods*. Cambridge University Press, Cambridge, 1994.

[5] O. Axelsson und P. S. Vassilevski: *Algebraic Multilevel Preconditioning Methods. I*. Numer. Math., 56:157–177, 1989.

[6] O. Axelsson und P. S. Vassilevski: *Algebraic Multilevel Preconditioning Methods. II*. SIAM J. Numer. Anal., 27:1569–1590, 1990.

[7] O. Axelsson und P. S. Vassilevski: *The AMLI Method: An Algebraic Multilevel Iteration Method for Positive Definite Sparse Matrices*. Techn. Ber. 9936, Department of Mathematics, University of Nijmegen, 1999.

[8] I. Babuška und A. K. Aziz: *Survey Lectures on the Mathematical Foundation of the Finite Element Method*. In: A. K. Aziz (Hrsg.): *The Mathematical Foundation of the Finite Element Method with Applications to Partial Differential Equations*. Academic Press, New York, 1972.

[9] S. Balay, K. Buschelman, W. D. Gropp, D. Kaushik, L. C. McInnes und B. F. Smith: *PETSc home page*. http://www.mcs.anl.gov/petsc, 2001.

[10] S. Balay, W. D. Gropp, L. C. McInnes und B. F. Smith: *Efficient Management of Parallelism in Object Oriented Numerical Software Libraries*. In: E. Arge, A. M. Bruaset und H. P. Langtangen (Hrsg.): *Modern Software Tools in Scientific Computing*, S. 163–202. Birkhäuser, 1997.

[11] S. Balay, W. D. Gropp, L. C. McInnes und B. F. Smith: *PETSc Users Manual*. Techn. Ber. ANL-95/11 - Revision 2.1.0, Argonne National Laboratory, 2001.

[12] R. E. Bank: *Hierarchical Bases and the Finite Element Method*. Acta Numerica, 5:1–43, 1996.

[13] R. E. Bank und M. Benbourenane: *The Hierarchical Basis Multigrid Method for Convection-Diffusion Equations*. Numer. Math., 61:7–37, 1992.

[14] R. E. Bank und S. Gutsch: *Hierarchical Basis for the Convection-Diffusion Equation on Unstructured Meshes*. In: P. Bjørstad, M. Espedal und D. Keyes (Hrsg.): *Ninth International Conference on Domain Decomposition Methods*, 1998.

[15] R. E. Bank und J. Xu: *The Hierarchical Basis Multigrid Method and Incomplete LU Decomposition*. In: D. Keyes und J. Xu (Hrsg.): *Seventh International Conference on Domain Decomposition Methods*, 1994.

[16] R. E. Bank, T. F. Dupont und H. Yserentant: *The Hierarchical Basis Multigrid Method*. Numer. Math., 52:427–458, 1988.

[17] J. Bey: *Finite-Volumen- und Mehrgitter-Verfahren für elliptische Randwertprobleme*. B. G. Teubner, Stuttgart, 1998.

[18] G. Beylkin: *On the Representation of Operators in Bases of Compactly Supported Wavelets*. SIAM J. Numer. Anal., 29:1716–1740, 1992.

[19] G. Beylkin, R. Coifman und V. Rokhlin: *Fast Wavelet Transforms and Numerical Algorithms. I*. Comm. Pure and Appl. Math., 44:141–183, 1991.

[20] F. A. Bornemann und H. Yserentant: *A Basic Norm Equivalence in the Theory of Multilevel Methods*. Numer. Math., 64:455–476, 1993.

[21] E. F. F. Botta, A. van der Ploeg und F. W. Wubs: *Nested Grids ILU-Decomposition (NGILU)*. Journal of Comp. and Appl. Math., 66:515–526, 1996.

[22] D. Braess: *Finite Elemente*. Springer, Berlin, Heidelberg, New York, 1992.

[23] J. H. Bramble, J. E. Pasciak und J. Xu: *The Analysis of Multigrid Algorithms for Nonsymmetric and Indefinite Problems*. Math. Comp., 51:389–414, 1988.

[24] J. H. Bramble, J. E. Pasciak und J. Xu: *Parallel Multilevel Preconditioners*. Math. Comp., 55:1–22, 1990.

[25] A. Brandt: *Multi-Level Adaptive Technique (MLAT) for Fast Numerical Solution to Boundary Value Problems*. In: H. Cabannes und R. Temam (Hrsg.): *Proceedings of the 3rd International Conference on Numerical Methods in Fluid Mechanics*, Lecture Notes in Physics Vol. 18. Springer, Berlin, Heidelberg, New York, 1973.

[26] A. Brandt und I. Yavneh: *Inadequacy of Some First-Order Upwind Difference Schemes for Some Recirculating Flows*. J. Comp. Phys., 93:128–143, 1991.

[27] W. L. Briggs, V. E. Henson und S. F. McCormick: *A Multigrid Tutorial*. SIAM, Phiadelphia, 2000.

[28] J. M. Carnicer, W. Dahmen und J. M. Peña: *Local Decomposition of Refinable Spaces*. Appl. Comput. Harm. Anal., 3:127–153, 1996.

[29] T. F. Chan, W. P. Tang und W. L. Wan: *Wavelet Sparse Approximate Inverse Preconditioners*. BIT, 37:644–660, 1997.

[30] A. Cohen und R. Masson: *Wavelet Methods for Second-Order Elliptic Problems, Preconditioning and Adaptivity*. SIAM J. Sci. Comput., 21:1006–1026, 1999.

[31] N. A. Coult: *A Multiresolution Strategy for Homogenization of Partial Differential Equations*. Dissertation, University of Colorado at Boulder, 1997.

[32] W. Dahmen: *Wavelet and Multiscale Methods for Operator Equations*. Acta Numerica, 6:55–228, 1997.

[33] W. Dahmen und A. Kunoth: *Multilevel Preconditioning*. Numer. Math., 63:315–344, 1992.

[34] W. Dahmen und R. Schneider: *Wavelets on Manifolds I: Construction and Domain Decomposition*. Techn. Ber. 149, RWTH Aachen, 1999.

[35] W. Dahmen, S. Müller und T. Schlinkmann: *On a Robust Adaptive Multigrid Solver for Convection-Dominated Problems*. Techn. Ber. 171, RWTH Aachen, 1999.

[36] I. Daubechies: *Ten Lectures on Wavelets*. SIAM, Phiadelphia, 1992.

[37] J. E. Dendy: *Black Box Multigrid*. J. Comput. Phys., 48:366–386, 1982.

[38] J. E. Dendy: *Black Box Multigrid for Nonsymmetric Problems*. Appl. Math. Comput., 13:261–283, 1983.

[39] W. Dörfler: *Uniform Error Estimates for an Exponentially Fitted Finite Element Method for Singularly Perturbed Elliptic Equations*. SIAM J. Numer. Anal., 36:1709–1738, 1999.

[40] I. S. Duff, A. M. Erisman und J. K. Reid: *Direct Methods for Sparse Matrices*. Clarendon Press, Oxford, 1986.

[41] M. Eiermann und O. G. Ernst: *Geometric Aspects in the Theory of Krylov Subspace Methods*. Erscheint in Acta Numerica 2001.

[42] V. L. Eijkhout und P. S. Vassilevski: *The Role of the Strengthened Cauchy–Buniakowskii–Schwarz Inequality in Multilevel Methods*. SIAM Rev., 33:405–419, 1991.

[43] M. Embree: *How Descriptive are GMRES Convergence Bounds?*. Techn. Ber. OUCL Numerical Analysis Group 99/08, University of Oxford, 1999.

[44] P. A. Farell, A. F. Hegarty, J. J. M. Miller, E. O'Riordan und G. I. Shishkin: *Robust Computational Techniques for Boundary Layers*. Chapman & Hall, Boca Raton, 2000.

[45] R. P. Fedorenko: *A Relaxation Method for Solving Elliptic Difference Equations*. USSR Comp. Math. and Math. Phys., 5:1092–1096, 1961.

[46] M. Floater und E. Quak: *Piecewise Linear Prewavelets on Arbitrary Triangulations*. Numer. Math., 82:221–252, 1999.

[47] J. Fuhrmann: *Zur Verwendung von Mehrgitterverfahren bei der numerischen Behandlung elliptischer partieller Differentialgleichungen mit variablen Koeffizienten*. Dissertation, TU Chemnitz-Zwickau, 1994.

[48] D. L. Gines: *Fast Electromagnetic Simulations Using Wavelets*. Dissertation, University of Colorado at Boulder, 1997.

[49] D. L. Gines, G. Beylkin und J. Dunn: *LU Factorization of Non-Standard-Forms and Direct Multiresolution Solvers*. Techn. Ber. 278, Program in Applied Mathematics, University of Colorado at Boulder, 1996.

[50] H. Goering, A. Felgenhauer, G. Lube, H.-G. Roos und L. Tobiska: *Singularly Perturbed Differential Equations*. Akademie Verlag, Berlin, 1983.

[51] G. H. Golub und C. F. van Loan: *Matrix Computations*. The Johns Hopkins University Press, Baltimore, 1994.

[52] T. Grauschopf, M. Griebel und H. Regler: *Additive Multilevel-Preconditioners Based on Bilinear Interpolation, Matrix Dependent Geometric Coarsening and Algebraic Multigrid Coarsening for Second Order Elliptic PDEs*. Applied Numerical Mathematics, 23:63–96, 1997. Auch als Techn. Ber. 342/02/96A, Institut für Informatik, TU München, 1996.

[53] A. Greenbaum: *Iterative Methods for Solving Linear Systems*. SIAM, Phiadelphia, 1997.

[54] M. Griebel: *Zur Lösung von Finite-Differenzen- und Finite-Element-Gleichungen mittels der Hierarchischen Transformations-Mehrgitter-Methode*. Techn. Ber. 342/4/90 A, TU München, 1990.

[55] M. Griebel: *Multilevel Algorithms Considered as Iterative Methods on Semidefinite Systems*. SIAM J. Sci. Comput., 15:547–565, 1994.

[56] M. Griebel: *Multilevelmethoden als Iterationsverfahren über Erzeugendensystemen*. B. G. Teubner, Stuttgart, 1994.

[57] M. Griebel und P. Oswald: *On The Abstract Theory of Additive and Multiplicative Schwarz Algorithms*. Numer. Math., 70:163–180, 1995. Auch als Techn. Ber. 342/6/93 A, Institut für Informatik, TU München, 1993.

[58] M. Griebel und P. Oswald: *Tensor Product Type Subspace Splitting and Multilevel Iterative Methods for Anisotropic Problems*. Adv. Comput. Math., 4:171–206, 1995. Auch als Techn. Ber. 342/15/94A, Institut für Informatik, TU München, 1994.

[59] M. Griebel und G. Starke: *Multilevel Preconditioning Based on Discrete Symmetrization for Convection-Diffusion Equations*. Journal of Computational and Applied Mathematics, 83:165–183, 1997.

[60] C. Großmann und H.-G. Roos: *Numerik partieller Differentialgleichungen*. B. G. Teubner, Stuttgart, 1994.

[61] W. Hackbusch: *Ein iteratives Verfahren zur schnellen Auflösung elliptischer Randwertprobleme*. Techn. Ber. 76-12, Universität Köln, 1976.

[62] W. Hackbusch: *Multigrid-Methods and Applications*. Springer, Berlin, Heidelberg, New York, 1985.

[63] W. Hackbusch: *Integralgleichungen – Theorie und Numerik*. B. G. Teubner, Stuttgart, 1989.

[64] W. Hackbusch: *Iterative Lösung großer schwachbesetzter Gleichungssysteme*. B. G. Teubner, Stuttgart, 1993.

[65] W. Hackbusch: *Theorie und Numerik elliptischer Differentialgleichungen*. B. G. Teubner, Stuttgart, 1996.

[66] W. Hackbusch: *A Sparse Matrix Arithmetic Based on $\mathcal{H}$-Matrices. Part I: Introduction to $\mathcal{H}$-Matrices*. Computing, 62:89–108, 1999.

[67] W. Hackbusch und B. N. Khoromskij: *A Sparse $\mathcal{H}$-matrix Arithmetic. Part II: Application to Multi-Dimensional Problems*. Computing, 64:21–47, 2000.

[68] W. Hackbusch und T. Probst: *Downwind Gauß–Seidel Smoothing for Convection Dominated Problems*. Num. Lin. Algebra Appl., 4:85–102, 1997.

[69] W. Hackbusch und G. Wittum: *Incomplete Decomposition (ILU): Theory, Technique and Application, Notes on Numerical Fluid Mechanics 41*. Vieweg, Braunschweig, 1992.

[70] W. Hackbusch, B. N. Khoromskij und S. Sauter: *On H^2-Matrices*. In: H.-J. Bungartz, R. Hoppe und C. Zenger (Hrsg.): *Lectures on Applied Mathematics*, S. 9–30. Springer, Berlin, Heidelberg, New York, 2000.

[71] A. F. Hegarty, E. O'Riordan und M. Stynes: *A Comparison of Uniformly Convergent Difference Schemes for Two-Dimensional Convection-Diffusion Problems*. J. Comp. Phys., 105:24–32, 1993.

[72] A. R. Horn und C. R. Johnson: *Topics in Matrix Analysis*. Cambridge University Press, Cambridge, 1991.

[73] T. Huckle: *Matrix Multilevel Methods and Preconditioning*. Techn. Ber. 342/11/98A, Institut für Informatik, TU München, 1998.

[74] A. M. Il'in: *Differencing Scheme for a Differential Equation with a Small Parameter Affecting the Highest Derivative*. Mat. Notes Acad. Sc. USSR, 6:596–602, 1969.

[75] S. Jaffard: *Wavelet Methods for Fast Resolution of Elliptic Problems*. SIAM J. Numer. Anal., 29:965–986, 1992.

[76] K. Johannsen: *Robuste Mehrgitterverfahren für die Konvektions-Diffusions Gleichung mit wirbelbehafteter Konvektion*. B. G. Teubner, Stuttgart, 2000.

[77] U. Kotyczka und P. Oswald: *Piecewise Linear Prewavelets of Small Support*. In: C. Chui und L. Schumaker (Hrsg.): *Approximation Theory VIII*. World Scientific, Singapore, 1995.

[78] Y. A. Kuznetsov: *Algebraic Multigrid Domain Decomposition Methods*. Sov. J. Numer. Anal. Math. Modelling, 4:351–379, 1989.

[79] S. Le Borne: *Multigrid Methods for Convection-Dominated Problems*. Dissertation, Universität Kiel, 1999.

[80] J. F. Maître und F. Musy: *The Contraction Number of a Class of Two-Level Methods; an Exact Evaluation for Some Finite Element Subspaces and Model Problems*. In: W. Hackbusch und U. Trottenberg (Hrsg.): *Multigrid Methods*, Lecture Notes in Mathematics Vol. 960. Proceedings der First European Multigrid Conference, Springer, Berlin, Heidelberg, New York, 1982.

[81] S. Mallat: *Multiresolution Approximation and Wavelet Orthonormal Bases of* $\mathcal{L}^2(\mathbb{R})$. Trans. Amer. Math. Soc., 315:69–87, 1989.

[82] Y. Meyer: *Wavelets and Operators*. Cambridge University Press, Cambridge, 1992.

[83] Y. Meyer: *Wavelets and Fast Numerical Algorithms*. In: P. G. Ciarlet und J. L. Lions (Hrsg.): *Handbook of Numerical Analysis, Vol. V*. Elsevier Science, 1997.

[84] Y. Meyer und R. Coifman: *Wavelets, Calderón–Zygmund and Multilinear Operators*. Cambridge University Press, Cambridge, 1997.

[85] J. J. M. Miller, E. O'Riordan und G. I. Shishkin: *Fitted Numerical Methods for Singular Perturbation Problems*. World Scientific, Singapore, 1996.

[86] K. W. Morton: *Numerical Solution of Convection-Diffusion Problems*. Chapman & Hall, Boca Raton, 1996.

[87] N. M. Nachtigal, S. C. Reddy und L. N. Trefethen: *How Fast are Nonsymmetric Matrix Iterations?*. SIAM J. Matrix Anal. Appl., 13:778–795, 1992.

[88] M. Neytcheva: *Arithmetic and Communication Complexity of Preconditioning Methods*. Dissertation, University of Nijmegen, 1995.

[89] Y. Notay: *A Robust Algebraic Preconditioner for Finite Difference Approximations of Convection-Diffusion Equations*. Techn. Ber. GANMN 99–01, University of Brussels, 1999.

[90] P. Oswald: *On Discrete Norm Estimates Related to Multilevel Preconditioners in the Finite Element Method*. In: K. G. Ivanov, P. Petrushev und B. Sendov (Hrsg.): *Constructive Theory of Functions*. Proceedings der Int. Conf. Varna 1991, Bulg. Acad. Sci., Sofia, 1992.

[91] P. Oswald: *Multilevel Finite Element Approximation*. B. G. Teubner, Stuttgart, 1994.

[92] C. Pflaum: *Fast and Robust Multilevel Algorithms*. Habilitation, Universität Würzburg, 1998.

[93] C. Pflaum: *Robust Convergence of Multilevel Algorithms for Convection-Diffusion Equations*. SIAM J. Numer. Anal., 37:443–469, 2000.

[94] T. Probst: *Mehrgitterverfahren für Konvektionsdiffusionsgleichungen*. Dissertation, Universität Kiel, 1999.

[95] M. H. Protter und H. F. Weinberger: *Maximum Principles in Differential Equations*. Prentice-Hall, Englewood Cliffs, 1967.

[96] A. Reusken: *On the Approximate Cyclic Reduction Preconditioner*. Techn. Ber. 144, Institut für Geometrie und Praktische Mathematik, RWTH Aachen, 1997.

[97] A. Reusken: *Approximate Cyclic Reduction Preconditioning*. In: W. Hackbusch und G. Wittum (Hrsg.): *Multigrid Methods V*, Lecture Notes in Computational Science and Engineering Vol. 3. Proceedings der Fifth European Multigrid Conference, Springer, Berlin, Heidelberg, New York, 1998.

[98] A. Reusken: *An Algebraic Multilevel Preconditioner for Symmetric Positive Definite and Indefinite Problems*. In: A. Frommer, T. Lippert, B. Meldeke und K. Schilling (Hrsg.): *Numerical Challenges in Lattice Quantum Chromodynamics*, Lecture Notes in Computational Science and Engineering Vol. 15. Springer, Berlin, Heidelberg, New York, 2000.

[99] A. Reusken: *Convergence Analysis of a Multigrid Method for Convection-Diffusion Equations*. Techn. Ber. 190, RWTH Aachen, 2000.

[100] A. Rieder, R. O. Wells und X. Zhou: *A Wavelet Approach to Robust Multilevel Solvers for Anisotropic Elliptic Problems*. Appl. Comput. Harm. Anal., 1:355–367, 1994.

[101] H.-G. Roos: *Ten Ways to Generate the Il'in and Related Schemes*. Journal of Comp. and Appl. Math., 53:43–59, 1994.

[102] H.-G. Roos, M. Stynes und L. Tobiska: *Numerical Methods for Singularly Perturbed Equations*. Springer, Berlin, Heidelberg, New York, 1996.

[103] J. W. Ruge und K. Stüben: *Algebraic Multigrid*. In: S. F. McCormick (Hrsg.): *Multigrid Methods*. SIAM, Phiadelphia, 1987.

[104] Y. Saad: *Iterative Methods for Sparse Linear Systems*. PWS Publishing, Boston, 1996.

[105] Y. Saad und J. Zhang: *Diagonal Threshold Techniques in Robust Multi-Level ILU, Preconditioners for General Sparse Linear Systems*. Techn. Ber. 97-98, University of Minnesota, Minneapolis, 1998.

[106] Y. Saad und J. Zhang: *Enhanced Multi-Level Block ILU Preconditioning Strategies for General Sparse Linear Systems*. Techn. Ber. 98-98, University of Minnesota, Minneapolis, 1998.

[107] D. Schittko: *Waveletbasierte LU-Faktorisierung für elliptische Probleme*. Diplomarbeit, Universität Bonn, 2000.

[108] G. Starke: *Multilevel Minimal Residual Methods for Nonsymmetric Elliptic Problems*. Num. Lin. Algebra Appl., 3:351–367, 1996.

[109] G. Starke: *Field-Of-Values Analysis of Preconditioned Iterative Methods for Nonsymmetric Elliptic Problems*. Numer. Math., 78:103–117, 1997.

[110] R. Stevenson: *Piecewise Linear (Pre-)wavelets on Non-uniform Meshes*. In: W. Hackbusch und G. Wittum (Hrsg.): *Multigrid Methods V*, Lecture Notes in Computational Science and Engineering Vol. 3. Proceedings der Fifth European Multigrid Conference, Springer, Berlin, Heidelberg, New York, 1998.

[111] R. Stevenson: *Stable Three-Point Wavelet Bases on General Meshes*. Numer. Math., 80:131–158, 1998.

[112] K. Stüben: *Algebraic Multigrid (AMG): An Introduction with Applications*. Techn. Ber. 53, GMD-Forschungszentrum Informationstechnik GmbH, St. Augustin, 1999.

[113] W. Sweldens: *The Lifting Scheme: A Construction of Second Generation Wavelets*. SIAM J. Math. Anal., 29:511–546, 1997.

[114] P. Tchamitchian: *Wavelets, Functions and Operators*. In: G. Erlebacher, M. Y. Hussaini und L. M. Jameson (Hrsg.): *Wavelets, Theory and Applications*. Oxford University Press, 1996.

[115] L. N. Trefethen: *Pseudospectra of the Convection-Diffusion Operator*. SIAM J. Appl. Math., 54:1634–1649, 1994.

[116] L. N. Trefethen: *Pseudospectra of Linear Operators*. SIAM Rev., 39:383–406, 1997.

[117] U. Trottenberg, C. Oosterlee und A. Schüller: *Multigrid*. Academic Press, San Diego, 2001.

[118] P. S. Vassilevski: *Nearly Optimal Iterative Methods for Solving Finite Element Elliptic Equations Based on the Multilevel Splitting of the Matrix*. Techn. Ber. 1989-09, Institute for Scientific Computation, University of Wyoming, Laramie, 1989.

[119] P. S. Vassilevski: *Hybrid V-Cycle Algebraic Multilevel Preconditioners*. Math. Comp., 58:489–512, 1992.

[120] P. S. Vassilevski: *On Two Ways of Stabilizing the Hierarchical Basis Multilevel Methods*. SIAM Rev., 39:18–53, 1997.

[121] P. S. Vassilevski und J. Wang: *Stabilizing the Hierachical Basis by Approximate Wavelets. I: Theory*. Numer. Lin. Algebra Appl., 4:103–126, 1997.

[122] P. S. Vassilevski und J. Wang: *Stabilizing the Hierachical Basis by Approximate Wavelets, II: Implementation and Numerical Results*. SIAM J. Sci. Comput., 20:490–514, 1998.

[123] J. Wang: *Convergence Analysis of Multigrid Algorithms for Nonselfadjoint and Indefinite Elliptic Problems*. SIAM J. Numer. Anal., 30:275–285, 1993.

[124] R. Weiss: *Error-Minimizing Krylov Subspace Methods*. SIAM J. Sci. Comput., 15:511–527, 1994.

[125] R. Weiss: *Parameter-Free Iterative Linear Solvers*. Akademie Verlag, Berlin, 1996.

[126] G. Wittum: *On the Robustness of ILU Smoothing*. SIAM J. Sci. Comput., 10:699–717, 1989.

[127] J. Xu: *Iterative Methods by Space Decomposition and Subspace Correction*. SIAM Rev., 34:581–613, 1992.

[128] I. Yavneh, C. H. Venner und A. Brandt: *Fast Multigrid Solution of the Advection Problem with Closed Characteristics*. SIAM J. Sci. Comput., 19:128–143, 1998.

[129] H. Yserentant: *On the Multi-Level Splitting of Finite Element Spaces*. Numer. Math., 49:379–412, 1986.

[130] H. Yserentant: *Two Preconditioners Based on the Multi-Level Splitting of Finite Element Spaces*. Numer. Math., 58:163–184, 1990.

[131] H. Yserentant: *Old and New Convergence Proofs for Multigrid Methods*. Acta Numerica, 2:285–326, 1993.

[132] P. M. de Zeeuw: *Matrix-Dependent Prolongations and Restrictions in a Black-Box Multigrid Solver*. J. Comput. Appl. Math., 33:1–27, 1990.

[133] P. M. de Zeeuw: *Acceleration of Iterative Methods by Coarse Grid Corrections*. Dissertation, University of Amsterdam, 1997.

[134] J. Zhang: *A Grid Based Multilevel Incomplete LU Factorization Preconditioning Technique for General Sparse Matrices*. Techn. Ber. 283-99, Department of Computer Science, University of Kentucky, Lexington, 1999.

[135] J. Zhang: *On Preconditioning Schur Complement and Schur Complement Preconditioning*. Techn. Ber. 287-99, Department of Computer Science, University of Kentucky, Lexington, 1999.

[136] J. Zhang: *RILUM: A General Framework for Robust Multilevel Recursive Incomplete LU Preconditioning Techniques*. Techn. Ber. 284-99, Department of Computer Science, University of Kentucky, Lexington, 1999.

[137] X. Zhang: *Multilevel Schwarz Methods*. Numer. Math., 63:521–539, 1992.