



Algebraische Multilevel-Verfahren für ungeordnete physikalische Systeme

Zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

am Fachbereich Mathematik der
Bergischen Universität Gesamthochschule Wuppertal
genehmigte

Dissertation

von

Dipl.-Math. Björn Uwe Medeke

aus Wuppertal

Tag der mündlichen Prüfung: 17. Januar 2001
Referent: Prof. Dr. A. Frommer
Korreferent: Prof. Dr. G. Heindl

Für Anke

Vorwort

Das Lösen linearer Gleichungssysteme gehört sicherlich zu den häufigsten Berechnungen in der numerischen linearen Algebra. Dabei sind in den Anwendungen gerade die durch eine Diskretisierung partieller Differentialgleichungen entstehenden Koeffizientenmatrizen von besonderem Interesse. Diese Matrizen sind üblicherweise sehr groß, stark strukturiert, jedoch dünn besetzt, d.h. sie beinhalten nur wenige Nichtnulleinträge.

Kann keine spezielle Struktur der Koeffizientenmatrix ausgenutzt werden, so sind iterative Lösungsverfahren den direkten Lösungsverfahren vorzuziehen. Diese bestimmen ausgehend von einer Näherungslösung schrittweise bis zum Erreichen einer Genauigkeitsanforderung Approximationen an die gesuchte exakte Lösung des linearen Gleichungssystems.

Die einfachsten Iterationsverfahren zur Lösung großer linearer Gleichungssysteme sind die einstufigen, stationären Iterationsverfahren. Hierzu zählen neben der klassischen Jacobi-, Gauß-Seidel- und symmetrischen Gauß-Seidel-Iteration beispielsweise auch die durch eine unvollständige Dreieckszerlegung induzierte ILU-Iteration sowie jeweils die relaxierten Varianten. Da die Konvergenzrate dieser Verfahren allerdings abhängig ist von der zugrunde liegenden Diskretisierungsschrittweite, konvergieren diese iterativen Lösungsverfahren mit zunehmender Problemgröße langsamer.

◇

Im Jahre 1961 wurde von Fedorenko [22] das erste Mehrgitterverfahren eingeführt. Die Grundidee eines Mehrgitterverfahrens ist es, den oszillierenden Anteil des Fehlers durch eine Glättungsiteration zu eliminieren und den verbleibenden glatten Anteil sukzessive auf ein gröberes Gitter abzubilden, bis auch er sich wie ein oszillierender Fehleranteil verhält. Mehrgitterverfahren

berücksichtigen so in jedem Iterationsschritt nicht nur Kopplungen zu den unmittelbar benachbarten Gitterpunkten, sondern es wird Information zwischen entfernt liegenden Gitterpunkten ausgetauscht. Brandt [9] erkannte 1973 die Effizienz der Mehrgitterverfahren, da unter gewissen Voraussetzungen an die Qualität des Glätters und an die Approximationseigenschaft der verwendeten Grobgittersysteme der Aufwand proportional zur Anzahl der Unbekannten und die Konvergenzrate (im Gegensatz zur Konvergenzrate klassischer Iterationsverfahren) unabhängig von der Diskretisierungsschrittweite ist. 1976 wurden Mehrgitterverfahren von Hackbusch [35] wiederentdeckt und seitdem von verschiedenen Arbeitsgruppen systematisch untersucht.

Da Gebietszerlegungsmethoden wie die (multiplikative) Schwarz-Iteration und Schurkomplement-Methoden auch als (verallgemeinerte) Mehrgitterverfahren beschrieben und analysiert werden können, werden Mehrgitterverfahren und Gebietszerlegungsmethoden zur Klasse der Multilevel-Verfahren zusammengefaßt. Potentiell gehören Multilevel-Verfahren zu den schnellsten iterativen Lösungsverfahren für lineare Gleichungssysteme. Weil aber kein als Black-Box einsetzbares und zugleich effizientes Multilevel-Verfahren existiert, sondern statt dessen die Verfahrenskomponenten problemspezifisch konstruiert werden müssen, wurden Multilevel-Verfahren bisher fast ausschließlich zur Lösung geordneter Systeme mit meist symmetrisch positiv definiten Koeffizientenmatrizen untersucht.

◇

Ende der 80er und Anfang der 90er Jahre gab es erste Versuche seitens der Theoretischen Physik, Mehrgitterverfahren zur Lösung von Problemen aus der Quantenchromodynamik (QCD) einzusetzen. Ziel numerischer Simulationen in der QCD ist es, die starken Wechselwirkungen von Elementarteilchen zu untersuchen und Eigenschaften physikalischer Observablen vorherzusagen. Eine Diskretisierung der zugrunde liegenden Kontinuumstheorie führt auf die innerhalb eines Hybrid Monte Carlo Prozesses (HMC) wiederholt zu bestimmende Lösung eines großen, jedoch dünn besetzten linearen Gleichungssystems der Gestalt

$$A^H(t)A(t)\psi(t) = \varphi, \quad A(t) \in \mathbb{C}^{n \times n}, \quad \psi(t), \varphi \in \mathbb{C}^n.$$

Dabei sind die Matrizen $A(t)$ dieser Gleichungssysteme ungeordnet, d.h. die Koeffizienten sind statistisch erzeugt, und die Einträge der Matrix fluktuieren zufällig von einem Gitterpunkt zum nächsten.

Die bislang veröffentlichten Arbeiten konzentrieren sich auf sogenannte Multigrid-Monte-Carlo-Verfahren [40, 41, 42, 71] projektive Multigrid-

Ansätze [12, 14, 11, 13, 21], neuronale Multigrid-Methoden [5, 43] und *Parallel Transported Multigrid* [8, 7] für zumeist einfache Diskretisierungen.

◇

Ziel der vorliegenden Arbeit ist es, explorativ das Potential von Schurkomplement-Methoden zur Prädiktionierung ungeordneter Systeme zu untersuchen, die eng verwandt sind mit den Systemen aus der QCD.

Im ersten Kapitel werden zunächst häufig auftretende grundlegende Begriffe und Notation eingeführt. Nach einigen prinzipiellen Überlegungen zur Herleitung iterativer Lösungsverfahren wird auf klassische Iterationsverfahren näher eingegangen. Es folgt ein allgemeiner Überblick über existierende Krylov-Unterraum-Verfahren sowie elementare problemunabhängige Prädiktionierungstechniken.

Zu Beginn des zweiten Kapitels stellen wir anhand des Dirichletschen Randwertproblems exemplarisch die einzelnen Komponenten eines Zwei- respektive Mehrgitterverfahrens vor. Im zweiten Teil werden Schurkomplement-Methoden behandelt. Nach einer Einführung geometrischer und algebraischer Auswahlkriterien der Grobgitter gehen wir auf den Fall hermitesch positiv definiten Matrizen ein und formulieren für geordnete Matrizen dieser Klasse geeignete Restriktionen, Prolongationen sowie Grobgittermatrizen.

Die verbleibenden Kapitel beschäftigen sich mit der Konstruktion von Schurkomplement-Methoden zur Prädiktionierung von drei ähnlich strukturierten ungeordneten Systemen. Die Koeffizientenmatrizen beschreiben jeweils Nächste-Nachbar-Kopplungen auf regulären zwei- bzw. vierdimensionalen Gittern.

Ausgangspunkt in Kapitel 3 ist die hermitesch positiv definite Laplace Matrix mit zufällig erzeugten Kopplungen. Nach einer anfänglichen Darstellung der wesentlichen Matriceigenschaften und einigen Bemerkungen zu Mehrgitterverfahren für das geordnete System (konstante Kopplungen) werden sowohl für die originäre als auch für die zugehörige odd-even reduzierte Laplace Matrix Schurkomplement-Verfahren hergeleitet. Diese Schurkomplement-Verfahren basieren auf

- a) einer geometrischen F/C-Einteilung der Gitterpunkte, welche im Fall der odd-even reduzierten Matrix sogar rein algebraisch durch starke und schwache Kopplungen motiviert werden kann,
- b) einer durch einfache Zerlegungen induzierten approximativen Inverse der Teilmatrix, welche die Kopplungen zwischen feinen Gitterpunkten repräsentiert sowie

- c) einer Schurkomplement-Approximation, welche die Wirkung des exakten Schurkomplements auf Eigenvektoren zu kleinen Eigenwerten hinreichend gut approximiert und dennoch eine rekursive Vorgehensweise erlaubt.

Die numerischen Ergebnisse veranschaulichen für das Schurkomplement-präkonditionierte odd-even reduzierte System eine um einen Faktor 5-6 verbesserte asymptotische Konvergenzgeschwindigkeit des CG-Verfahrens gegenüber der für das unpräkonditionierte odd-even reduzierte System. Mit proportional zur Gitterdimension N wachsender Kondition verbessert sich sowohl für das ungeordnete als auch für das geordnete System die asymptotische Konvergenzgeschwindigkeit des CG-Verfahrens für das Schurkomplement-präkonditionierte odd-even reduzierte System im Vergleich zur asymptotischen Konvergenzgeschwindigkeit des CG-Verfahrens für das unpräkonditionierte System.

In Kapitel 4 wird für die odd-even reduzierte, positiv reelle Schwinger Matrix demonstriert, daß sich die zuvor für die Laplace Matrix entwickelten Schurkomplement-Präkonditionierer erfolgreich zur Präkonditionierung der Schwinger Matrix übertragen lassen. Dabei wird die in Kapitel 3 verwendete Einteilung der Kopplungen in starke und schwache modifiziert und auf Kopplungen repräsentierende Nichtnullblöcke angewendet. Es stellt sich insbesondere heraus, daß der verwendete Präkonditionierer wesentliche Eigenschaften der Schwinger Matrix respektiert und eine nicht-triviale γ_5 -Symmetrie erhält. Für die hermitesche Schwinger Matrix wird der Nachweis der Glättungseigenschaft für die modifizierte Jacobi-Glättung erbracht. Anschließend wird eine optimale semiiterative Glättungsiteration bestimmt. Vorgestellt wird des weiteren eine einfache Methode der (Eich-)fixierung, welche es offenbar vermag, Eigenvektoren zu betragskleinen Eigenwerten im geometrischen Sinne zu glätten.

In Kapitel 5 zeigen wir, daß beim Übergang von der Schwinger Matrix hin zur Wilson Fermion Matrix eine auf algebraisch motivierte Auswahlkriterien der Gitter beruhende Schurkomplement-Präkonditionierung des odd-even reduzierten Systems nicht mehr sinnvoll ist. Statt dessen werden Schurkomplement-Präkonditionierer der originären Wilson Fermion Matrix entworfen, welche zum einen die Symmetrieeigenschaften erhalten und zum anderen zu Grobgittersystemen führen, die wiederum als Wilson Fermion Matrizen interpretiert werden können [48].

Numerische Experimente lassen das Potential der Verfahren für Wilson Fermionen erahnen, jedoch ist der funktionale Zusammenhang der bei den Re-

striktionen, Prolongationen und Grobgittersystemen verwendeten Parameter nicht klar. Ein tieferes Verständnis dieses Ansatzes versprechen wir uns durch eine systematische Untersuchung der physikalischen Beziehung des ursprünglichen und der resultierenden Wilson Fermion Matrix seitens der Vertreter der Theoretischen Physik.

Völlig analog zum Fall der hermiteschen Schwinger Matrix kann eine optimale semiiterative Glättung für die hermitesche Wilson Fermion Matrix hergeleitet werden.

Abschließend wird der Zusammenhang zwischen Schwarz-Iterationen, welche auf hermitesch positiv definiten Teilräumen operieren und der Zerlegung der Wilson Fermion Matrix in ihren hermiteschen und schiefhermiteschen Anteil aufgezeigt. Obgleich nachgewiesen wird, daß sich Schwarz-Iterationen zur Präkonditionierung der Wilson Fermion Matrix nicht eignen, gaben diese Untersuchungen in der Diskussion mit Vertretern der Theoretischen Physik den Anstoß zur Entwicklung sogenannter *Single-Flavour*-Algorithmen [46].

◇

Die vorliegende Dissertation entstand im Zeitraum von Oktober 1996 bis September 2000 am Fachbereich Mathematik an der Bergischen Universität Gesamthochschule Wuppertal.

Mein besonderer Dank gilt meinem Doktorvater Prof. Dr. Andreas Frommer für die Aufnahme in seine Arbeitsgruppe 'Scientific Computing', für die interessante Themenstellung und für seine Diskussionsbereitschaft. Seine wertvollen Anmerkungen und Vorschläge sowie seine kritischen Kommentare begleiteten und erleichterten die Durchführung der Arbeit, und sie trugen zu ihrem Gelingen bei.

Für die Übernahme des Korreferats möchte ich Prof. Dr. Gerhard Heindl herzlich danken.

Dr. Dr. Thomas Lippert und Dr. Peer Ueberholz danke ich für die Beantwortung unzähliger Fragen zur Quantenchromodynamik, Dr. Rob Stevenson für seine Erklärungen zu algebraischen Wavelet-Präkonditionierern. Dr. Janos Korzak, Dr. Peter Fiebach und Dipl.-Math. Georg Weuffen gilt mein Dank für viele anregende Gespräche.

Ich danke allen Mitarbeiterinnen und Mitarbeitern der Arbeitsgruppe für die angenehme Arbeitsatmosphäre.

Wuppertal, im Januar 2001

Björn Medeke

Inhaltsverzeichnis

Vorwort	i
1 Grundlagen	1
1.1 Notation und Definitionen	1
1.2 Lineare Gleichungssysteme	3
1.3 Iterationsverfahren	5
1.3.1 Klassische Iterationsverfahren	5
1.3.2 Krylov-Unterraum-Verfahren	12
1.4 Präkonditionierung	15
2 Multilevel-Verfahren	19
2.1 Mehrgitterverfahren	19
2.1.1 Glättungsiteration	20
2.1.2 Grobgitterkorrektur	22
2.2 Schurkomplement-Methoden	26
2.3 Grobgitter	30
2.3.1 Geometrisch motivierte Grobgitter	31
2.3.2 Algebraisch motivierte Grobgitter	32
2.4 Restriktion und Prolongation	34
2.5 Grobgittersysteme	35
2.6 Konvergenzanalyse	36
2.7 Zusammenfassung	37
3 Das ungeordnete Laplace-System	39
3.1 Einführung	39
3.2 Eigenschaften	40
3.3 Das odd-even reduzierte System	45

3.4	Multilevel-Verfahren	47
3.4.1	Mehrgitterverfahren	48
3.4.2	Schurkomplement-Verfahren	54
3.4.3	Schurkomplement-Verfahren zur Lösung des odd-even reduzierten Systems	68
3.5	Numerische Ergebnisse	81
3.6	Zusammenfassung	87
4	Das Schwinger-Modell	89
4.1	Einführung	89
4.2	Eigenschaften	91
4.3	Schurkomplement-Verfahren	96
4.4	Numerische Ergebnisse	108
4.5	Die hermitesche Schwinger Matrix	108
4.5.1	Schurkomplement-Verfahren	109
4.5.2	Glättungseigenschaft	111
4.6	Zusammenfassung	117
5	Wilson-Fermionen	119
5.1	Einführung	119
5.2	Eigenschaften	121
5.3	Schurkomplement-Verfahren	123
5.4	Schwarz-Iteration	138
5.5	Zusammenfassung	142
A	Statistik	145
	Literaturverzeichnis	149

Abbildungsverzeichnis

1.1	Lexikographische und odd-even Anordnung der Gitterpunkte	11
2.1	Glatte und oszillierende Eigenvektoren	21
2.2	Das Standardgrobgitter L_2 und das Gitter $L_0 \cup L_2$	32
3.1	Der 5-Punkte-Stern der Laplace Matrix $A = I - \kappa D$	41
3.2	Das Besetztheitsmuster der Matrix D bei lexikographischer und odd-even Anordnung der Gitterpunkte	42
3.3	Das Spektrum der Matrix D für eine kalte und einer heiße Konfiguration	44
3.4	Das Spektrum der Laplace Matrix A und der zugehörigen odd-even reduzierten Matrix $A^{(e)}$ sowie das typische Konvergenzverhalten des CG-Verfahrens	46
3.5	Der 9-Punkte-Stern der odd-even reduzierten Matrix	47
3.6	Die Matrix $D_{L_i, L_1} D_{L_1, L_i}$ für $i = 0, 2$	50
3.7	Die Matrix $D_{L_i, L_1} D_{L_1, L_j} D_{L_j, L_1} D_{L_1, L_i}$ für $i, j = 0, 2$ mit $i \neq j$	52
3.8	Der 9-Punkte-Stern der Galerkin-Approximation	53
3.9	Die Kondition des präkonditionierten Schurkomplements für eine kalte Konfiguration	59
3.10	Die Funktion $p(t) = (1 + \omega t)(1 - t)$	64
3.11	Der 7-Punkte-Stern der Matrix $D_{L_1, L_0} D_{L_0, L_1}$ und der 9-Punkte-Stern der Schurkomplement-Approximation \tilde{S}_A	67
3.12	Die Kondition des präkonditionierten Schurkomplements für eine heiße Konfiguration	68
3.13	Die Zerlegung des Graphen $G_{A^{(e)}}(V, E)$ in die Graphen $G_{A^{(e)}}(V, E \setminus E_S^*)$ und $G_{A^{(e)}}(V, E_S^*)$	72

3.14	Ein Vergleich der Spektren der Jacobi- und der ILU-präkonditionierten Teilmatrix $A_{ff}^{(e)}$	77
3.15	Die relaxierte ILU-Präkonditionierung der Matrix $A_{ff}^{(e)}$	79
4.1	Das Spektrum der Matrix D für eine kalte und eine heiße Konfiguration	94
4.2	Die (Eich-)Fixierung einer heißen Konfiguration	95
4.3	Das Besetztheitsmuster der Teilmatrix $A_{fc}^{(e)}$	101
4.4	Die Spektren der Matrizen A und $A_{ff}^{(e)}$	102
4.5	Die Diagonale der Matrix $(D_{L_1, L_0} D_{L_0, L_1})^k$ für $1 \leq k \leq 4$	106
4.6	Die 9-Punkte-Sterne der ersten vier (Grobgitter-)Matrizen	107
4.7	Das typische Konvergenzverhalten des BiCGStab-Verfahrens	109
5.1	Das Spektrum der Wilson Fermion Matrix $A = I - \kappa D$ und der odd-even reduzierten Wilson Fermion Matrix $A^{(e)}$ für eine realistische Konfiguration	123
5.2	Das Besetztheitsmuster der odd-even reduzierten Wilson Fermion Matrix $A^{(e)}$	128
5.3	Das Besetztheitsmuster der Wilson Fermion Matrix A	130
5.4	Die Lösung Ψ_f des Systems $A_{ff} \Psi_f = \Phi_f$ bei einer Punktquelle Φ_f	132
5.5	Das Besetztheitsmuster der Wilson Fermion Systeme auf Ω_c	135
5.6	Die Extrapolation des kritischen Wertes κ_c für eine realistische Konfiguration auf Ω und für die korrespondierende Konfiguration auf Ω_c	136
5.7	Die Anzahl benötigter Iterationsschritte einer einfachen Zweigitteriteration für eine kalte Konfiguration in Abhängigkeit vom Grobgittergewicht	136
5.8	Das Konvergenzverhalten einer einfachen Zweigitteriteration und des GS-Verfahrens (bei einer odd-even Anordnung) für eine realistische Konfiguration	137
5.9	Die Ausführungszeiten des BiCGStab-Verfahrens und des Schurkomplement-präkonditionierten BiCGStab-Verfahrens	137
5.10	Das Spektrum des hermiteschen und des schief-hermiteschen Anteils von D	141

Tabellenverzeichnis

3.1	Der kritische Wert κ_c einer heißen Konfiguration	43
3.2	Numerische Ergebnisse für die Schurkomplement-präkonditionierte odd-even reduzierte Laplace Matrix auf Ω_h mit $N = 16$ und $m = 0.01$	83
3.3	Numerische Ergebnisse für die Schurkomplement-präkonditionierte odd-even reduzierte Laplace Matrix auf Ω_h mit $N = 32$ und $m = 0.01$	83
3.4	Numerische Ergebnisse für die Schurkomplement-präkonditionierte odd-even reduzierte Laplace Matrix auf Ω_h mit $N = 64$ und $m = 0.01$	84
3.5	Mittelwert und Standardabweichung für verschiedene Gittergrößen	84
3.6	Numerische Ergebnisse für die Schurkomplement-präkonditionierte odd-even reduzierte Laplace Matrix auf Ω_h mit $N = 16$ und $\kappa(A) = 255$	85
3.7	Numerische Ergebnisse für die Schurkomplement-präkonditionierte odd-even reduzierte Laplace Matrix auf Ω_h mit $N = 32$ und $\kappa(A) = 1023$	85
3.8	Numerische Ergebnisse für die Schurkomplement-präkonditionierte odd-even reduzierte Laplace Matrix auf Ω_h mit $N = 64$ und $\kappa(A) = 4095$	86
3.9	Numerische Ergebnisse für die Schurkomplement-präkonditionierte odd-even reduzierte Laplace Matrix auf Ω_h mit $N = 16$, $N = 32$, $N = 64$ für eine kalte Konfiguration und $\kappa(A) = N^2 - 1$	86

4.1	Numerische Ergebnisse für die Schurkomplement-präkonditionierte odd-even reduzierte Schwinger Matrix und verschiedene Gittergrößen	110
4.2	Ein Vergleich der Glättungsfaktoren des (modifizierten) Jacobi- und des semiiterativen Glätters für die hermitesche Schwinger Matrix Q	115
5.1	Der kritischer Wert κ_c für A sowie für den hermiteschen Anteil A_+ von A	141

Kapitel 1

Grundlagen

In diesem Kapitel werden zunächst einige häufig auftretende grundlegende Begriffe und Notation bereitgestellt. Anschließend werden die zur Lösung von linearen Gleichungssystemen existierenden, konzeptionell verschiedenen Ansätze erläutert. Nach einer Gegenüberstellung direkter Methoden und iterativer Verfahren klassifizieren wir die im Rahmen dieser Arbeit ausschließlich betrachteten iterativen Verfahren. Die resultierende Einteilung in klassische Iterationsverfahren und moderne Krylov-Unterraum-Verfahren erweist sich in vielen praktischen Anwendungen als unflexibel. Einen Ausweg bietet die Kombination von Krylov-Unterraum-Verfahren und klassischen Iterationsverfahren als Prädiktionierer.

1.1 Notation und Definitionen

Matrizen werden stets mit Groß-, Vektoren hingegen mit Kleinbuchstaben bezeichnet. Eine Ausnahme bildet der Vektor $\mathbf{1} = (1, \dots, 1)^T$. Alle Größen sind, sofern nicht explizit angegeben, komplex. Skalare Größen werden durch griechische Kleinbuchstaben gekennzeichnet.

Definition 1.1. Für $A \in \mathbb{C}^{m \times n}$ bezeichnet A^T die Transponierte und A^H die Adjungierte der Matrix A . Die Inverse einer regulären Matrix $A \in \mathbb{C}^{n \times n}$ wird durch A^{-1} gekennzeichnet.

Definition 1.2. Für die Einheitsmatrix schreiben wir I_n oder I , falls die Dimension dem Kontext zu entnehmen ist. Ist $A = (a_{ij}) \in \mathbb{C}^{m \times n}$ eine Matrix

mit $a_{ij} = 0$, so schreiben wir abkürzend $A = 0$.

Definition 1.3. Es sei $A = (a_{ij}) \in \mathbb{C}^{m \times n}$. Dann heißt $|A| = (|a_{ij}|)$ die Betragsmatrix von A .

Definition 1.4. Das Euklidische Innenprodukt zweier Vektoren $x, y \in \mathbb{C}^n$ ist definiert durch

$$\langle x, y \rangle = y^H x = \sum_{i=1}^n \bar{y}_i x_i,$$

und die Euklidische Norm $\|\cdot\|_2$ ist gegeben durch $\|x\|_2 = \sqrt{\langle x, x \rangle}$.

Definition 1.5. Eine hermitesche Matrix $A \in \mathbb{C}^{n \times n}$ ist positiv (semi)definit, falls $\langle x, Ax \rangle = x^H Ax > 0$ ($x^H Ax \geq 0$) ist für alle $x \neq 0$.

Definition 1.6. Eine hermitesche Matrix $A \in \mathbb{C}^{n \times n}$, deren quadratische Form $x^H Ax$ positive und negative Werte annimmt, heißt indefinit.

Durch eine hermitesch positiv definite Matrix A wird sowohl ein Innenprodukt als auch eine Norm induziert [19]:

Definition 1.7. Es sei $A \in \mathbb{C}^{n \times n}$ hermitesch positiv definit. Das Innenprodukt $\langle \cdot, \cdot \rangle_A$ ist definiert durch

$$\langle x, y \rangle_A = \langle Ax, y \rangle = y^H Ax, \quad x, y \in \mathbb{C}^n,$$

die Vektornorm $\|\cdot\|_A$ ist gegeben durch $\|x\|_A = \sqrt{\langle x, x \rangle_A}$.

Definition 1.8. Es sei $A \in \mathbb{C}^{n \times n}$. Das Spektrum der Matrix A ist definiert durch

$$\sigma(A) = \{\lambda \in \mathbb{C} : \det(A - \lambda I) = 0\}.$$

Jedes $\lambda \in \sigma(A)$ heißt Eigenwert von A , und ein Vektor $x \in \mathbb{C}^n$ heißt Eigenvektor von A zum Eigenwert λ , falls $x \neq 0$ und $Ax = \lambda x$ gilt.

Eine hermitesche Matrix $A \in \mathbb{C}^{n \times n}$ ist genau dann positiv definit, falls alle Eigenwerte von A positiv sind. Eine hermitesche Matrix $A \in \mathbb{C}^{n \times n}$ ist indefinit, falls sie positive und negative Eigenwerte besitzt.

Definition 1.9. Für eine Matrix $A \in \mathbb{C}^{n \times n}$, welche ausschließlich reelle Eigenwerte besitzt, bezeichnen $\lambda_{\min}(A)$ und $\lambda_{\max}(A)$ den kleinsten respektive den größten Eigenwert von A .

Definition 1.10. Der Spektralradius $\rho(A)$ einer Matrix $A \in \mathbb{C}^{n \times n}$ ist definiert durch

$$\rho(A) = \max_{\lambda \in \sigma(A)} |\lambda|.$$

Definition 1.11. Es sei $\|\cdot\|$ eine Norm auf \mathbb{C}^n . Dann ist die zugehörige Operatornorm in $\mathbb{C}^{n \times n}$ definiert als

$$\|A\| = \max_{\|x\|=1} \|Ax\| = \max_{x \neq 0} \frac{\|Ax\|}{\|x\|}, \quad A \in \mathbb{C}^{n \times n}.$$

Definition 1.12. Die (Spektral-)Kondition $\text{cond}(A)$ einer Matrix $A \in \mathbb{C}^{n \times n}$ ist definiert als

$$\text{cond}(A) = \|A\|_2 \|A^{-1}\|_2.$$

Für hermitesch positiv definite Matrizen A gilt [19]

$$\|A\|_2 = \lambda_{\max}(A), \quad \|A^{-1}\|_2 = \lambda_{\min}(A)$$

und daher auch $\text{cond}(A) = \kappa(A)$ mit

$$\kappa(A) = \frac{\lambda_{\max}(A)}{\lambda_{\min}(A)}.$$

1.2 Lineare Gleichungssysteme

Das Lösen linearer Gleichungssysteme ist in vielen praktischen Anwendungen der Mathematik eine häufig wesentliche Teilaufgabe. Die dabei zu betrachtenden Systeme

$$Ax = b, \quad A \in \mathbb{C}^{n \times n} \text{ regulär}, \quad x, b \in \mathbb{C}^n \quad (1.1)$$

sind typischerweise sehr groß, die Matrix A hingegen ist meist dünn besetzt, beinhaltet also nur wenige Nichtnullelemente. Matrizen dieser Art entstehen u.a. durch Diskretisierungen partieller Differentialgleichungen. Als Beispiel hierfür möge das Dirichletsche Randwertproblem dienen:

Beispiel 1.13. Es sei $\Omega = (0, 1) \times (0, 1)$ das offene Einheitsquadrat, $\delta\Omega$ dessen topologischer Rand und f eine auf $\bar{\Omega} = \Omega \cup \delta\Omega$ stetige Funktion. Gesucht ist eine auf $\bar{\Omega}$ stetige und auf Ω zweimal stetig differenzierbare Funktion u , welche das Dirichletsche Randwertproblem

$$\begin{aligned} -u_{xx} - u_{yy} &= f(x, y), & 0 < x, y < 1, \\ u(x, y) &= 0, & (x, y) \in \delta\Omega \end{aligned} \quad (1.2)$$

erfüllt. Eine Diskretisierung mittels der Methode der finiten Differenzen auf einem äquidistanten Gitter

$$\Omega_h = \left\{ (x_i, x_j) \mid (x_i, x_j) = (ih, jh), \quad 1 \leq i, j \leq N, \quad h = \frac{1}{N+1} \right\}$$

führt (bei lexikographischer Anordnung der Gitterpunkte, s. Def. 1.31) zu einem linearen Gleichungssystem

$$Au = f, \quad A \in \mathbb{R}^{n \times n}, \quad u, f \in \mathbb{R}^n \quad (1.3)$$

mit $n = N^2$ und $f_{(i-1)N+j} = h^2 f((x_i, x_j))$, wobei die Koeffizientenmatrix A Blocktridiagonalstruktur besitzt,

$$A = \begin{pmatrix} B & -I & 0 & 0 \\ -I & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & -I \\ 0 & 0 & -I & B \end{pmatrix} \in \mathbb{R}^{n \times n}, \quad B = \begin{pmatrix} 4 & -1 & 0 & 0 \\ -1 & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & -1 \\ 0 & 0 & -1 & 4 \end{pmatrix} \in \mathbb{R}^{N \times N}.$$

Die Matrix A ist offensichtlich dünn besetzt, denn es gibt höchstens 5 Nicht-nulleinträge pro Zeile. Die Dimension n der Matrix A ist abhängig von der Diskretisierungsschrittweite h ,

$$n = \left(\frac{1}{h} - 1 \right)^2.$$

Direkte Methoden

Auf den ersten Blick bereitet die Lösung von sehr großen, dünn besetzten, linearen Gleichungssystemen keinerlei Schwierigkeiten: Nach der Berechnung einer LU-Zerlegung $A = LU$ der Koeffizientenmatrix A (oder einer Permutation von A) mit Hilfe des Gaußschen Eliminationsverfahrens kann der Lösungsvektor auf einfache Art und Weise über die resultierenden Dreieckssysteme bestimmt werden,

$$Ly = b, \quad Ux = y. \quad (1.4)$$

In der Praxis ist diese direkte Vorgehensweise i. a. nicht geeignet. Zum einen ist der hohe Rechenaufwand des Gaußschen Eliminationsverfahrens für hinreichend große Matrixdimension prohibitiv, zum anderen sind die Faktoren L und U im Gegensatz zu A meist voll besetzt. Die bei der Faktorisierung entstehende Auffüllung kann durch intelligente Pivotstrategien auf ein Minimum beschränkt werden. Dazu sind jedoch geeignete Datenstrukturen erforderlich, und eine effiziente Realisierung des so modifizierten Eliminationsverfahrens ist oft schwierig. Daher betrachten wir ab jetzt ausschließlich iterative Verfahren.

1.3 Iterationsverfahren

Iterative Verfahren berechnen sukzessive bis zum Erreichen eines vorgegebenen Abbruchkriteriums wie beispielsweise einer Genauigkeitsanforderung Approximationen an die gesuchte exakte Lösung des linearen Gleichungssystems. Allen iterativen Verfahren ist gemein, daß in jedem Iterationsschritt im wesentlichen eine Matrix-Vektor-Multiplikation mit der dünn besetzten Koeffizientenmatrix auszuführen ist. Der Rechenaufwand ist daher in jedem einzelnen Iterationsschritt vergleichbar mit dem Aufwand für eine Matrix-Vektor-Multiplikation mit A . Eine Ausnahme bildet das GMRES-Verfahren [60], bei dem der Aufwand zur Lösung eines Kleinste-Quadrate-Problems dominiert.

1.3.1 Klassische Iterationsverfahren

Die ältesten und einfachsten Iterationsverfahren zur Lösung linearer Gleichungssysteme sind die einstufigen, stationären Iterationsverfahren. Eine ausführliche Einführung dieser und verwandter Iterationsverfahren findet man in verschiedenen klassischen Lehrbüchern [69, 74].

Iterationsverfahren

Definition 1.14. Ein einstufiges, stationäres Iterationsverfahren ist eine Iteration der Gestalt

$$x^{(k+1)} = Hx^{(k)} + c, \quad k = 0, 1, \dots \quad (1.5)$$

mit einer Iterationsmatrix $H \in \mathbb{C}^{n \times n}$ und einem Vektor $c \in \mathbb{C}^n$ zu einem gegebenen Startvektor $x^{(0)} \in \mathbb{C}^n$.

Iterative Verfahren der Gestalt (1.5) werden als stationär bezeichnet, da die Iterationsmatrix H nicht vom jeweiligen Iterationsschritt k abhängt. Der Begriff „einstufig“ gibt an, daß zur Berechnung der neuen Iterierten $x^{(k+1)}$ lediglich die unmittelbar zuvor berechnete Iterierte $x^{(k)}$ herangezogen wird.

Zerlegungen

Zur Konstruktion geeigneter einstufiger, stationärer Iterationsverfahren zur Lösung des linearen Gleichungssystems (1.1) benötigen wir zunächst die folgende Definition.

Definition 1.15. Es sei $A \in \mathbb{C}^{n \times n}$. Die Darstellung $A = M - N$ von A heißt Zerlegung von A , falls M regulär ist.

Ist $A = M - N$ eine Zerlegung der Matrix $A \in \mathbb{C}^{n \times n}$, so erfüllt die gesuchte exakte Lösung die Fixpunktgleichungen

$$Mx^* = Nx^* + b, \quad x^* = M^{-1}Nx^* + M^{-1}b. \quad (1.6)$$

Daher induziert jede Zerlegung $A = M - N$ einer regulären Matrix $A \in \mathbb{C}^{n \times n}$ ein Iterationsverfahren der Gestalt (1.5) zur Lösung des Systems (1.1),

$$x^{(k+1)} = M^{-1}Nx^{(k)} + M^{-1}b, \quad k = 0, 1, \dots$$

Im Falle der Konvergenz strebt die Folge der Iterierten $x^{(k)}$ nach den vorhergehenden Überlegungen gegen die gesuchte Lösung,

$$x^{(k)} \rightarrow x^* = A^{-1}b, \quad k \rightarrow \infty. \quad (1.7)$$

Beim praktischen Rechnen wird natürlich weder die Inverse M^{-1} der Matrix M noch die Iterationsmatrix $H = M^{-1}N$ explizit konstruiert, sondern die Iterierte $x^{(k+1)}$ wird über das durch (1.6) induzierte System gemäß der Vorschrift

$$Mx^{(k+1)} = Nx^{(k)} + b, \quad k = 0, 1, \dots$$

aus der Iterierten $x^{(k)}$ berechnet.

Bevor wir uns Konvergenzaussagen einfacher, stationärer Iterationsverfahren zuwenden, werden wir exemplarisch einige häufig verwendete, Iterationsverfahren induzierende Zerlegungen angegeben. Abkürzend bezeichnet dabei D den Diagonalanteil, $-L$ den strikten unteren und $-U$ den strikten oberen Dreiecksanteil der Matrix A .

Jacobi-Zerlegung

Definition 1.16. Ist $A = D - L - U \in \mathbb{C}^{n \times n}$, D regulär und $\omega \in \mathbb{R} \setminus \{0\}$, so heißt $A = P - Q$ mit

$$\begin{aligned} P &= \omega^{-1}D \\ Q &= \omega^{-1}(1 - \omega)D + L + U \end{aligned} \quad (1.8)$$

relaxierte Jacobi-Zerlegung von A und das induzierte Iterationsverfahren relaxiertes Jacobi-Verfahren.

Beispiel 1.17. Die Iterierte $x^{(k+1)}$ des relaxierten Jacobi-Verfahrens zur Lösung des linearen Gleichungssystems (1.3) ist komponentenweise gegeben durch

$$x_i^{(k+1)} = (1 - \omega)x_i^{(k)} + \frac{\omega}{4} \left(b_i + x_{i-1}^{(k)} + x_{i+1}^{(k)} + x_{i-N}^{(k)} + x_{i+N}^{(k)} \right). \quad (1.9)$$

Dabei sind Komponenten außerhalb des zulässigen Indexbereichs identisch Null zu setzen.

SOR- und SSOR-Zerlegung

Definition 1.18. Ist $A = D - L - U \in \mathbb{C}^{n \times n}$, D regulär und $\omega \in \mathbb{R} \setminus \{0\}$, so heißt $A = P - Q$ mit

$$\begin{aligned} P &= \omega^{-1}(D - \omega L) \\ Q &= \omega^{-1}((1 - \omega)D + \omega U) \end{aligned} \quad (1.10)$$

SOR-Zerlegung von A und das induzierte Iterationsverfahren SOR-Verfahren (engl.: Successive Over-Relaxation) oder auch relaxiertes Gauß-Seidel-Verfahren.

Definition 1.19. Es sei $A = D - L - U \in \mathbb{C}^{n \times n}$, D regulär und $\omega \in \mathbb{R} \setminus \{0, 2\}$. Dann heißt $A = P - Q$ mit

$$\begin{aligned} P &= (\omega(2 - \omega))^{-1}(D - \omega L)D^{-1}(D - \omega U) \\ Q &= (\omega(2 - \omega))^{-1}(((1 - \omega)D + \omega L)D^{-1}((1 - \omega)D + \omega U)) \end{aligned} \quad (1.11)$$

SSOR-Zerlegung von A und das induzierte Iterationsverfahren SSOR-Verfahren (engl.: Symmetric Successive Over-Relaxation) oder auch relaxiertes symmetrisches Gauß-Seidel-Verfahren.

Für $\omega = 1$ wird das SOR-Verfahren (SSOR-Verfahren) auch (symmetrisches) Gauß-Seidel-Verfahren genannt und abkürzend GS-Verfahren (SGS-Verfahren) geschrieben.

Eine weitere Klasse einfacher, stationärer Iterationsverfahren wird durch die sogenannte unvollständige LU-Zerlegung (engl.: Incomplete LU decomposition; ILU-Zerlegung) induziert:

ILU-Zerlegung

Die Idee der ILU-Zerlegung und ihrer Varianten beruht darauf, innerhalb des Gaußschen Eliminationsprozesses die Auffüllung (engl.: fill-in) der Faktoren L und U lediglich an bestimmten Positionen zuzulassen.

Definition 1.20. Es sei $A = (a_{ij}) \in \mathbb{C}^{n \times n}$ und $\mathcal{I} = \{1, \dots, n\}$ eine angeordnete Indexmenge. Die Menge $E \subseteq \mathcal{I} \times \mathcal{I}$ heißt Muster, falls gilt

$$\{(i, i) \mid i \in \mathcal{I}\} \subseteq E.$$

Definition 1.21. Eine unvollständige LU-Zerlegung von $A = (a_{ij}) \in \mathbb{C}^{n \times n}$ zum Muster E ist eine Zerlegung der Form $A = LU - R$. Dabei ist $L =$

(l_{ij}) eine normierte untere und $U = (u_{ij})$ eine obere Dreiecksmatrix mit der Eigenschaft

$$l_{ij} = u_{ij} = 0 \text{ für alle } (i, j) \notin E.$$

Gilt zusätzlich

$$\sum_{k=1}^n l_{ik} u_{kj} = a_{ij} \text{ für alle } (i, j) \in E,$$

so heißt die ILU-Zerlegung direkt.

Üblicherweise besteht das Muster E mindestens aus den Positionen der Nichtnullelemente der Matrix $A = (a_{ij}) \in \mathbb{C}^{n \times n}$ und dem minimalen Muster,

$$E = \{(i, j) \mid a_{i,j} \neq 0\} \cup \{(i, i) \mid i \in \mathcal{I}\}.$$

Relaxierte ILU-Zerlegung

Bei einer ILU-Zerlegung $A = LU - R$ werden Nichtnullelemente der Matrix A außerhalb des Musters E nicht berücksichtigt. Bei der folgenden Modifikation werden diese Elemente nicht ersatzlos ignoriert, sondern relaxiert auf die Diagonale der Matrix LU aufgetragen:

Definition 1.22. Es sei $\omega \in \mathbb{R}$. Eine relaxierte ILU-Zerlegung von $A = (a_{ij}) \in \mathbb{C}^{n \times n}$ zum Muster E ist eine ILU-Zerlegung von A zum Muster E , für die gilt

$$\begin{aligned} a_{ij} &= \sum_{k=1}^n l_{ik} u_{kj} \text{ für alle } (i, j) \in E, i \neq j, \\ r_{ii} &= \omega \sum_{j=1, j \neq i}^n r_{ij} \text{ für alle } i. \end{aligned} \tag{1.12}$$

Für $\omega = 0$ sind die relaxierte ILU-Zerlegung und die ILU-Zerlegung identisch, für $\omega = -1$ gilt $LU\mathbf{1} = A\mathbf{1}$.

Konvergenz

Definition 1.23. Ist $x^{(k)}$ Iterierte eines Iterationsverfahrens zur Lösung des Systems $Ax = b$, so bezeichnet $r^{(k)} = b - Ax^{(k)}$ das Residuum und $e^{(k)} = x^* - x^{(k)}$ mit $x^* = A^{-1}b$ den Fehler. Zwischen Fehler und Residuum gilt die Beziehung $Ae^{(k)} = r^{(k)}$.

Die Konvergenzeigenschaften einstufiger, stationärer Iterationsverfahren wird vollständig bestimmt durch die zugehörige Iterationsmatrix

$$H = M^{-1}N = M^{-1}(M - A) = I - M^{-1}A, \quad (1.13)$$

denn für den Fehler gilt offensichtlich

$$e^{(k)} = He^{(k-1)} = H^k e^{(0)}. \quad (1.14)$$

Wir werden diese Aussage nun präzisieren.

Definition 1.24. Das Iterationsverfahren (1.5) heißt konvergent, falls ein Vektor $x \in \mathbb{C}^n$ existiert, so daß unabhängig von der Wahl des Startvektors $x^{(0)} \in \mathbb{C}^n$ die Folge der Iterierten $x^{(k)}$ aus (1.5) gegen x konvergiert.

Satz 1.25. Das Iterationsverfahren (1.5) ist genau dann konvergent, wenn für den Spektralradius $\rho(H)$ der Iterationsmatrix H die Ungleichung $\rho(H) < 1$ erfüllt ist.

Beweis. [26], Satz A.2.4. □

Eine hinreichende Bedingung für die Konvergenz des Iterationsverfahrens (1.5) ist $\|H\| < 1$, da die Abschätzung $\rho(H) \leq \|H\|$ für jede Operatornorm in $\mathbb{C}^{n \times n}$ gilt.

Konvergenzrate

Wir benötigen nun die Definition der Konvergenzrate einer Nullfolge, um anschließend als Maß für die Konvergenzgeschwindigkeit von Iterationsverfahren die asymptotische Konvergenzrate definieren zu können. Es wird sich herausstellen, daß die asymptotische Konvergenzrate mit dem Spektralradius der Iterationsmatrix identisch ist.

Definition 1.26. Es sei $\{a_k\}$ eine Nullfolge in \mathbb{R}_0^+ . Dann bezeichnet $R(a_k)$ die Konvergenzrate

$$R(a_k) = \inf \{r > 0 \mid \exists C > 0 : a_k \leq Cr^k \forall k\}.$$

Definition 1.27. Das Iterationsverfahren (1.5) sei konvergent. Dann bezeichnet $R = R((1.5))$ die Konvergenzrate von (1.5),

$$R = R((1.5)) = \sup \{R(\|e^{(k)}\|) \mid \{x^{(k)}\} \text{ ist Iteriertenfolge von (1.5)}\}.$$

Satz 1.28. Das Iterationsverfahren (1.5) sei konvergent. Dann gilt

$$R((1.5)) = \rho(H).$$

Beweis. [26], Satz A.2.6. □

Definition 1.29. Die asymptotische Konvergenzgeschwindigkeit des Iterationsverfahrens (1.5) ist definiert als

$$A = A((1.5)) = -\log_{10} R((1.5)).$$

Die asymptotische Konvergenzgeschwindigkeit läßt sich dahingehend interpretieren, daß für hinreichend großes k die Anzahl der korrekten Dezimalstellen sich pro Iteration um A erhöht.

Beispiel 1.30. Die Iterationsmatrix H des relaxierten Jacobi-Verfahrens zur Lösung des linearen Gleichungssystems (1.3) besitzt die n Eigenwerte

$$\lambda^{(k,l)} = (1 - \omega) + \frac{\omega}{2} \left(\cos \left(\frac{k\pi}{N+1} \right) + \cos \left(\frac{l\pi}{N+1} \right) \right), \quad 1 \leq k, l \leq N$$

mit den zugehörigen Eigenvektoren $z^{(k,l)}$,

$$z_{i,j}^{(k,l)} = \sin \left(\frac{k\pi i}{N+1} \right) \sin \left(\frac{l\pi j}{N+1} \right), \quad 1 \leq i, j \leq N.$$

Der Spektralradius $\rho(H)$ der Iterationsmatrix H ist daher

$$\rho(H) = 1 - \omega \left(1 - \cos \left(\frac{\pi}{N+1} \right) \right) = 1 - \mathcal{O}(N^{-2}). \quad (1.15)$$

Das Beispiel zeigt, daß für das relaxierte Jacobi-Verfahren die Konvergenzrate von der Diskretisierungsschrittweite h abhängt, d.h. das relaxierte Jacobi-Verfahren konvergiert um so langsamer, je feiner die Diskretisierung gewählt wird. Dieses Phänomen wird in der Theoretischen Physik *critical slowing down* genannt.

Anordnungen der Gitterpunkte

Abgesehen von der Diskussion in Kapitel 5 werden im Rahmen der vorliegenden Arbeit ausschließlich Koeffizientenmatrizen A betrachtet, welche Kopplungen zwischen Gitterpunkten auf dem sogenannten zweidimensionalen Standard-Gitter

$$\Omega_h = \left\{ (x_i, x_j) \mid (x_i, x_j) = (ih, jh), \quad 1 \leq i, j \leq N, \quad h = \frac{1}{N+1} \right\} \quad (1.16)$$

repräsentieren. Unterschiedliche Anordnungen der Gitterpunkte führen zu unterschiedlichen Koeffizientenmatrizen. Eine Matrix erhalten wir aus einer

anderen durch eine Transformation der Gestalt $A \rightarrow P^T A P$, wobei P eine Permutationsmatrix ist. Da sich mit den Koeffizientenmatrizen im allgemeinen auch die Iterationsmatrizen ändern, können sich die Konvergenzraten eines einfachen, stationären Iterationsverfahrens bei einem Wechsel von einer Anordnung der Gitterpunkte zu einer anderen unterscheiden. Daher werden nun zwei wichtige Anordnungen eingeführt.

Eine natürliche Reihenfolge ist die sogenannte lexikographische Anordnung:

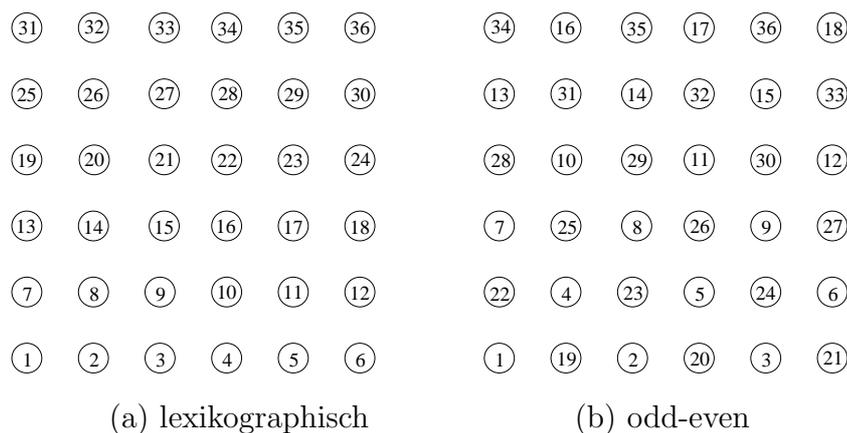
Definition 1.31. Für zwei Gitterpunkte $x = (ih, jh)$ und $y = (kh, lh)$ gelte $x < y$ genau dann, wenn entweder $i < k$ ist oder $i = k$ und $j < l$. Die Anordnung einer Menge von Gitterpunkten bezüglich „ $<$ “ heißt lexikographische Anordnung.

Im Rahmen dieser Arbeit benötigen wir desweiteren die odd-even Anordnung. Dazu definieren wir:

Definition 1.32. Der Gitterpunkt $x = (ih, jh)$ heißt gerade (engl.: even), falls die Summe $i + j$ gerade ist, andernfalls ungerade (engl.: odd).

Definition 1.33. Permutieren wir die Gitterpunkte so, daß zunächst alle ungeraden und anschließend alle geraden Gitterpunkte gezählt werden, so nennen wir die resultierende Anordnung auch odd-even Anordnung.

Abbildung 1.1 veranschaulicht eine lexikographische Anordnung und eine odd-even Anordnung der Gitterpunkte.



(a) lexikographisch

(b) odd-even

Abbildung 1.1: Lexikographische und odd-even Anordnung der Gitterpunkte.

Die Definition sowohl der lexikographischen als auch der odd-even Anordnung der Gitterpunkte läßt sich ohne Mühe auf d -dimensionale Gitter mit $d > 2$ erweitern, vgl. Definition 5.4 und Definition 5.5.

1.3.2 Krylov-Unterraum-Verfahren

Krylov-Unterraum-Verfahren sind nicht-stationäre, iterative Lösungsverfahren. Die Grundidee aller Krylov-Unterraum-Verfahren ist stets die gleiche. Es werden sukzessive Vektorräume höherer Dimension erzeugt, gemäß eines vorgegebenen Auswahlkriteriums darin ein Vektor als Approximation an die gesuchte Lösung des Gleichungssystems bestimmt und überprüft, ob diese Näherung bereits den Genauigkeitsanforderungen genügt. Die dabei verwendeten Vektorräume sind die sogenannten Krylov-Unterräume.

Definition 1.34. Es sei $A \in \mathbb{C}^{n \times n}$, $r \in \mathbb{C}^n$ und $m \in \mathbb{N}$. Dann heißt

$$K_m(A, r) = \text{span} \{r, Ar, \dots, A^{m-1}r\}$$

der Krylov-Unterraum der Stufe m bezüglich A und r .

Offensichtlich ist die Dimension der Krylov-Unterräume nach oben beschränkt, insbesondere spannt nicht jeder Krylov-Unterraum den \mathbb{C}^n vollständig auf.

Lemma 1.35. Es sei $A \in \mathbb{C}^{n \times n}$ und $r \in \mathbb{C}^n \setminus \{0\}$. Dann existiert ein $m_0 \in \mathbb{N}$, $m_0 \leq n$ mit $\dim K_m(A, r) = m$ für alle $m = 1, \dots, m_0$ und $\dim K_m(A, r) = m_0$ für alle $m \geq m_0$.

Beweis. [15], Abschnitt 3.1. □

Satz 1.36. Es sei $A \in \mathbb{C}^{n \times n}$ regulär, $b \in \mathbb{C}^n$. Dann gilt für jedes $x^{(0)} \in \mathbb{C}^n$

$$\begin{aligned} x^* &= A^{-1}b \in x^{(0)} + K_{m_0}(A, r^{(0)}), \\ A^{-1}b &\notin x^{(0)} + K_m(A, r^{(0)}), \quad m < m_0 \end{aligned}$$

mit $r^{(0)} = b - Ax^{(0)}$ und m_0 aus Lemma 1.35.

Beweis. [15], Abschnitt 3.1. □

Mit der Definition der Krylov-Unterräume ist es nun möglich, die Klasse der Krylov-Unterraum-Verfahren zu charakterisieren:

Definition 1.37. Ein Krylov-Unterraum-Verfahren zur Lösung von (1.1) ist ein Verfahren, für dessen Iterierte $x^{(m)}$ gilt, daß $x^{(m)} \in x^{(0)} + K_m(A, r^{(0)})$ ist.

Das CG-Verfahren

Krylov-Unterraum-Verfahren unterscheiden sich durch die Auswahl der Iterierten. So kann beispielsweise unter geeigneten Voraussetzungen an die Matrix erreicht werden, daß die Iterierte gewissen Optimalitätsbedingungen genügt. Exemplarisch sei hier auf das Anfang der 50er Jahre entwickelte CG-Verfahren [39] hingewiesen, welches den Erfolg von Krylov-Unterraum-Verfahren als moderne iterative Löser einleitete.

Für hermitesch positiv definite Matrizen $A \in \mathbb{C}^{n \times n}$ wird die CG-Iterierte in jedem Schritt so bestimmt, daß der Fehler in der A -Norm unter allen möglichen Iterierten aus dem Krylov-Unterraum minimal ist,

$$\|e^{(k)}\|_A = \|x^* - x^{(k)}\|_A = \min_{x \in x^{(0)} + K_k(A, r^{(0)})} \|x^* - x\|_A. \quad (1.17)$$

Diese Minimalitätseigenschaft kann auch dahingehend interpretiert werden, daß das Residuum orthogonal zum bereits erzeugten Krylov-Unterraum ist (s. [15], Abschnitt 3.3.1),

$$r^{(k)} \perp K_k(A, r^{(0)}). \quad (1.18)$$

Ausgehend von (1.17) kann der Fehler in jedem Iterationsschritt nach oben abgeschätzt werden:

Satz 1.38. Es sei $A \in \mathbb{C}^{n \times n}$ hermitesch positiv definit. Der Fehler der CG-Iterierten genügt der Ungleichung

$$\|e^{(k)}\|_A \leq 2 \left(\frac{\sqrt{\kappa(A)} - 1}{\sqrt{\kappa(A)} + 1} \right)^k \|e^{(0)}\|_A. \quad (1.19)$$

Beweis. [66], Theorem 38.5. □

Das CG-Verfahren besitzt daher höchstens die Konvergenzrate

$$R_{CG} = \frac{\sqrt{\kappa(A)} - 1}{\sqrt{\kappa(A)} + 1}$$

und die asymptotische Konvergenzgeschwindigkeit $A_{CG} \propto 1/\sqrt{\kappa(A)}$.

Weitere Krylov-Unterraum-Verfahren

Die Idee des CG-Verfahrens kann erfolgreich auf weitere Klassen von Matrizen übertragen werden. Für hermitesch indefinite Systeme generiert MINRES [51] die Iterierten so, daß das Residuum minimal ist,

$$\|r^{(k)}\|_2 = \|b - Ax^{(k)}\|_2 = \min_{x \in x^{(0)} + K_k(A, r^{(0)})} \|b - Ax\|_2. \quad (1.20)$$

Sowohl das CG-Verfahren als auch MINRES basieren auf dem symmetrischen Lanczos-Verfahren [44], welches für nicht-hermitesche Matrizen nicht geeignet ist. In diesem Fall werden statt dessen Krylov-Unterraum-Verfahren verwendet, die auf dem unsymmetrischen Lanczos-Verfahren beruhen. Prototyp dieser Krylov-Unterraum-Verfahren ist das Bi-CG [45], [24], in dem die Iterierte so konstruiert wird, daß das Residuum orthogonal zum Krylov-Unterraum der Adjungierten steht,

$$r^{(k)} \perp K_k(A^H, \hat{r}), \quad \hat{r} \in \mathbb{C}^n. \quad (1.21)$$

Da jedoch die Galerkin-Bedingung (1.21) im allgemeinen keine Minimalitätseigenschaft des Residuums $r^{(k)}$ nach sich zieht, kann und wird $\|r^{(k)}\|_2$ starke Schwankungen aufweisen.

Mit QMR [25] steht ein weiteres auf dem unsymmetrischen Lanczos-Verfahren aufbauendes Krylov-Unterraum-Verfahren bereit, das zumindest eine „Quasi-Optimalität“ des Residuums erzwingt. Aus diesem Grund ist das Konvergenzverhalten des QMR typischerweise glatter als das des Bi-CG.

Eine Verallgemeinerung von MINRES ist GMRES [60], das auf der Arnoldi-Iteration [3] beruht und auch bei allgemeinen Matrizen Iterierte mit minimalen Residuen generiert. Mit Ausnahme von GMRES können die Iterierten aller zuvor erwähnten Krylov-Unterraum-Verfahren mit Hilfe einer kurzen Rekursionsvorschrift aus einer oder einigen wenigen zuvor bestimmten Iterierten berechnet werden. Aus praktischen Gründen wird daher das GMRES-Verfahren nach k Iterationen mit der zuletzt bestimmten Iterierten als Startwert erneut gestartet. Auch bei dieser GMRES(k) genannten Variante des GMRES ist die Konvergenz nicht garantiert.

Durch eine Kombination der bisher angesprochenen Krylov-Unterraum-Verfahren stehen hybride Krylov-Unterraum-Verfahren zur Verfügung. So ist beispielsweise das BiCGStab [67] als Kombination des Bi-CG und des GMRES(1) ein Krylov-Unterraum-Verfahren, bei dem das häufig irreguläre Konvergenzverhalten des Bi-CG durch eine lokale Minimierung mittels

GMRES(1) geglättet wird. Eine Erweiterung zu BiCGStab(l) wurde in [64] vorgestellt.

Einen Überblick über Krylov-Unterraum-Verfahren findet man u.a. in [6, 15, 59, 62, 61]. Des weiteren existieren eine Vielzahl von Krylov-Unterraum-Verfahren, welche insbesondere für Koeffizientenmatrizen mit speziellen Eigenschaften angepaßt bzw. entwickelt wurden [30]. So wurden beispielsweise moderne Krylov-Unterraum-Verfahren wie QMR und GMRES besonders effizient für Familien geschifteter Systeme [28, 27, 29], polynomiale Systeme [23] und Systeme mit nicht-trivialen Symmetrien [53] erweitert und modifiziert.

1.4 Präkonditionierung

Moderne Krylov-Unterraum-Verfahren haben sich als iterative Lösungsverfahren großer linearer Gleichungssysteme für gut konditionierte Matrizen bewährt. In der Praxis treten jedoch häufig Systeme auf, die mit den modernen Krylov-Unterraum-Verfahren allein nur schwer zu lösen sind. Ihre Popularität verdanken Krylov-Unterraum-Verfahren aber der Tatsache, daß durch eine Präkonditionierung genannte Transformation des Systems die Konvergenzgeschwindigkeit der Verfahren selbst bei diesen realen Problemen meist drastisch verbessert werden kann [4, 59].

Präkonditioniertes System

Die Konvergenzgeschwindigkeit iterativer Verfahren hängt im wesentlichen von den Spektraleigenschaften der beteiligten Koeffizientenmatrix ab, welche in vielen Anwendungen durch eine Präkonditionierung, d.h. durch geeignete Transformation vom ursprünglichen System hin zu einem äquivalenten System

$$\widehat{A}\widehat{x} = \widehat{b}, \quad \widehat{A} = M_1^{-1}AM_2^{-1}, \quad \widehat{x} = M_2x \quad \widehat{b} = M_1^{-1}b, \quad (1.22)$$

beschleunigt werden kann.

Definition 1.39. Die Matrizen M_1 und M_2 aus (1.22) heißen linker bzw. rechter Präkonditionierer, die Matrix $M = M_1M_2$ Präkonditionierer von A .

Die Spaltung des Präkonditionierers M in M_1 und M_2 ist für gewöhnlich nicht nötig, da sich Krylov-Unterraum-Verfahren im allgemeinen so umformulieren lassen, daß pro Iterationsschritt die Systeme

$$Mu = v \quad \text{bzw.} \quad M^H u = v \quad (1.23)$$

zu lösen sind. Jedoch ist dies nicht immer möglich, so daß es in einigen Fällen sinnvoll sein kann, ausschließlich eine rechte ($M_1 = I$) oder linke ($M_2 = I$) Präkonditionierung zu wählen, um den durch die Präkonditionierung entstehenden zusätzlichen Aufwand pro Iteration gering zu halten.

Eigenschaften eines Präkonditionierers

Ziel bei der Präkonditionierung iterativer Verfahren ist die Konstruktion eines Präkonditionierers M , so daß der Übergang von dem ursprünglichen (1.1) zu dem präkonditionierten System (1.22) einen Geschwindigkeitsgewinn bedeutet. Aus diesem Grund fordern wir von einem Präkonditionierer insbesondere die folgenden Eigenschaften:

- a) Die Matrix M ist eine Approximation von A .
- b) Die Konstruktion von M ist nicht prohibitiv.
- c) Das Lösen der Systeme mit M ist einfach.

Die erste Forderung führt zu einer Verbesserung der Konvergenzrate, die zweite beschränkt den Initialisierungsaufwand des Präkonditionierers und die dritte Forderung limitiert den durch die Präkonditionierung resultierenden zusätzlichen Aufwand pro Iteration. Ferner sind weitergehende Forderungen, beispielsweise die der parallelen Realisierung des Präkonditionierers oder das Respektieren einer vorhandenen Symmetrieeigenschaft denkbar.

Abgebrochene Reihenentwicklungen

Als eine erste, naheliegende Möglichkeit zur Beschleunigung von Krylov-Unterraum-Verfahren bietet es sich an, als Präkonditionierer eine direkte Approximation der Inversen der Koeffizientenmatrix zu konstruieren.

Lemma 1.40. Es sei $B \in \mathbb{C}^{n \times n}$ mit $\rho(B) < 1$. Dann gilt

$$(I - B)^{-1} = \sum_{\nu=0}^{\infty} B^{\nu}. \quad (1.24)$$

Beweis. [26], Satz A.2.2. □

Lemma 1.41. Es sei $A = P - Q$ eine Zerlegung von A mit $\rho(P^{-1}Q) < 1$. Dann gilt

$$A^{-1} = \sum_{\nu=0}^{\infty} (P^{-1}Q)^{\nu} P^{-1}. \quad (1.25)$$

Beweis. Es ist $A = P(I - P^{-1}Q)$ mit $\rho(P^{-1}Q) < 1$. Die Behauptung folgt mit Lemma 1.40. \square

Trunkieren wir die Reihenentwicklung (1.25), so erhalten wir als Präkonditionierer

$$M_k^{-1} = \sum_{\nu=0}^{k-1} (P^{-1}Q)^\nu P^{-1}. \quad (1.26)$$

Die Präkonditionierung durch eine abgebrochene Reihenentwicklung (1.26), also das Berechnen des Vektors $u = M_k^{-1}v$, kann auch interpretiert werden als Ausführen von k Schritten eines einfachen Iterationsverfahrens:

Lemma 1.42. Es sei $A = P - Q$ eine Zerlegung von $A \in \mathbb{C}^{n \times n}$. Ist $x^{(k)}$ Iterierte des Iterationsverfahrens

$$Px^{(k+1)} = Qx^{(k)} + b, \quad k = 0, 1, \dots$$

mit Startwert $x^{(0)} = 0$, so gilt

$$x^{(k)} = \sum_{\nu=0}^{k-1} (P^{-1}Q)^\nu P^{-1}b \quad k = 0, 1, \dots$$

Beweis. Per Induktion über k . \square

Weitere Präkonditionierungstechniken

Neben der abgebrochenen Reihenentwicklung als eine problemunabhängige Standardmöglichkeit zur Beschleunigung von Krylov-Unterraum-Verfahren und den im nächsten Kapitel ausführlicher dargestellten Multilevel-Verfahren existieren andere Präkonditionierungstechniken.

Genannt seien hier neben der blockzyklischen Reduktion [16] und der Präkonditionierung durch dünn besetzte approximative Inverse [17, 33] insbesondere solche problemspezifischen Methoden, die beispielsweise auf einer Zerlegung der Matrix in zwei oder mehrere Operatoren basieren [31] und so das zugrunde liegende kontinuierliche Problem besser reflektieren.

Da diese Vorgehensweisen jedoch nicht Gegenstand der vorliegenden Arbeit sind, verweisen wir für eine detaillierte Beschreibung dieser und anderer Methoden auf die einschlägige Literatur.

Kapitel 2

Multilevel-Verfahren

Multilevel-Verfahren wie Mehrgitterverfahren (engl.: multigrid methods) und Gebietszerlegungsmethoden (engl.: domain decomposition methods) sind weitere Iterationsverfahren, welche sich insbesondere zur numerischen Lösung partieller Differentialgleichungen eignen. Der Grund hierfür ist, daß unter gewissen Voraussetzungen der Aufwand bei diesen Methoden proportional zur Anzahl der Unbekannten und die Konvergenzrate unabhängig von der Diskretisierungsschrittweite ist. Damit zählen Multilevel-Verfahren potentiell zu den schnellsten iterativen Lösungsverfahren für lineare Gleichungssysteme. Aufgrund dieser Optimalität wird meist auf die Kombination von Multilevel- und Krylov-Unterraum-Verfahren verzichtet, wobei aber gerade die Verbindung von Krylov-Unterraum-Verfahren mit Multilevel-Iterationen zur Prädiktionierung für viele in der Praxis relevante Aufgabenstellungen sehr effiziente Iterationsverfahren liefert.

Da kein universal als Black-Box einsetzbares Multilevel-Verfahren existiert, sondern die Komponenten von Multilevel-Verfahren problemspezifisch gewählt werden müssen, werden wir uns in diesem Kapitel eher konzeptionell mit Multilevel-Verfahren beschäftigen.

2.1 Mehrgitterverfahren

In diesem Abschnitt werden wir die Idee der Zwei- und Mehrgitterverfahren [36, 37, 10] beschreiben. Dazu stellen wir die zwei wesentlichen Komponen-

ten „Glättungsiteration“ und „Grobgitterkorrektur“ vor und zeigen auf, wie diese sich als Bestandteile eines Zweigitterverfahrens ergänzen. Anschließend werden wir echte Mehrgitterverfahren motivieren.

2.1.1 Glättungsiteration

Der Ausgangspunkt zur Herleitung von Mehrgitterverfahren ist eine Analyse der Fehler einer Folge von Iterierten, welche mittels eines einfachen, stationären Iterationsverfahrens generiert wurde. Exemplarisch stellen wir das relaxierte Jacobi-Verfahren zur Lösung des Dirichletschen Randwertproblems (1.2) vor und zeigen, daß es gute Glättungseigenschaften besitzt.

Lemma 2.1. Die Matrix A aus (1.3) besitzt die n Eigenwerte

$$\lambda^{(k,l)} = 4 - 2 \left(\cos \left(\frac{k\pi}{N+1} \right) + \cos \left(\frac{l\pi}{N+1} \right) \right), \quad 1 \leq k, l \leq N$$

mit den zugehörigen Eigenvektoren $z^{(k,l)}$,

$$z_{i,j}^{(k,l)} = \sin \left(\frac{k\pi i}{N+1} \right) \sin \left(\frac{l\pi j}{N+1} \right), \quad 1 \leq i, j \leq N. \quad (2.1)$$

Beweis. [38], Abschnitt 4.1. □

Betrachten wir die Eigenvektoren $z^{(k,l)}$ als Gitterfunktionen, vgl. Abb. 2.1, so ist es naheliegend, die Eigenvektoren in zwei Klassen, nämlich in niedrig- und hochfrequente Eigenvektoren, einzuteilen:

Definition 2.2. Wir bezeichnen den durch die hochfrequenten Eigenvektoren aufgespannten Unterraum

$$X = \text{span} \left(z^{(i,j)} \mid 1 \leq i, j \leq N, \max\{i, j\} > \frac{N}{2} \right) \quad (2.2)$$

als Raum der oszillierenden Gitterfunktionen und den durch die niedrigfrequenten Eigenvektoren aufgespannten Unterraum

$$Y = \text{span} \left(z^{(i,j)} \mid 1 \leq i, j \leq \frac{N}{2} \right) \quad (2.3)$$

als Raum der glatten Gitterfunktionen.

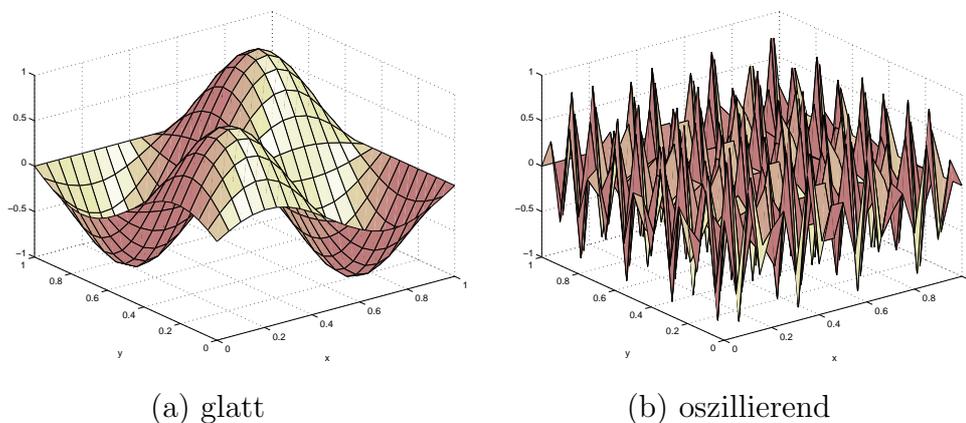


Abbildung 2.1: Glatter Eigenvektor $z^{(2,2)}$ und oszillierender Eigenvektor $z^{(17,17)}$ auf einem Gitter Ω_h mit $N = 20$.

Unser vorläufiges Ziel ist es, das Konvergenzverhalten des relaxierten Jacobi-Verfahrens auf dem Raum der oszillierenden Gitterfunktionen zu analysieren. Dazu bemerken wir, daß die Eigenvektoren (2.1) der Matrix A und die der Iterationsmatrix

$$H = I - \frac{\omega}{4}A, \quad \omega \in \mathbb{R} \setminus \{0\} \quad (2.4)$$

des relaxierten Jacobi-Verfahrens offensichtlich identisch sind, und da diese Eigenvektoren eine Orthogonalbasis bilden [38], läßt sich der Fehler in der Eigenvektorbasis schreiben als

$$e^{(k)} = \sum_{i,j=1}^N \alpha_{i,j} z^{(i,j)}. \quad (2.5)$$

Unter der Voraussetzung, daß der Fehler ausschließlich aus hochfrequenten Eigenvektoren besteht, liefert das folgende Lemma die Konvergenz des Jacobi-Verfahrens unabhängig von der zugrunde liegenden Diskretisierungsschrittweite h .

Lemma 2.3. Gilt $e^{(0)} \in X$, so folgt $e^{(k)} \in X$, und für $\omega = \frac{4}{5}$ genügen die Fehler $e^{(k)}$ der Abschätzung

$$\|e^{(k)}\|_2 \leq \left(\frac{3}{5}\right)^k \|e^{(0)}\|_2. \quad (2.6)$$

Beweis. Analog zum Beweis von Lemma 10.1.1 in [38]. □

Spalten wir den Anfangsfehler auf in einen niedrig- und einen hochfrequenten Anteil, so können wir folgern, daß durch einige wenige Schritte des relaxierten Jacobi-Verfahrens die hochfrequenten Anteile des Fehlers stark reduziert werden. Die niedrigfrequenten Anteile des Fehlers variieren als Funktion auf dem Gitter vergleichsweise langsam. Anstelle einer Reduktion erfolgt also eine Glättung des Gesamtfehlers. Iterationsverfahren mit dieser Eigenschaft bezeichnen wir als *Glättungsiteration*, die zugehörige Iterationsmatrix als *Glätter*.

Lemma 2.4. Nach ν Schritten einer auf der Zerlegung $A = M - N$ der Matrix A basierenden Glättungsiteration erhalten wir als Iterierte

$$x^{(\nu)} = S^\nu x^{(0)} + \sum_{l=0}^{\nu-1} S^l M^{-1} b, \quad S = M^{-1} N,$$

und für den Fehler gilt $e^{(\nu)} = S^\nu e^{(0)}$.

Beweis. Per Induktion über ν . □

Algorithmus 2.1 Glättungsiteration

```

1: procedure  $S(x, b, \nu)$ 
2:    $x^{(0)} = x$ 
3:   for  $l = 0, \dots, \nu - 1$  do
4:      $Mx^{(l+1)} = Nx^{(l)} + b$ 
5:   end for
6:    $x = x^{(\nu)}$ 
7: end

```

Wir werden später den Begriff der Glättungseigenschaft präzisieren, vgl. Satz 2.23.

2.1.2 Grobgitterkorrektur

Im letzten Abschnitt haben wir gezeigt, daß eine Glättungsiteration den oszillierenden Anteil des Fehlers in einigen wenigen Iterationsschritten zu reduzieren vermag. Dieser Teil des Fehlers liegt in dem Raum der Eigenvektoren zu großen Eigenwerten, also hohen Frequenzen. Die Anteile zu niedrigen Frequenzen hingegen werden nicht oder kaum reduziert.

Das Ziel ist nun die Konstruktion eines Zweigitterverfahrens als eine Produktiteration, bestehend aus einigen Schritten einer Glättungsiteration gefolgt

von einem komplementären Verfahren, der sogenannten Grobgitterkorrektur, welches die niedrigfrequenten Anteile des Fehlers schnell reduziert. Dazu werden niedrige Frequenzen von dem originären Gitter auf ein gröberes Gitter transferiert, wo diese sich wie hohe Frequenzen verhalten und somit durch einfache Iterationsverfahren wieder schnell geglättet werden können. Eine rekursive Vorgehensweise dieser Idee führt zu echten Mehrgitterverfahren.

Im Gegensatz zu den zuvor in Kapitel 1 vorgestellten Iterationsverfahren (mit Ausnahme von GMRES) berücksichtigt ein Mehrgitterverfahren (bei Nächste-Nachbar-Kopplungen repräsentierenden Koeffizientenmatrizen) bei der Aufdatierung der Iterierten nicht nur die auf dem nächsten Gitterpunkt vorliegende Information, sondern auch Information auf bzgl. des originären Gitters weit entfernt liegenden Gitterpunkten. Dies ist der Grund dafür, daß bei Mehrgitterverfahren das Phänomen des *critical slowing down* eliminiert werden kann [19].

Zweigitterverfahren

Zur Berechnung der exakten Lösung $x_{h_0}^* = A_{h_0}^{-1}b_{h_0}$ des linearen Gleichungssystems

$$A_{h_0}x_{h_0} = b_{h_0}, \quad A_{h_0} = A, \quad b_{h_0} = b \quad (2.7)$$

auf dem Gitter Ω_{h_0} betrachten wir die äquivalente Residualgleichung

$$A_{h_0}e_{h_0} = r_{h_0}, \quad e_{h_0} = x_{h_0}^* - x_{h_0}, \quad r_{h_0} = b_{h_0} - A_{h_0}x_{h_0}. \quad (2.8)$$

Ist die Gitterfunktion $x_{h_0}^{(k)} \in \Omega_{h_0}$ das Resultat einiger Schritte einer einfachen Glättungsiteration, so ist der zugehörige Fehler $e_{h_0}^{(k)}$ nach obigen Überlegungen glatt und kann deshalb auf einem gröberen Gitter Ω_{h_1} approximiert werden.

Zur Konstruktion eines approximativen Fehlers wird das Residuum $r_{h_0}^{(k)}$ durch eine Restriktion $R_{h_0}^{h_1} : \Omega_{h_0} \rightarrow \Omega_{h_1}$ auf das Gitter Ω_{h_1} transferiert. Anschließend wird eine Approximation $e_{h_1}^{(k)}$ des Fehlers $e_{h_0}^{(k)}$ auf dem groben Gitter Ω_{h_1} als Lösung eines geeigneten Grobgittersystems mit einer Grobgittermatrix A_{h_1} bestimmt,

$$A_{h_1}e_{h_1}^{(k)} = r_{h_1}^{(k)}, \quad r_{h_1} = R_{h_0}^{h_1}r_{h_0}^{(k)}. \quad (2.9)$$

Danach wird die Lösung $e_{h_1}^{(k)}$ des Grobgittersystems mittels einer Prolongation $P_{h_1}^{h_0} : \Omega_{h_1} \rightarrow \Omega_{h_0}$ auf das originäre Gitter Ω_{h_0} interpoliert, und als neue

Iterierte die Gitterfunktion

$$x_{h_0}^{(k+1)} = x_{h_0}^{(k)} + P_{h_1}^{h_0} e_{h_1}^{(k)} \quad (2.10)$$

berechnet.

Zusammenfassend erhalten wir:

Lemma 2.5. Für eine Gitterfunktion $x_{h_0}^{(k)} \in \mathbb{C}^n$ liefert die Grobgitterkorrektur als neue Iterierte

$$x_{h_0}^{(k+1)} = (I - P_{h_1}^{h_0} A_{h_1}^{-1} R_{h_0}^{h_1} A_{h_0}) x_{h_0}^{(k)} + P_{h_1}^{h_0} A_{h_1}^{-1} R_{h_0}^{h_1} b_{h_0}.$$

Beweis. Per Konstruktion. □

Algorithmus 2.2 Grobgitterkorrektur

- 1: **procedure** GGK(x_{h_0}, b_{h_0})
 - 2: $r_{h_0} = b_{h_0} - A_{h_0} x_{h_0}$ {Residuum}
 - 3: $r_{h_1} = R_{h_0}^{h_1} r_{h_0}$ {Restriktion}
 - 4: $A_{h_1} e_{h_1} = r_{h_1}$ {Grobgittersystem}
 - 5: $x_{h_0} = x_{h_0} + P_{h_1}^{h_0} e_{h_1}$ {Prolongation}
 - 6: **end**
-

Kombinieren wir die Idee der Glättungsiteration und die der Grobgitterkorrektur, so erhalten wir als Produktiteration eine Zweigitteriteration, s. Alg. 2.3, und wir können gemäß unseren vorherigen Überlegungen beweisen:

Satz 2.6. Das Zweigitterverfahren ist ein Iterationsverfahren der Gestalt (1.5), und die zugehörige Iterationsmatrix H lautet

$$H = S_{post}^{\nu_2} (I - P_{h_1}^{h_0} A_{h_1}^{-1} R_{h_0}^{h_1} A_{h_0}) S_{pre}^{\nu_1}.$$

Dabei bezeichnen ν_1 und ν_2 die Anzahl der Vor- und Nachglättungsschritte mit dem Vorglätter S_{pre} und Nachglätter S_{post} .

Beweis. Mit Lemma 2.4 und Lemma 2.5 ist der Beweis trivial. □

Algorithmus 2.3 Zweigitteriteration

```

1: procedure ZG( $x_{h_0}, b_{h_0}$ )
2:    $x_{h_0} = S^{pre}(x_{h_0}, b_{h_0}, \nu_1)$            {Vorglättung}
3:    $r_{h_0} = b_{h_0} - A_{h_0}x_{h_0}$              {Residuum}
4:    $r_{h_1} = R_{h_0}^{h_1}r_{h_0}$                  {Restriktion}
5:    $A_{h_1}e_{h_1} = r_{h_1}$                        {Grobgittersystem}
6:    $x_{h_0} = x_{h_0} + P_{h_1}^{h_0}e_{h_1}$        {Prolongation}
7:    $x_{h_0} = S^{post}(x_{h_0}, b_{h_0}, \nu_2)$    {Nachglättung}
8: end

```

Mehrgitterverfahren

Es ist naheliegend, anstelle der exakten Lösung des Grobgittersystems (2.9) rekursiv eine Approximation durch Anwenden von $\gamma(l)$ Iterationsschritten eines so modifizierten Zweigitterverfahrens zu berechnen. Dies ist das Mehrgitterverfahren.

Natürlich muß auch im Falle einer Mehrgitteriteration ein Gleichungssystem auf dem größten Gitter exakt gelöst werden. Dies stellt in der Praxis jedoch kein Hindernis dar, sofern das Gitter und somit auch das entsprechende System hinreichend klein ist.

Algorithmus 2.4 Mehrgitteriteration

```

1: procedure MG( $x_{h_l}, b_{h_l}, l$ )
2:   if  $l = l_{\max}$  then
3:      $x_{h_l} = A_{h_l}^{-1}b_{h_l}$ 
4:   else
5:      $x_{h_l} = S_l^{pre}(x_{h_l}, b_{h_l}, l, \nu_1(l))$    {Vorglättung}
6:      $r_{h_{l+1}} = R_{l+1,l}(b_{h_l} - A_{h_l}x_{h_l})$    {Residuum}
7:      $e_{h_{l+1}} = 0$                                  {Grobgittersystem}
8:     for  $i = 1, \dots, \gamma(l)$  do
9:       MG( $e_{h_{l+1}}, r_{h_{l+1}}, l + 1$ )
10:    end for
11:     $x_{h_l} = x_{h_l} + P_{l,l+1}e_{h_{l+1}}$            {Prolongation}
12:     $x_{h_l} = S_l^{post}(x_{h_l}, b_{h_l}, \nu_2(l))$    {Nachglättung}
13:  end if
14: end

```

2.2 Schurkomplement-Methoden

Schurkomplement-Methoden zählen zu den Gebietszerlegungsmethoden. Ausgehend von einer sogenannten F/C-Anordnung der Gitterpunkte werden wir zunächst mit Hilfe der exakten Schurkomplement-Zerlegung die Inverse der Koeffizientenmatrix angeben. Anschließend erhalten wir durch eine Approximation der exakten Inversen die sogenannten Schurkomplement-Präkonditionierer.

F/C-Anordnung der Gitterpunkte

Wir partitionieren die Gitterpunkte $\Omega = \Omega_f \cup \Omega_c$ in eine Menge von *feinen* (engl.: fine) Gitterpunkten Ω_f und eine Menge von *groben* (engl.: coarse) Gitterpunkten Ω_c ,

$$\Omega = \Omega_f \cup \Omega_c, \quad \Omega_f \cap \Omega_c = \emptyset. \quad (2.11)$$

Werden zunächst alle feinen Gitterpunkte und anschließend alle groben Gitterpunkte gezählt, so besitzt die Matrix $A \in \mathbb{C}^{n \times n}$ bei dieser F/C-Anordnung der Gitterpunkte die Blockgestalt

$$A = \begin{pmatrix} A_{ff} & A_{fc} \\ A_{cf} & A_{cc} \end{pmatrix}. \quad (2.12)$$

Im folgenden setzen wir stets voraus, daß A_{ff} nichtsingulär ist.

Bemerkung 2.7. Die Annahme eines Gitters ist „rein willkürlich“, denn die hier vorgestellten Schurkomplement-Methoden lassen sich für jede beliebige Einteilung der Komponenten bzw. Unbekannten formulieren.

Die Schurkomplement-Zerlegung

Lemma 2.8. Ist A eine Matrix der Gestalt (2.12), so ist das Schurkomplement $S_A = A_{cc} - A_{cf}A_{ff}^{-1}A_{fc}$ nichtsingulär.

Beweis. Nach Voraussetzung sind die Matrizen A und A_{ff} nichtsingulär. Somit ist die Behauptung eine unmittelbare Folgerung aus der Schurkomplement-Zerlegung

$$A = \begin{pmatrix} I & 0 \\ A_{cf}A_{ff}^{-1} & I \end{pmatrix} \begin{pmatrix} A_{ff} & 0 \\ 0 & S_A \end{pmatrix} \begin{pmatrix} I & A_{ff}^{-1}A_{fc} \\ 0 & I \end{pmatrix} \quad (2.13)$$

der Matrix A . □

Die Inverse der Matrix A kann explizit angegeben werden:

Lemma 2.9. Es sei A eine Matrix der Gestalt (2.12). Dann gilt

$$A^{-1} = \begin{pmatrix} A_{ff}^{-1} & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} -A_{ff}^{-1}A_{fc} \\ I \end{pmatrix} S_A^{-1} \begin{pmatrix} -A_{cf}A_{ff}^{-1} & I \end{pmatrix}. \quad (2.14)$$

Beweis. Ausgehend von der Zerlegung (2.13) der Matrix A folgt

$$A^{-1} = \begin{pmatrix} I & -A_{ff}^{-1}A_{fc} \\ 0 & I \end{pmatrix} \begin{pmatrix} A_{ff}^{-1} & 0 \\ 0 & S_A^{-1} \end{pmatrix} \begin{pmatrix} I & 0 \\ -A_{cf}A_{ff}^{-1} & I \end{pmatrix}. \quad (2.15)$$

Durch Ausrechnen erhalten wir schließlich die Behauptung. \square

Block-Gauß-Elimination

Sind die Matrizen A_{ff}^{-1} und S_A explizit bekannt, so erhalten wir durch eine Transformation der Matrix A auf Blockdiagonalgestalt,

$$\begin{pmatrix} I & 0 \\ -A_{cf}A_{ff}^{-1} & I \end{pmatrix} \begin{pmatrix} A_{ff} & A_{fc} \\ A_{cf} & A_{cc} \end{pmatrix} \begin{pmatrix} I & -A_{ff}^{-1}A_{fc} \\ 0 & I \end{pmatrix} = \begin{pmatrix} A_{ff} & 0 \\ 0 & S_A \end{pmatrix},$$

und anschließendes Lösen des zu $Ax = b$ äquivalenten Systems,

$$\begin{pmatrix} A_{ff} & 0 \\ 0 & S_A \end{pmatrix} \begin{pmatrix} I & A_{ff}^{-1}A_{fc} \\ 0 & I \end{pmatrix} \begin{pmatrix} x_f \\ x_c \end{pmatrix} = \begin{pmatrix} I & 0 \\ -A_{cf}A_{ff}^{-1} & I \end{pmatrix} \begin{pmatrix} b_f \\ b_c \end{pmatrix},$$

ein auf der exakten Block-Gauß-Elimination beruhendes, prinzipiell direktes Lösungsverfahren, vgl. Alg. 2.5.

Algorithmus 2.5 Direkter Löser des Systems $Ax = b$

- 1: $A_{ff}z_f = b_f$
 - 2: $S_Ax_c = b_c - A_{cf}z_f$
 - 3: $A_{ff}x_f = b_f - A_{fc}x_c$
-

Zur Lösung sind insgesamt zwei Matrix-Vektor-Multiplikationen mit A_{ff}^{-1} und das Lösen eines kleineren Systems mit S_A auf den groben Gitterpunkten erforderlich. Diese direkte Vorgehensweise läßt sich als optimale matrix-abhängige Grobgitterkorrektur auffassen.

Lemma 2.10. Die Matrizen R und P seien definiert als

$$R = \begin{pmatrix} -A_{cf}A_{ff}^{-1} & I \end{pmatrix}, \quad P = \begin{pmatrix} -A_{ff}^{-1}A_{fc} \\ I \end{pmatrix}.$$

Dann gilt $RAP = S_A$.

Beweis. Aus der Identität

$$\begin{aligned} RAP &= \begin{pmatrix} -A_{cf}A_{ff}^{-1} & I \end{pmatrix} \begin{pmatrix} A_{ff} & A_{fc} \\ A_{cf} & A_{cc} \end{pmatrix} \begin{pmatrix} -A_{ff}^{-1}A_{fc} \\ I \end{pmatrix} \\ &= \begin{pmatrix} -A_{cf}A_{ff}^{-1} & I \end{pmatrix} \begin{pmatrix} 0 \\ S_A \end{pmatrix} \end{aligned}$$

folgt sofort die Behauptung. \square

Schurkomplement-Präkonditionierer

Da die Matrix A_{ff} im allgemeinen nicht diagonal ist, ist die zugehörige Inverse dicht besetzt und die oben beschriebene direkte Vorgehensweise nicht praktikabel. Daher ersetzen wir sowohl das exakte Schurkomplement S_A durch eine Schurkomplement-Approximation S als auch die Inverse der Matrix A_{ff} durch eine lokal wirkende approximative Inverse. Als resultierenden Präkonditionierer M bekommen wir die Matrix

$$M = \begin{pmatrix} I & 0 \\ A_{cf}\tilde{A}_{ff}^{-1} & I \end{pmatrix} \begin{pmatrix} \tilde{A}_{ff} & 0 \\ 0 & S \end{pmatrix} \begin{pmatrix} I & \tilde{A}_{ff}^{-1}A_{fc} \\ 0 & I \end{pmatrix}. \quad (2.16)$$

Ein Blick auf (2.16) zeigt, daß Schurkomplement-Präkonditionierer auf einer unvollständigen Block-Faktorisierung der Koeffizientenmatrix basieren.

Analog zu Lemma 2.9 erhalten wir als Inverse

$$M^{-1} = \begin{pmatrix} \tilde{A}_{ff}^{-1} & 0 \\ 0 & 0 \end{pmatrix} + PS^{-1}R, \quad (2.17)$$

wobei die Matrizen R und P definiert sind als

$$R = \begin{pmatrix} -A_{cf}\tilde{A}_{ff}^{-1} & I \end{pmatrix}, \quad P = \begin{pmatrix} -\tilde{A}_{ff}^{-1}A_{fc} \\ I \end{pmatrix}. \quad (2.18)$$

Die Matrix R kann als *matrixabhängige Restriktion*, P als *matrixabhängige Prolongation* und die Schurkomplement-Approximation S als Grobgittermatrix interpretiert werden. Der verbleibende Term entspricht der Glättung im Mehrgitter-Kontext, sie wird hier aber nur auf den feinen Gitterpunkten durchgeführt. Matrixabhängige Restriktionen und Prolongationen wurden für spezielle (symmetrisch positiv definite) Probleme und bestimmte Klassen von Matrizen (M-Matrizen) intensiv studiert [2, 20, 18, 55, 73].

Im Vergleich zu klassischen Mehrgitterverfahren zur Präkonditionierung ist das Zusammenspiel von Glättungsiteration und Grobgitterkorrektur bei Schurkomplement-Präkonditionierern nicht wesentlich, statt dessen liegt der Schlüssel in der Konstruktion geeigneter Schurkomplement-Approximationen.

Lemma 2.11. Ist $\tilde{A}_{ff} = A_{ff}$, so gilt $\sigma(M^{-1}A) = \sigma(S^{-1}S_A) \cup \{1\}$.

Beweis. Aus der Ähnlichkeit der beteiligten Matrizen,

$$M^{-1}A = \begin{pmatrix} I & -A_{ff}^{-1}A_{fc} \\ 0 & I \end{pmatrix} \begin{pmatrix} I & 0 \\ 0 & S^{-1}S_A \end{pmatrix} \begin{pmatrix} I & A_{ff}^{-1}A_{fc} \\ 0 & I \end{pmatrix},$$

folgt unmittelbar die Behauptung. \square

Eine geeignete Schurkomplement-Approximation vermag es, daß das Spektrum des präkonditionierten Schurkomplements hinreichend gut um Eins *geclustert* wird. Dies bedeutet insbesondere, daß die Schurkomplement-Approximation die Wirkung des exakten Schurkomplements auf Eigenvektoren zu betragskleinen Eigenwerten möglichst genau widerspiegeln sollte.

Natürlich ist die in Lemma 2.11 getroffene Annahme $\tilde{A}_{ff} = A_{ff}$ in der Praxis nicht haltbar. Statt dessen wird die Matrix A_{ff} approximiert.

Lemma 2.12. Es sei $J = \begin{pmatrix} 0 & I \end{pmatrix}$ die triviale Restriktion. Die Matrizen R , P , Q , E und F seien definiert durch

$$P = \begin{pmatrix} I & P_{fc} \\ 0 & I \end{pmatrix}, \quad R = \begin{pmatrix} I & 0 \\ R_{cf} & I \end{pmatrix}, \quad Q = \begin{pmatrix} \tilde{A}_{ff} & 0 \\ 0 & J(RAP)J^T \end{pmatrix}$$

sowie

$$E = \begin{pmatrix} \tilde{A}_{ff} - A_{ff} & 0 \\ 0 & 2R_{cf}A_{ff}P_{fc} + R_{cf}A_{fc} + A_{cf}P_{fc} \end{pmatrix},$$

$$F = \begin{pmatrix} 0 & A_{fc} + A_{ff}P_{fc} \\ A_{cf} + R_{cf}A_{ff} & 0 \end{pmatrix}.$$

Dann gilt $R^{-1}(Q - E)P^{-1} = A - F$.

Beweis. [63]. \square

Das Lemma 2.12 zeigt, daß der Schurkomplement-Präkonditionierer

$$\begin{aligned} (PQ^{-1}R)^{-1} &= R^{-1}QP^{-1} \\ &= \begin{pmatrix} I & 0 \\ -R_{cf} & I \end{pmatrix} \begin{pmatrix} \tilde{A}_{ff} & 0 \\ 0 & J(RAP)J^T \end{pmatrix} \begin{pmatrix} I & -P_{fc} \\ 0 & I \end{pmatrix} \end{aligned}$$

mit

$$J(RAP)J^T = A_{cf}P_{fc} + R_{cf}A_{fc} + R_{cf}A_{ff}P_{fc}$$

wegen

$$R^{-1}QP^{-1} = A - F + R^{-1}EP^{-1}$$

eine Störung der Matrix A ist (s. (2.13)),

$$A = \begin{pmatrix} I & 0 \\ A_{cf}A_{ff}^{-1} & I \end{pmatrix} \begin{pmatrix} A_{ff} & 0 \\ 0 & A_{cc} - A_{cf}A_{ff}^{-1}A_{fc} \end{pmatrix} \begin{pmatrix} I & A_{ff}^{-1}A_{fc} \\ 0 & I \end{pmatrix},$$

bei der die Matrizen A_{ff} , $A_{ff}^{-1}A_{fc}$ und $A_{cf}A_{ff}^{-1}$ durch die Matrizen \tilde{A}_{ff} , $-P_{fc}$ bzw. $-R_{cf}$ ersetzt werden.

2.3 Grobgitter

Allen Multilevel-Verfahren ist gemein, daß sie eine Hierarchie von Grobgittern benötigen. Prinzipiell kann die Definition eines Grobgitters aufgrund geometrischer Information oder rein algebraischer Kriterien erfolgen.

Für die im Rahmen dieser Arbeit betrachteten Probleme erweisen sich die aus einer uniformen Diskretisierung des Einheitswürfels $I = [0, 1]^d$ mit $d = 2, 4$ resultierenden Gitter Ω_h als ausreichend. Der Einfachheit halber betrachten wir hier den zweidimensionalen Fall

$$\Omega_h = \left\{ (x_i, x_j) \mid (x_i, x_j) = (ih, jh), \quad 1 \leq i, j \leq N, \quad h = \frac{1}{N+1} \right\}.$$

Einfärbung der Gitterpunkte

Zur mathematischen Definition in den Anwendungen häufig verwendeter Grobgitter und der späteren Definition von geeigneten Restriktionen und Prolongationen für die im Rahmen dieser Arbeit untersuchten Matrizen wird es sich als überaus hilfreich erweisen, eine spezielle Einfärbung der Gitterpunkte vorzunehmen:

Definition 2.13. Die Menge $L_i(\Omega_h)$ von Gitterpunkten sei definiert durch

$$\begin{aligned} L_0(\Omega_h) &= \{x = (ih, jh) \in \Omega_h \mid i, j \text{ ungerade}\} \\ L_2(\Omega_h) &= \{x = (ih, jh) \in \Omega_h \mid i, j \text{ gerade}\} \\ L_1(\Omega_h) &= \{x = (ih, jh) \in \Omega_h \mid i + j \text{ ungerade}\}. \end{aligned}$$

Offensichtlich ist $L_1(\Omega_h)$ die Menge der ungeraden Gitterpunkte, und die verbleibende Menge $L_0(\Omega_h) \cup L_2(\Omega_h)$ besteht aus den geraden Gitterpunkten.

Lemma 2.14. Durch die Mengen $L_i(\Omega_h)$ ist eine Partition der Gitterpunkte Ω_h gegeben,

$$\Omega_h = \bigcup_{i=0}^2 L_i(\Omega_h), \quad L_i(\Omega_h) \cap L_j(\Omega_h) = \emptyset \quad (i \neq j).$$

Beweis. Klar. □

Lemma 2.15. Es sei N gerade, $n = N^2$ und Ω_h mit $|\Omega_h| = n$ Gitterpunkten. Dann gilt

$$|L_0(\Omega_h)| = \frac{1}{4}n, \quad |L_1(\Omega_h)| = \frac{1}{2}n, \quad |L_2(\Omega_h)| = \frac{1}{4}n.$$

Beweis. Abzählen. □

2.3.1 Geometrisch motivierte Grobgitter

Geometrische Auswahlkriterien des Grobgitters Ω_H respektieren die regelmäßige Struktur des Gitters Ω_h und ermöglichen es, a priori die Grobgitterhierarchie zu erzeugen. Im folgenden werden zwei wichtige Grobgitter für das zweidimensionale Standard-Gitter Ω_h (1.16) vorgestellt.

Standard-Grobgitter

Das wohl beliebteste uniforme Grobgitter, das sogenannte Standard-Grobgitter,

$$\Omega_H = L_2(\Omega_h), \tag{2.19}$$

basiert auf einer Ausdünnung der Gitterpunkte Ω_h entlang jeder Dimension um einen Faktor 2. Das resultierende Grobgitter Ω_H mit Diskretisierungsschrittweite $H = 2h$ ist in Abb. 2.2 dargestellt.

Odd-Even-Grobgitter

Ein weiteres uniformes Gitter erhalten wir durch eine Ausdünnung um einen Faktor $\sqrt{2}$, vgl. Abb. 2.2. Dazu erweitern wir das Standard-Grobgitter um die verbleibenden geraden Gitterpunkte $L_0(\Omega_h)$, also

$$\Omega_H = L_0(\Omega_h) \cup L_2(\Omega_h). \tag{2.20}$$

Das resultierende Grobgitter Ω_H besteht aus allen geraden Gitterpunkten.

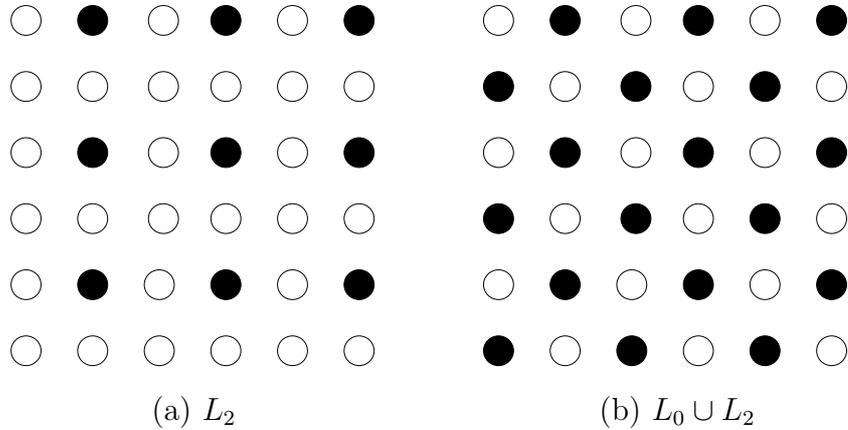


Abbildung 2.2: Standard-Grobgridter mit Grobgitterpunkten aus L_2 sowie Grobgitter bestehend aus allen geraden Gitterpunkten $L_0 \cup L_2$.

Unter der Voraussetzung, daß die Anzahl der Gitterpunkte in jeder Gitterrichtung gerade ist, können die vorgestellten Strategien zur Auswahl der Grobgitterpunkte auch bei Nächsten-Nachbar-Kopplungen mit periodischen Randbedingungen herangezogen werden. Sowohl das Standard-Grobgridter als auch das Odd-Even-Grobgridter lassen sich in kanonischer Weise auf höhere Dimensionen übertragen, daher stellt der mehrdimensionale Fall keine Schwierigkeit dar.

Weitere Grobgitter

Für spezielle Probleme ist es denkbar, das Grobgitter so zu konstruieren, daß die Grobgitterpunkte Ω_H nicht zu Ω_h gehören, sondern auf dem sogenannten dualen Gitter liegen. Da diese sogenannten gestaffelten Grobgitter im Rahmen dieser Arbeit nicht verwendet werden, verweisen wir auf die entsprechende Literatur [36, 37]. Des weiteren sind Grobgitter mit einem größeren Ausdünnungsfaktor vorstellbar, jedoch erweisen sich diese in der Praxis oft als nicht vorteilhaft.

2.3.2 Algebraisch motivierte Grobgitter

Die den algebraischen Auswahlkriterien [58, 68, 65] zugrunde liegende Heuristik ist die Glättung entlang sogenannter starker Kopplungen innerhalb von Mehrgitterverfahren. Um diese Strategie näher vorzustellen, werden einige graphentheoretische Begriffe benötigt.

Definition 2.16. Es sei $A = (a_{ij}) \in \mathbb{C}^{n \times n}$ gegeben. Dann bezeichnet $G_A(V, E)$ den Graphen der Matrix A , bestehend aus einer Knotenmenge

$V = \{v_1, \dots, v_n\}$ und einer Kantenmenge E mit

$$e_{ij} = (v_i, v_j) \in E \iff a_{ij} \neq 0, \quad i \neq j.$$

Die Kantenmenge E entspricht offensichtlich den Nichtnullelementen der Matrix A und die Knotenmenge V den Unbekannten. Wir definieren nun den Begriff der Nachbarschaft eines Knoten.

Definition 2.17. Es sei $G_A(V, E)$ der Graph der Matrix $A = (a_{ij}) \in \mathbb{C}^{n \times n}$. Die Nachbarn N_i eines Knoten $v_i \in V$ sind definiert durch

$$N_i = \{v_j \in V \mid e_{ij} \in E\}.$$

Der nächste Schritt ist eine Charakterisierung der Kanten, welche es ermöglicht, die Kantenmenge E zu zerlegen in eine Kantenmenge E_S starker Kopplungen (engl.: strong edges) und eine Kantenmenge E_W schwacher Kopplungen (engl.: weak edges),

$$E = E_S \cup E_W, \quad E_S \cap E_W = \emptyset.$$

Definition 2.18. Es sei $A = (a_{ij}) \in \mathbb{C}^{n \times n}$ und $G_A(V, E)$ der zugehörige Graph. Per Definition repräsentiert jede Kante $e_{ii} \in E$ eine starke Kopplung. Des weiteren ist die Kante $e_{ij} \in E$ eine starke Kopplung, falls die Bedingung

$$|a_{ij}| > \beta \max_{j \in N_i} |a_{ij}| \tag{2.21}$$

für einen gegebenen Parameter $0 \leq \beta < 1$ erfüllt ist. Eine typische Wahl ist $\beta = 0.5$. Jede nicht starke Kopplung $e_{ij} \in E$ ist eine schwache Kopplung.

Definition 2.19. Es sei $G_A(V, E)$ der Graph der Matrix $A = (a_{ij}) \in \mathbb{C}^{n \times n}$. Zwei Knoten $v_i, v_j \in V$ sind unabhängig (bzgl. der Kantenmenge E), falls $e_{ij} \notin E$ und $e_{ji} \notin E$ ist. Eine Teilmenge $M \subset V$ heißt unabhängig, falls je zwei Knoten aus M unabhängig sind. M ist eine maximal unabhängige Menge, falls es keine unabhängige Menge M' mit $M \subset M'$ gibt.

Um robuste Verfahren zu konstruieren, ist es wünschenswert, in Richtung der starken Kopplungen zu glätten [54]. Aus diesem Grund wird die Menge Ω_f der feinen Gitterpunkte als eine maximal unabhängige Menge des reduzierten Graphen $G_A(V, E_S)$ bzgl. der Kantenmenge E_S starker Kopplungen gewählt. Offensichtlich ist eine maximal unabhängige Menge nicht eindeutig bestimmt, so daß zur Konstruktion von Ω_f zwischen unterschiedlichen Vorgehensweisen gewählt werden kann.

2.4 Restriktion und Prolongation

Da im Rahmen klassischer Mehrgitterverfahren geometrisch glatte Gitterfunktionen auf einem groben Gitter approximiert werden sollen, ist es in diesem Kontext sinnvoll, die Restriktion und die Prolongation so zu wählen, daß zumindest konstante Gitterfunktionen auf konstante Gitterfunktionen abgebildet werden,

$$R_h^H \mathbf{1}_h = \mathbf{1}_H, \quad P_H^h \mathbf{1}_H = \mathbf{1}_h. \quad (2.22)$$

Setzen wir das Standard-Grobgrid voraus, so sind aus diesem Grund für Matrizen der Gestalt (1.3) die beiden wichtigsten Restriktionen mit den zugehörigen Prolongationen in der üblichen Stern-Notation [37] die 5-Punkt-Restriktion (engl.: half weighting)

$$R_5 = \frac{1}{8} \begin{bmatrix} & 1 & \\ 1 & 4 & 1 \\ & 1 & \end{bmatrix}, \quad P_5 = \frac{1}{2} \begin{bmatrix} & 1 & \\ 1 & 2 & 1 \\ & 1 & \end{bmatrix} \quad (2.23)$$

und die 9-Punkt-Restriktion (engl.: full weighting)

$$R_9 = \frac{1}{16} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}, \quad P_9 = \frac{1}{4} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}. \quad (2.24)$$

Für matrixabhängige Prolongationen (2.18) ist die Forderung $\tilde{A}_{ff} \mathbf{1} = A_{fc} \mathbf{1}$ äquivalent zu $P_H^h \mathbf{1}_H = \mathbf{1}_h$.

Der bisher verwendete Begriff einer glatten Funktion wurde im Zusammenhang mit einer geometrisch glatten Gitterfunktion verwendet. Algebraisch kann dieser Begriff verallgemeinert werden zu:

Definition 2.20. Eine Gitterfunktion $x_h \in \Omega_h$ heißt glatt, falls $\|Ax_h\|_2 \ll \|x_h\|_2$ ist, andernfalls heißt x_h oszillierend.

Gemäß dieser Definition sind offensichtlich Eigenvektoren zu betragskleinen Eigenwerten algebraisch glatt, und Eigenvektoren zu betragsmäßig großen Eigenwerten sind algebraisch oszillierende Gitterfunktionen.

Da algebraisch glatte Eigenvektoren nicht notwendig geometrisch glatt sind, ist es naheliegend, im algebraischen Kontext eine zu (2.22) analoge Forderung aufzustellen. Für hermitesch positiv definiten Matrizen A führt dies zu der Forderung einer exakten Eigenvektor-Interpolation,

$$Pv_c = v, \quad Av = \lambda_{\min}(A)v, \quad v = (v_f^T, v_c^T)^T. \quad (2.25)$$

Sowohl durch (2.22) als auch durch (2.25) wird die Wirkung der Prolongation auf spezielle Gitterfunktionen vorgeschrieben. Man spricht in diesem Zusammenhang von sogenannten *Filter-Bedingungen*. Es sei darauf hingewiesen, daß „sinnvolle“ Filter-Bedingungen nicht notwendigerweise auch von der „exakten“ algebraischen Restriktion bzw. Prolongation,

$$R = (-A_{cf}A_{ff}^{-1} \quad I), \quad P = \begin{pmatrix} -A_{ff}^{-1}A_{fc} \\ I \end{pmatrix}, \quad (2.26)$$

erfüllt werden.

2.5 Grobgittersysteme

Bei Mehrgitterverfahren für hermitesch positiv definite Matrizen werden zu einer gegebenen Prolongation optimale Grobgittersysteme verwendet [47, 37]. Grundlage hierfür ist der folgende Satz.

Satz 2.21. Für hermitesch positiv definite Matrizen ist die Lösung der Gleichung $A_h x_h = b_h$ äquivalent zu dem Problem der Minimierung des Funktionals

$$f(x_h) = \frac{1}{2} \langle A_h x_h, x_h \rangle - \operatorname{Re} \langle b_h, x_h \rangle. \quad (2.27)$$

Beweis. [32], Abschnitt 10.2.1. □

Ersetzen wir die Näherung $x_h \in \Omega_h$ der Lösung von (2.27) durch $x_h + P_H^h x_H$ mit $x_H \in \Omega_H$, so erhalten wir:

Satz 2.22. Es sei $A_h \in \mathbb{C}^{n \times n}$ hermitesch positiv definit. Dann gilt

$$f(x_h + P_H^h x_H) = f(x_h) + \frac{1}{2} \langle A_H x_H, x_H \rangle - \operatorname{Re} \langle r_H, x_H \rangle$$

mit

$$A_H = (P_H^h)^H A_h P_H^h, \quad r_H = (P_H^h)^H (b_h - A_h x_h). \quad (2.28)$$

Beweis. [38], Abschnitt 10.1.5. □

Nach Satz 2.22 ist für hermitesch positiv definite Matrizen A_h zu einer gegebenen Prolongation P_H^h die optimale Grobgittermatrix A_H die sogenannte *Galerkin-Approximation*

$$R_h^H A_h P_H^h,$$

wobei als Restriktion R_h^H die Adjungierte $(P_H^h)^H$ der Prolongation verwendet wird.

Für Schurkomplement-Methoden hingegen werden als Grobgittermatrizen geeignete Approximationen des exakten Schurkomplements benötigt. Über Lemma 2.10 und Lemma 2.12 ist ein Zusammenhang zur Galerkin-Approximation gegeben. Speziell für M -Matrizen gibt es eine Vielzahl von Filter-Bedingungen und Probing-Techniken, um eine Schurkomplement-Approximation zu definieren [72].

2.6 Konvergenzanalyse

Die Konvergenz von Mehrgitterverfahren hängt von zahlreichen Parametern, beispielsweise die Anzahl der Glättungsschritte, die Auswahl des Glätters, der Restriktion und Prolongation sowie der Grobgittersysteme, ab. Die Konvergenztheorie klassischer Mehrgitterverfahren basiert auf einer Normabschätzung der Iterationsmatrix des Zweigitterverfahrens [37, 38].

Satz 2.23. Die Iterationsmatrix des Zweigitterverfahrens sei gegeben durch

$$H = (I - P_H^h A_H^{-1} R_h^H A_h) S^\nu.$$

Ferner gebe es eine Nullfolge $\eta(\nu)$ und eine Konstante C , so daß die Bedingungen

$$\|A_h S^\nu\| \leq \eta(\nu) \|A_h\| \tag{2.29}$$

$$\|A_h^{-1} - P_H^h A_H^{-1} R_h^H\| \leq C / \|A_h\| \tag{2.30}$$

erfüllt sind. Dann ist die Norm der Iterationsmatrix beschränkt durch

$$\|H\| \leq C\eta(\nu) \leq 1, \quad \nu > \nu_0.$$

Beweis. [38], Satz 10.6.15. □

Die Glättungseigenschaft (2.29), welche die Effizienz des Glätters beurteilt, kann meist mit Mitteln der Linearen Algebra bewiesen werden. Selbst wenn das zu lösende System aus einer Diskretisierung einer partiellen Differentialgleichung resultiert und Eigenschaften des kontinuierlichen Problems bekannt sind, ist der Beweis der Approximationseigenschaft (2.30) im allgemeinen sehr schwierig zu führen.

2.7 Zusammenfassung

Die theoretischen Resultate und numerischen Ergebnisse zeigen, daß klassische geometrische Mehrgitterverfahren zur Lösung von Gleichungssystemen der Gestalt (1.2) auf Standard-Grobgittern mit 9-Punkt-Restriktionen $R = R_9$ und zugehörigen Prolongationen $P = P_9$ sowie Galerkin-Approximationen RAP als Grobgittermatrizen schnell und unabhängig von der Diskretisierungsschrittweite h konvergieren, sofern die durch die symmetrisch positiv definite Koeffizientenmatrix A repräsentierten Kopplungen konstant sind. In diesem Zusammenhang wird das Gleichungssystem auch als *geordnetes System* bezeichnet.

Für *ungeordnete Systeme*, insbesondere für Matrizen mit zufällig erzeugten Kopplungen, sind klassische Mehrgitterverfahren ungeeignet, da die Eigenvektoren zu betragskleinen Eigenwerten im allgemeinen nicht geometrisch glatt sind. Einen Ausweg bieten in dieser Situation Schurkomplement-Ansätze zur Präkonditionierung, welche auf algebraisch motivierten Grobgittern in Verbindung mit matrixabhängigen Restriktionen und Prolongationen wirken. Im Rahmen dieser Arbeit werden solche Schurkomplement-Verfahren und ihre einzelnen Komponenten systematisch für drei verschiedene ungeordnete Systeme untersucht und effiziente Schurkomplement-Präkonditionierer hergeleitet.

Kapitel 3

Das ungeordnete Laplace-System

In diesem Kapitel werden wir mit dem ungeordneten Laplace-System ein erstes, einfaches Modellproblem vorstellen und einige wesentliche Eigenschaften dieses Modellproblems herleiten. Eine wichtige Rolle kommt dabei der odd-even Symmetrie zu, denn sie erlaubt es, als Ausgangspunkt zur Konstruktion von Multilevel-Verfahren neben dem originären System auch das odd-even reduzierte System zu verwenden. Im Anschluß an die Entwicklung verschiedener Multilevel-Verfahren für die zwei äquivalenten Systeme werden wir diese hinsichtlich ihrer Effizienz kritisch beurteilen.

3.1 Einführung

Das Gleichungssystem des Modellproblems beschreibt eine Nächste-Nachbar-Kopplung auf einem regulären zweidimensionalen Gitter mit periodischer Fortsetzung an den Rändern. Um das Gleichungssystem präzise formulieren zu können, benötigen wir einige Vorbereitungen.

Definition 3.1. Es sei Ω einem reguläres zweidimensionales Gitter,

$$\Omega = \{x = (x_1, x_2) \mid x_i \in \{1, \dots, n_i\} \text{ für alle } i\}. \quad (3.1)$$

Aufgrund der Periodizität sind Komponenten des Gitterpunktes x außerhalb des jeweiligen Bereiches modulo Indexgrenze zu verstehen. Der Einfachheit

halber setzen wir voraus, daß die Anzahl der Gitterpunkte in jeder Richtung eine hinreichend große Zweierpotenz ist,

$$n_i = 2^{N_i}, \quad N_i \geq 2, \quad i = 1, 2. \quad (3.2)$$

Definition 3.2. Eine Konfiguration $U = \{U_\mu(x)\}$ ist eine Sammlung von $2n_1n_2$ unitären Zahlen $U_\mu(x)$,

$$U_\mu(x) = \exp(-2\pi i \phi(x, \mu)), \quad \phi(x, \mu) \in [0, 1),$$

wobei x Gitterpunkt und $\mu \in \{1, 2\}$ ist.

Wir unterscheiden drei verschiedene Arten von Konfigurationen:

Definition 3.3. Eine Konfiguration heißt kalt, falls alle $U_\mu(x) = 1$ sind. Sie heißt heiß, falls alle $U_\mu(x)$ gleichverteilt sind. Eine Konfiguration, die weder kalt noch heiß ist, heißt realistisch.

Nun formulieren wir das Gleichungssystem des Modellproblems präzise:

Definition 3.4. Das lineare Gleichungssystem des Modellproblems für eine gegebene Konfiguration $\{U_\mu(x)\}$ und den reellen Parameter $\kappa \geq 0$ ist gegeben durch

$$A\psi = \varphi, \quad A = I - \kappa D, \quad (3.3)$$

wobei die Matrix $D = (D_{x,y})$ gegeben ist durch

$$D_{x,y} = U_\mu(x)\delta_{x,y-e_\mu} + U_\mu^H(x-e_\mu)\delta_{x,y+e_\mu}. \quad (3.4)$$

Abbildung 3.1 veranschaulicht die Kopplung eines Gitterpunktes x mit seinen direkten Nachbarn $y = x \pm e_\mu$ in Gitterrichtung μ sowie die zugehörigen Kopplungskoeffizienten.

Bemerkung 3.5. Für eine kalte Konfiguration bezeichnen wir das System (3.3) auch als geordnet, für eine realistische und für eine heiße Konfiguration hingegen als ungeordnet. Die Begriffe „geordnet“ und „kalt“ werden ebenso wie die Begriffe „ungeordnet“ und „heiß“ bzw. „realistisch“ synonym verwendet.

3.2 Eigenschaften

Ohne Mühe können wir nun bereits einige wichtige Eigenschaften wie beispielsweise die odd-even Symmetrie des Modellproblems festhalten und die Eigenwertverteilung der Matrix A diskutieren.

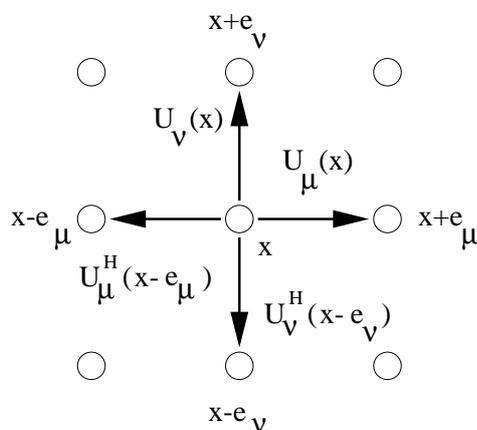


Abbildung 3.1: Die Nächste-Nachbar-Kopplung des Modellproblems als 5-Punkte-Stern mit zugehörigen Gewichten.

Das Besetztheitsmuster

Lemma 3.6. Die Matrix D besitzt genau vier Nichtnulleinträge pro Zeile, nämlich an den Positionen $(\mu = 1, 2)$

$$\begin{aligned} (x, x + e_\mu) &: U_\mu(x) \\ (x, x - e_\mu) &: U_\mu^H(x - e_\mu). \end{aligned}$$

Beweis. Klar. □

Die odd-even Symmetrie

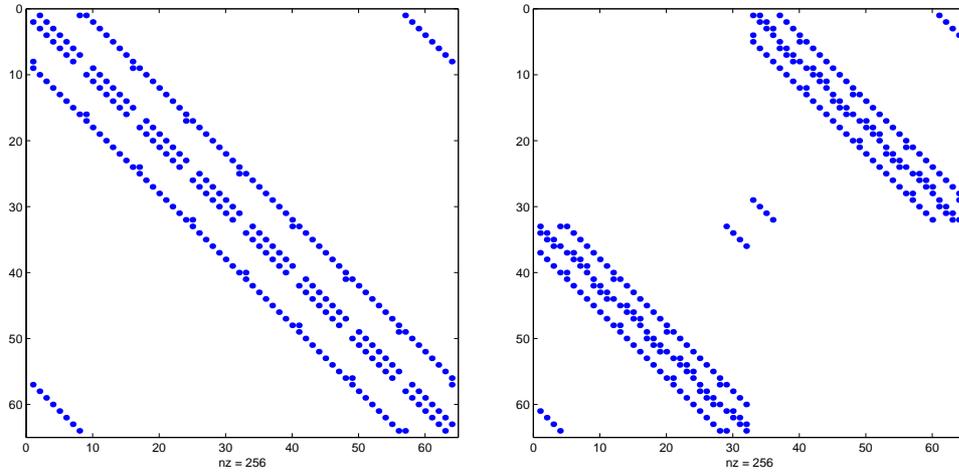
Da die Anzahl der Gitterpunkte in jeder Richtung nach der eingangs festgehaltenen Annahme eine hinreichend große Zweierpotenz ist, koppeln gerade Gitterpunkte x_e nur mit ungeraden Gitterpunkten x_o und umgekehrt. Diese Eigenschaft bezeichnen wir als *odd-even Symmetrie*.

Aufgrund der odd-even Symmetrie besitzt die Matrix D bei einer odd-even Anordnung der Gitterpunkte die Gestalt

$$D = \begin{pmatrix} 0 & D_{oe} \\ D_{eo} & 0 \end{pmatrix}.$$

Dabei kennzeichnet der Index o ungerade, der Index e hingegen gerade Gitterpunkte.

Die Abbildung 3.2 veranschaulicht das Besetztheitsmuster der Matrix D bei einer lexikographischen und einer odd-even Anordnung der Gitterpunkte.



(a) lexikographisch

(b) odd-even

Abbildung 3.2: Das Besetztheitsmuster der Matrix D bei lexikographischer und odd-even Anordnung der Gitterpunkte.

Das Spektrum

Lemma 3.7. Die Matrix D ist hermitesch.

Beweis. Die Matrix D besitzt vier Nichtnulleinträge pro Spalte ($\mu = 1, 2$),

$$\begin{aligned} (x - e_\mu, x) &: U_\mu(x - e_\mu) \\ (x + e_\mu, x) &: U_\mu^H(x). \end{aligned}$$

Daher besitzt die Matrix D^H vier Nichtnulleinträge pro Zeile ($\mu = 1, 2$),

$$\begin{aligned} (x, x - e_\mu) &: U_\mu^H(x - e_\mu) \\ (x, x + e_\mu) &: U_\mu(x), \end{aligned}$$

und ein Vergleich mit den Einträgen der Matrix D liefert die Behauptung. \square

Als unmittelbare Folgerung bemerken wir, daß das Spektrum der Matrix D reell ist. Da die Matrix D nur reelle Eigenwerte besitzt, sind somit auch alle Eigenwerte von A reell.

Definition 3.8. Der kritische Wert κ_c ist definiert als $\kappa_c = 1/\lambda_{\max}(D)$.

Wird der Parameter κ geeignet gewählt, ist die Matrix A hermitesch positiv definit.

Lemma 3.9. Die Matrix A ist hermitesch positiv definit für alle $0 \leq \kappa < \kappa_c$.

Beweis. Für $\kappa < 1/\lambda_{\max}(D)$ gilt $\lambda_{\min}(A) = 1 - \kappa\lambda_{\max}(D) > 0$. \square

Lemma 3.10. Es gilt $\rho(D) \leq 4$.

Beweis. $\rho(D) \leq \|D\|_1 = \|D\|_\infty = 4$. \square

Im Falle einer kalten Konfiguration ist der Spektralradius $\rho(D)$ explizit bekannt.

Lemma 3.11. Für eine kalte Konfiguration gilt $\rho(D) = 4$.

Beweis. Aus Lemma 3.10 und $D\mathbf{1} = 4\mathbf{1}$ folgt $\rho(D) = 4$. \square

Es ist $\kappa_c = 1/4$ der kritische Wert einer kalten Konfiguration. Die kritischen Werte κ_c einer heißen Konfiguration lassen sich lediglich approximieren. Für jeweils 100 Messungen sind der Mittelwert $M(\cdot)$, die Standardabweichung $\sigma(\cdot)$ sowie der minimale und maximale Wert von κ_c in Abhängigkeit von der Gittergröße $n = N^2$ in Tabelle 3.1 zusammengestellt.

N	$M(\kappa_c)$	$\sigma(\kappa_c)$	$\min(\kappa_c)$	$\max(\kappa_c)$
4	0.31346	0.014174	0.27418	0.35198
8	0.2978	0.005564	0.28722	0.3137
16	0.29211	0.0034659	0.2836	0.29938
32	0.28859	0.00247	0.28248	0.29358
64	0.28594	0.0019428	0.28085	0.28914
128	0.28375	0.001977	0.2786	0.28756

Tabelle 3.1: Der kritische Wert κ_c im Falle einer heißen Konfiguration.

Der Tabelle 3.1 ist zu entnehmen, daß für hinreichend große Gitterdimensionen der kritische Wert $\kappa_c \approx 0.28$ ist. Daher ist die folgende Annahme gerechtfertigt:

Annahme 3.12. Es gilt stets $\kappa < \kappa_s$ mit $\kappa_s = 0.3$.

Der Wert κ_s ermöglicht es insbesondere, beispielsweise bei Abschätzungen der Kondition konkrete obere Schranken anzugeben. Aus diesem Grund werden wir im folgenden wiederholt auf κ_s zurückgreifen.

Ausgehend von einer odd-even Anordnung der Gitterpunkte können weitere Aussagen über das Spektrum der Matrix D getroffen werden.

Lemma 3.13. Das Spektrum $\sigma(D)$ der Matrix D ist symmetrisch um den Ursprung, insbesondere ist die Matrix D maximal indefinit.

Beweis. Wegen

$$\begin{pmatrix} 0 & D_{oe} \\ D_{eo} & 0 \end{pmatrix} \begin{pmatrix} v_o \\ v_e \end{pmatrix} = \lambda \begin{pmatrix} v_o \\ v_e \end{pmatrix} \implies \begin{pmatrix} 0 & D_{oe} \\ D_{eo} & 0 \end{pmatrix} \begin{pmatrix} -v_o \\ v_e \end{pmatrix} = -\lambda \begin{pmatrix} -v_o \\ v_e \end{pmatrix}$$

ist mit λ auch $-\lambda$ ein Eigenwert von D . \square

Ist der Eigenvektor v zum Eigenwert λ der Matrix D eine konstante Gitterfunktion, so ist nach Lemma 3.13 der Eigenvektor zum Eigenwert $-\lambda$ oszillierend.

Lemma 3.14. Es sei $\kappa_c = 1/\lambda_{\max}(D)$. Dann gilt

$$\kappa(A) = \left(1 + \frac{\kappa}{\kappa_c}\right) / \left(1 - \frac{\kappa}{\kappa_c}\right).$$

Beweis. Wegen Lemma 3.13 gilt

$$\lambda_{\max}(A) = 1 + \frac{\kappa}{\kappa_c}, \quad \lambda_{\min}(A) = 1 - \frac{\kappa}{\kappa_c}.$$

Nun folgt die Behauptung unmittelbar. \square

Abbildung 3.3 zeigt das Spektrum einer kalten und einer heißen Konfiguration.

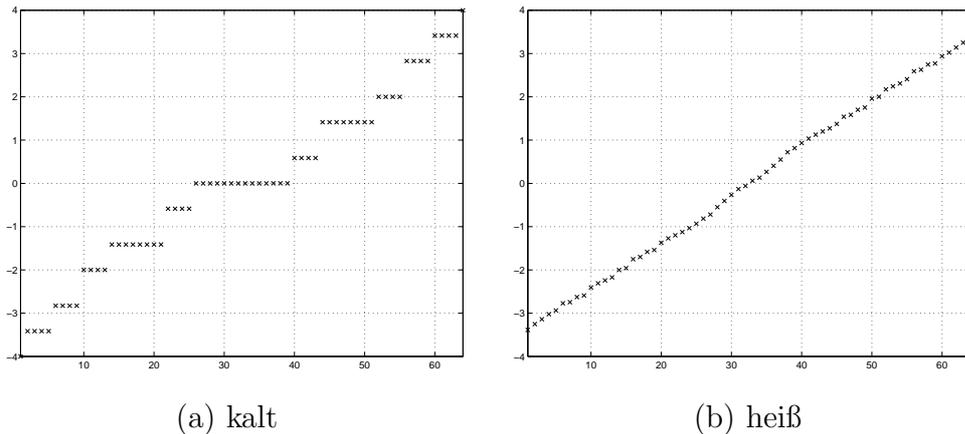


Abbildung 3.3: Das Spektrum $\sigma(D)$ der Matrix D im Falle einer kalten und einer heißen Konfiguration ($N = 8$). Die Eigenwerte sind der Größe nach sortiert aufgetragen.

3.3 Das odd-even reduzierte System

Aufgrund der odd-even Symmetrie des Problems ist es naheliegend, zur Prakonditionierung zunachst eine sogenannte odd-even Reduktion des Gleichungssystems vorzunehmen.

Bei einer odd-even Anordnung der Gitterpunkte ist die Schurkomplement-Zerlegung der Matrix

$$A = \begin{pmatrix} I & -\kappa D_{oe} \\ -\kappa D_{eo} & I \end{pmatrix}$$

nach (2.13) gegeben durch

$$A = \begin{pmatrix} I & 0 \\ -\kappa D_{eo} & I \end{pmatrix} \begin{pmatrix} I & 0 \\ 0 & I - \kappa^2 D_{eo} D_{oe} \end{pmatrix} \begin{pmatrix} I & -\kappa D_{oe} \\ 0 & I \end{pmatrix}.$$

Diese Beobachtung fuhrt zu der folgenden Definition.

Definition 3.15. Das zum Modellproblem (3.3) gehorige sogenannte *odd-even-reduzierte Gleichungssystem* ist gegeben durch

$$A_e \psi_e = \tilde{\varphi}_e, \quad A_e = I - \kappa^2 D_{eo} D_{oe}, \quad (3.5)$$

mit $\tilde{\varphi}_e = \varphi_e + \kappa D_{eo} \varphi_o$ und $\psi_o = \varphi_o + \kappa D_{oe} \psi_e$.

Der Ubergang vom ursprunglichen System hin zum odd-even reduzierten System entspricht im Multilevel-Kontext genau einem Zyklus bestehend aus einer exakten, matrixabhangigen Restriktion auf das aus allen geraden Gitterpunkten bestehende Grobgitter, dem einmaligen Losen eines Grobgittersystems, und der anschließenden Prolongation der Grobgitterlosung auf das originare Gitter.

Der Aufwand zur Berechnung einer Matrix-Vektor-Multiplikation mit dem Schurkomplement A_e entspricht dem Aufwand einer Matrix-Vektor-Multiplikation mit A , da in beiden Fallen im wesentlichen je eine Matrix-Vektor-Multiplikation mit der Matrix D_{eo} und der Matrix D_{oe} durchzufuhren ist.

Satz 3.16. λ ist ein Eigenwert von A genau dann, wenn $\lambda(2 - \lambda)$ Eigenwert von A_e ist.

Beweis. Da $\det(\lambda I - A) = \det((\lambda - 1)^2 I - \kappa^2 D_{eo} D_{oe}) = \det((\lambda^2 - 2\lambda)I + A_e)$ ist, gilt auch $\det(\lambda I - A) = 0 \iff \det(\lambda(2 - \lambda)I - A_e) = 0$. \square

Das odd-even reduzierte System bietet gegenüber dem ursprünglichen System drei wesentliche Vorteile:

- a) Die Kondition verbessert sich nach Satz 3.16, denn kleine Eigenwerte von A werden nahezu verdoppelt, große Eigenwerte hingegen annähernd halbiert, vgl. Abb. 3.4. In der Praxis führt dies zu einer Halbierung der Iterationszahlen für alle Krylov-Unterraum-Verfahren.
- b) Es kann Speicherplatz eingespart werden, da lediglich halbierte Vektorlängen gespeichert werden müssen.
- c) Es kann eine Einteilung in starke und schwache Kopplungen (s. Def. 2.18) vorgenommen werden.

Der zuletzt genannte Punkt ist im weiteren Verlauf wesentlich und bildet die Grundlage für die in Abschnitt 3.4.3 untersuchten Schurkomplement-Verfahren zur Lösung des odd-even reduzierten Systems.

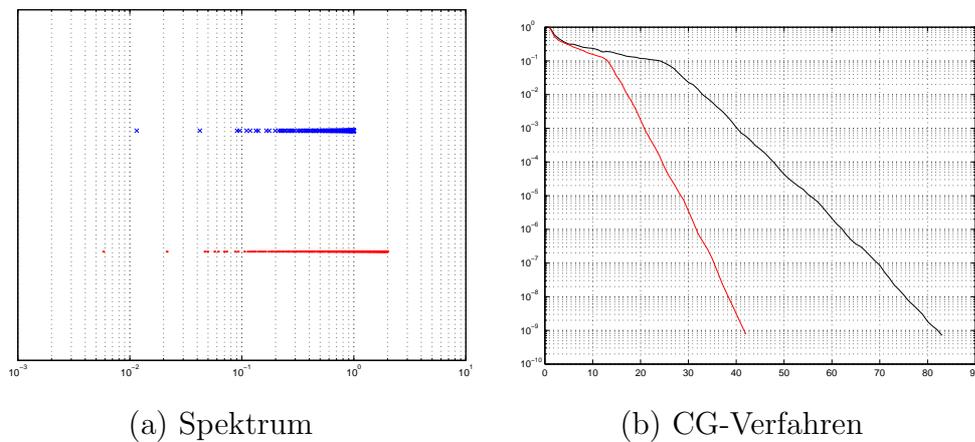


Abbildung 3.4: Die Eigenwerte der originären Matrix (\cdot) und der zugehörigen odd-even reduzierten Matrix (\times) sowie das typische Konvergenzverhalten des CG-Verfahrens für das originäre System und das äquivalente odd-even reduzierte System bei einer heißen Konfiguration.

Betrachten wir die Matrix A_e genauer, so stellen wir fest, daß die Matrix A_e einen 9-Punkte-Stern beschreibt, vgl. Abb. 3.5. Die beobachtete Auffüllung ist der Grund dafür, daß wir zur Lösung des Modellproblems nicht sukzessive eine odd-even Reduktion anwenden können.

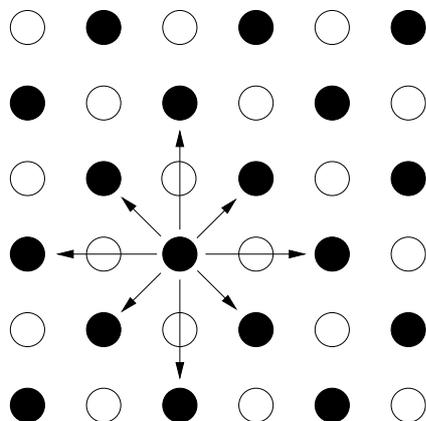


Abbildung 3.5: Die Kopplungen der odd-even reduzierten Matrix.

3.4 Multilevel-Verfahren

Bei der folgenden Konstruktion von Multilevel-Verfahren unterscheiden wir zwischen dem Fall eines geordneten und dem eines ungeordneten Systems. Wir werden zunächst den einfacheren Fall geordneter Systeme betrachten und uns anschließend dem schwierigeren Fall ungeordneter Systeme zuwenden. Es wird sich zeigen, daß die im einfachen Fall gewonnenen theoretischen Resultate zum Beweis entsprechender Aussagen für ungeordnete Systeme herangezogen werden können.

F/C-Anordnung der Gitterpunkte

Wir setzen im folgenden das Standard-Grobgitter voraus. Mit der in Def. 2.13 eingeführten Notation sind die groben Gitterpunkte Ω_c und die verbleibenden feinen Gitterpunkte Ω_f somit

$$\Omega_c = L_2(\Omega_h), \quad \Omega_f = \Omega_h \setminus \Omega_c = L_1(\Omega_h) \cup L_0(\Omega_h). \quad (3.6)$$

Repräsentiert die Matrix D_{L_i, L_j} die Kopplungen der Gitterpunkte $x \in L_i(\Omega_h)$ mit Gitterpunkten $y \in L_j(\Omega_h)$, so besitzt die Matrix D bei einer F/C-Anordnung der Gitterpunkte die Gestalt

$$D = \begin{pmatrix} D_{ff} & D_{fc} \\ D_{cf} & D_{cc} \end{pmatrix} = \begin{pmatrix} D_{ff} & D_{fc} \\ D_{cf} & 0 \end{pmatrix} \quad (3.7)$$

mit

$$D_{ff} = \begin{pmatrix} 0 & D_{L_1, L_0} \\ D_{L_0, L_1} & 0 \end{pmatrix} \quad (3.8)$$

und

$$D_{fc} = \begin{pmatrix} D_{L_1, L_2} \\ 0 \end{pmatrix}, \quad D_{cf} = (D_{L_2, L_1} \quad 0). \quad (3.9)$$

Dabei setzen wir voraus, daß alle Gitterpunkte aus $L_1(\Omega_h)$ vor Gitterpunkten aus $L_0(\Omega_h)$ gezählt werden. Sofern das Gitter Ω dem Kontext zu entnehmen ist, schreiben wir oft auch L_i statt $L_i(\Omega)$.

3.4.1 Mehrgitterverfahren

Im Falle einer kalten Konfiguration, $U_\mu(x) = 1$, sind die Koeffizienten der Matrix D offensichtlich konstant. Das System (3.3) ist geordnet, und die Matrix A ist nach Lemma 3.9 hermitesch positiv definit für alle $0 \leq \kappa < \kappa_c$. Daher sind klassische Mehrgitterverfahren (s. Kapitel 2) erfolgreich anwendbar.

Restriktion und Prolongation

Nach Lemma 3.11 ist die konstante Gitterfunktion $\mathbf{1}$ ein Eigenvektor zum kleinsten Eigenwert der Matrix A . Daher bietet sich die 9-Punkt-Restriktion (2.24) und die zugehörige Prolongation an. Da diese Restriktion und Prolongation konstante Gitterfunktionen auf konstante Gitterfunktionen abbilden, entspricht der Ansatz der in Kapitel 2 dargestellten exakten Eigenvektor-Interpolation, vgl. (2.25).

Ohne Mühe verifizieren wir, daß die Restriktion und Prolongation auch geschrieben werden können als

$$R_9 = \frac{1}{16} (2D_{L_2, L_1} \quad \frac{1}{2}D_{L_2, L_1}D_{L_1, L_0} \quad 4I), \quad P_9 = 4R_9^H. \quad (3.10)$$

Analog erhalten wir für die 5-Punkt-Formeln (2.23) die Darstellung

$$R_5 = \frac{1}{8} (D_{L_2, L_1} \quad 0 \quad 4I), \quad P_5 = \frac{1}{2} \begin{pmatrix} D_{L_1, L_2} \\ 0 \\ 2I \end{pmatrix}. \quad (3.11)$$

Die 5-Punkt-Formeln sind aber offensichtlich lediglich eingeschränkt geeignet, da hier der Austausch von Information zwischen Gitterpunkten aus L_0 und L_2 nur durch den Glättungsschritt erfolgen kann, nicht jedoch durch die Restriktion oder Prolongation.

Grobgittermatrix

Gemäß Satz 2.22 ist für hermitesch positiv definite Matrizen A die Galerkin-Approximation $A_c = RAP$ mit $R = P^H$ als Grobgittermatrix optimal. Eine analoge Aussage gilt für $R = cP^H$ mit $c > 0$.

Für $R = R_9$ und $P = P_9$ erhalten wir als Galerkin-Approximation:

Lemma 3.17. Es sei $R = R_9$, $P = P_9$ und $A_c = RAP$. Dann gilt

$$A_c = \frac{1}{4} \left(I + \frac{1-4\kappa}{4} D_{L_2, L_1} \left(I + \frac{1-8\kappa}{16(1-4\kappa)} D_{L_1, L_0} D_{L_0, L_1} \right) D_{L_1, L_2} \right). \quad (3.12)$$

Beweis. Ausrechnen. \square

Im folgenden werden wir genauer auf die Matrix $A_c = RAP$ eingehen. Dazu untersuchen wir aber zunächst die beteiligten Teilmatrizen:

Lemma 3.18. Es gilt $D_{L_0, L_1} \mathbf{1} = 4\mathbf{1}$, $D_{L_1, L_0} \mathbf{1} = 2\mathbf{1}$, $D_{L_1, L_2} \mathbf{1} = 2\mathbf{1}$ und $D_{L_2, L_1} \mathbf{1} = 4\mathbf{1}$.

Beweis. Trivial. \square

Lemma 3.19. Für $i = 0, 2$ gilt in der Stern-Notation

$$D_{L_i, L_1} D_{L_1, L_i} = \begin{bmatrix} & 1 & \\ 1 & 4 & 1 \\ & 1 & \end{bmatrix}.$$

Beweis. Um die Diagonaleinträge der Matrix $D_{L_i, L_1} D_{L_1, L_i}$ zu bestimmen, wählen wir einen Gitterpunkt $x \in L_i$ aus und berechnen

$$\begin{aligned} (D_{L_i, L_1} D_{L_1, L_i})_{x, x} &= (D_{L_i, L_1})_{x, \bullet} (D_{L_1, L_i})_{\bullet, x} \\ &= \sum_{\mu=1}^2 (D_{L_i, L_1})_{x, x+e_\mu} (D_{L_1, L_i})_{x+e_\mu, x} + (D_{L_i, L_1})_{x, x-e_\mu} (D_{L_1, L_i})_{x-e_\mu, x} = 4. \end{aligned}$$

Als Kopplung des Gitterpunktes x zum Nachfolger $y = x + 2e_\mu \in L_i$ in Gitterrichtung μ erhalten wir

$$\begin{aligned} (D_{L_i, L_1} D_{L_1, L_i})_{x, y} &= (D_{L_i, L_1})_{x, \bullet} (D_{L_1, L_i})_{\bullet, x+2e_\mu} \\ &= (D_{L_i, L_1})_{x, x+e_\mu} (D_{L_1, L_i})_{x+e_\mu, x+2e_\mu} = 1, \end{aligned}$$

und entsprechend zum Vorgänger $y = x - 2e_\mu \in L_i$ in Gitterrichtung μ

$$\begin{aligned} (D_{L_i, L_1} D_{L_1, L_i})_{x, y} &= (D_{L_i, L_1})_{x, \bullet} (D_{L_1, L_i})_{\bullet, x-2e_\mu} \\ &= (D_{L_i, L_1})_{x, x-e_\mu} (D_{L_1, L_i})_{x-e_\mu, x-2e_\mu} = 1. \end{aligned}$$

Weitere Kopplungen zu Gitterpunkten $y \in L_i$ existieren nicht, da diese vom Gitterpunkt x zu weit entfernt liegen. \square

Der Beweis des Lemmas 3.19 offenbart, daß die entsprechenden Matrixeinträge bereits bei einfachen Kopplungen ziemlich aufwendig zu bestimmen sind. Im folgenden werden wir daher der Einfachheit halber wiederholt anschaulich argumentieren: Dazu zeichnen wir zunächst die durch die Matrix D_{L_i, L_j} repräsentierte Kopplung von einem Gitterpunkt $x \in L_i$ zu einem benachbarten Gitterpunkt $y \in L_j$ als Pfeil von y zu x auf der Verbindungslinie der Gitterpunkte ein und zählen anschließend die vorhandenen Wege erlaubter Länge.

Für $i = 0, 2$ sind in Abb. 3.6(a) die durch die Matrix $D_{L_i, L_1} D_{L_1, L_i}$ repräsentierte Kopplungen eines Gitterpunktes $x \in L_i$ mit sich selbst und in Abb. 3.6(b) die Kopplungen mit seinen benachbarten Gitterpunkten $y \in L_i$ mit $y \neq x$ zu sehen. Dabei sind alle Gitterpunkte aus L_i schwarz und alle Gitterpunkte aus L_j grau markiert. Die verbleibenden Gitterpunkte sind aus L_1 .

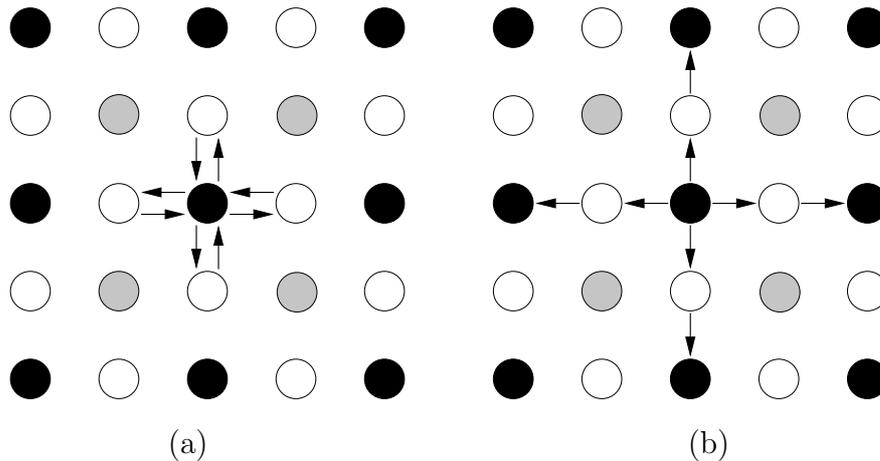


Abbildung 3.6: Die Matrix $D_{L_i, L_1} D_{L_1, L_i}$ für $i = 0, 2$.

Das Resultat des Lemmas 3.19 ist nun auch folgendermaßen leicht nachzuvollziehen:

- Es existieren genau vier verschiedene Wege der Länge 2 von einem Gitterpunkt $x \in L_i$ zu sich selbst, $(D_{L_i, L_1} D_{L_1, L_i})_{x,x} = 4$.
- Es existiert lediglich ein Weg der Länge 2 von einem Gitterpunkt $x \in L_i$ zu einem benachbarten Gitterpunkt $y \in L_i$, $(D_{L_i, L_1} D_{L_1, L_i})_{x,y} = 1$.

Auch der Beweis des folgenden Lemmas ist mit dieser Argumentationsweise einfach:

Lemma 3.20. Es gilt in der Stern-Notation

$$D_{L_i, L_1} D_{L_1, L_j} D_{L_j, L_1} D_{L_1, L_i} = 4 \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$$

für $i, j = 0, 2$ mit $i \neq j$.

Beweis. Wir betrachten die Abbildung 3.7. Wie zuvor sind alle Gitterpunkte aus L_i schwarz und alle Gitterpunkte aus L_j grau markiert. Die verbleibenden Gitterpunkte sind aus L_1 . Alle Wege besitzen die Länge 4. Es sei $x \in L_i$.

- a) Für $y = x + e_\mu + e_\nu \in L_j$ ($\mu \neq \nu$) zeigt Abb. 3.7(a) und Abb. 3.7(b) alle vier verschiedenen Wege von x über y zurück zu x . Insgesamt gibt es also 16 Wege von x über Gitterpunkte aus L_j zurück zu x .
- b) Für $y = x + 2e_\mu \in L_i$ sind alle 8 Wege von x zu y in Abb. 3.7(c) und Abb. 3.7(d) zu sehen. Ein entsprechendes Resultat erhalten wir für $y = x - 2e_\mu$.
- c) Des weiteren existieren 4 verschiedene Wege von x zum Gitterpunkt $y = x + 2e_\mu + 2e_\nu \in L_i$ mit $\mu \neq \nu$, siehe Abb. 3.7(e) und Abb. 3.7(f). Ein entsprechendes Resultat erhalten wir für die Gitterpunkte $y = x + 2e_\mu - 2e_\nu$, $y = x - 2e_\mu + 2e_\nu$ und $y = x - 2e_\mu - 2e_\nu$.

Insgesamt erhalten wir die Behauptung. □

Erinnern wir uns an die Grobgittermatrix (3.12),

$$A_c = \frac{1}{4} \left(I + \frac{1-4\kappa}{4} D_{L_2, L_1} \left(I + \frac{1-8\kappa}{16(1-4\kappa)} D_{L_1, L_0} D_{L_0, L_1} \right) D_{L_1, L_2} \right),$$

so kommt es offensichtlich im Vergleich zur originären Matrix zur Auffüllung, denn die Matrix A_c beschreibt einen 9-Punkte-Stern, vgl. Lemma 3.19 und Lemma 3.20.

Um praktikable Mehrgitterverfahren zu definieren, ist es jedoch wichtig, daß die Anzahl der Nichtnullelemente der Grobgittermatrizen beschränkt und damit der Aufwand auf der entsprechenden Stufe akzeptabel bleibt. Die folgende Überlegung zeigt, daß in dem hier diskutierten Fall lediglich im ersten Schritt, also beim Übergang vom originären 5-Punkte-Stern A zur Grobgittermatrix RAP Auffüllung eintritt, nicht jedoch in den darauf folgenden Schritten:

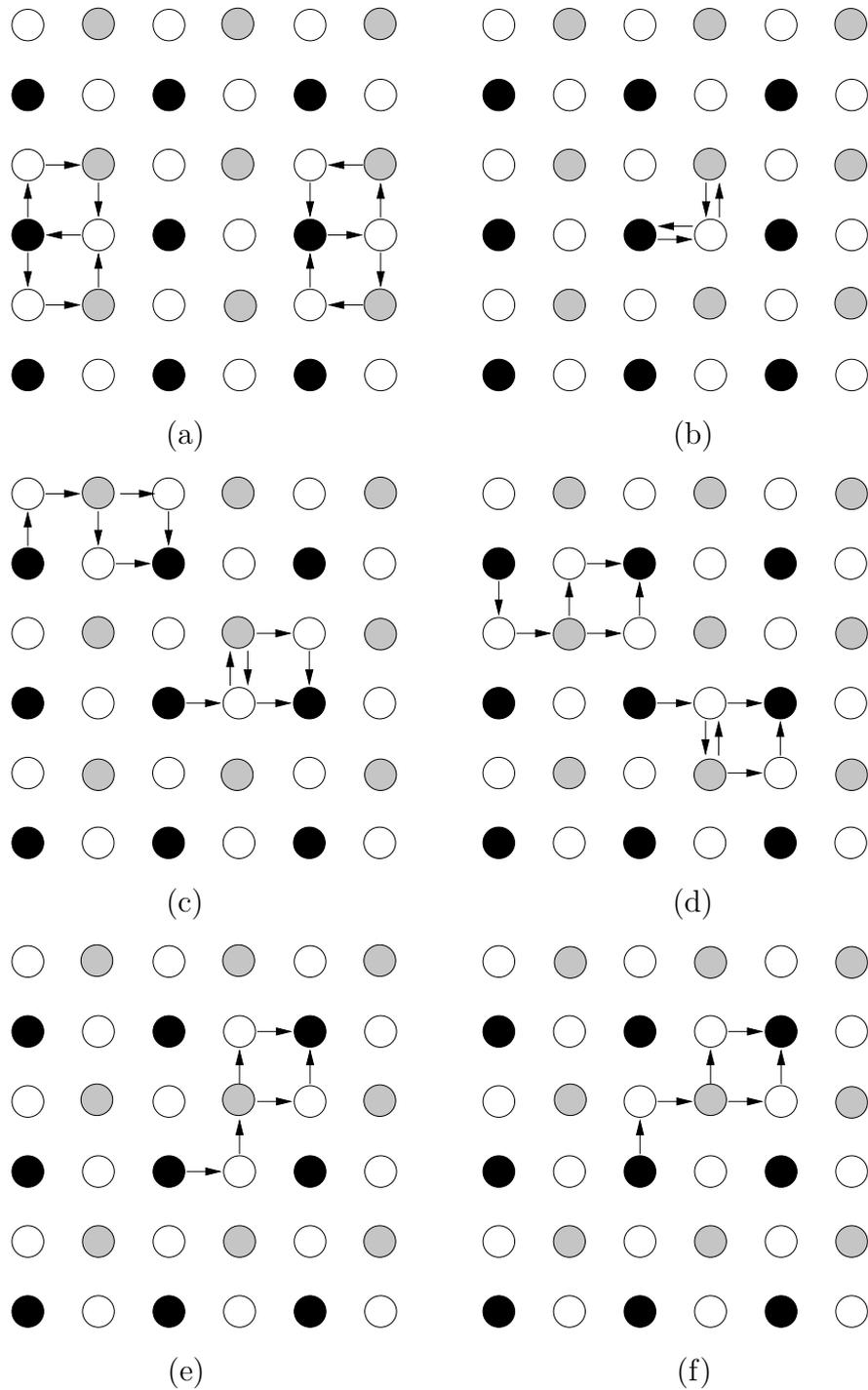


Abbildung 3.7: Die Matrix $D_{L_i, L_1} D_{L_1, L_j} D_{L_j, L_1} D_{L_1, L_i}$ für $i, j = 0, 2$ mit $i \neq j$.

Sind die Restriktion R , die Prolongation P und auch die Matrix A jeweils 9-Punkte-Sterne, so auch die Matrix RAP , denn die Matrizen R , A und P erlauben es jeweils höchstens, einen Weg der Länge 1 in einer Gitterrichtung μ zu gehen, vgl. Abb. 3.8. Dies bedeutet aber gerade, daß die Matrix RAP wiederum einen 9-Punkte-Stern beschreibt.

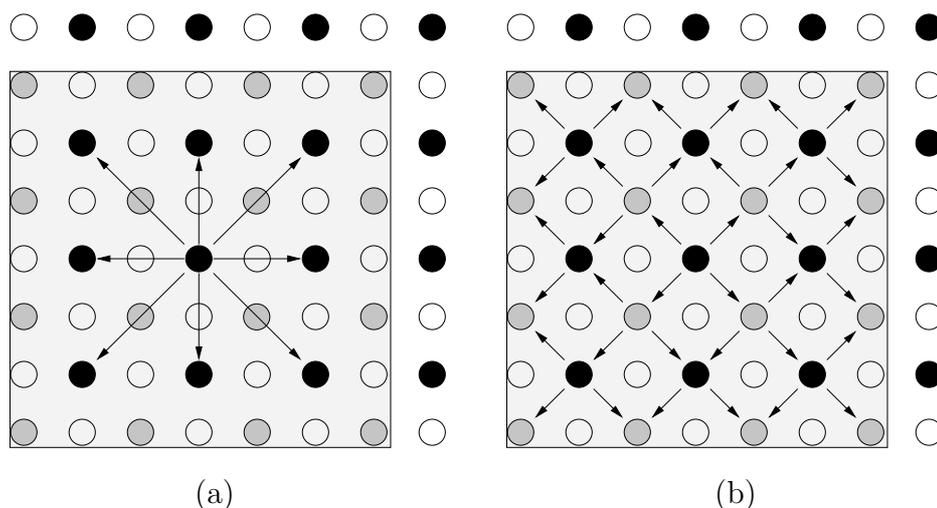


Abbildung 3.8: Beschreiben Restriktion R , Prolongation P und Matrix A jeweils 9-Punkte-Sterne, so ist die Galerkin-Approximation RAP wiederum ein 9-Punkte-Stern (a), da das Verlassen des grau eingefärbten Bereiches unmöglich ist (b).

Durch die Wahl der Restriktion $R = R_9$ und Prolongation $P = P_9$, vgl. (2.24) und auch (3.10), wird die konstante Gitterfunktion, welche Eigenvektor zum kleinsten Eigenwert der Matrix A ist, exakt auf das Grobgitter abgebildet.

Auf dem Grobgitter gilt:

Lemma 3.21. Es ist $A_c \mathbf{1} = \lambda_{\min}(A) \mathbf{1}$.

Beweis. Mit Hilfe der Lemmata 3.19 und 3.20 erhalten wir

$$D_{L_2, L_1} D_{L_1, L_2} \mathbf{1} = 8 \mathbf{1}, \quad D_{L_2, L_1} D_{L_1, L_0} D_{L_0, L_1} D_{L_1, L_2} \mathbf{1} = 64 \mathbf{1}.$$

Es folgt für die Galerkin-Approximation

$$4A_c \mathbf{1} = \left(1 + 8 \frac{1 - 4\kappa}{4} + 64 \frac{1 - 4\kappa}{4} \frac{1 - 8\kappa}{16(1 - 4\kappa)} \right) \mathbf{1} = 4(1 - 4\kappa) \mathbf{1},$$

also $A_c \mathbf{1} = (1 - 4\kappa) \mathbf{1}$. □

Nach Lemma 3.21 ist die konstante Gitterfunktion auch Eigenvektor der Galerkin-Approximation, so daß durch eine rekursive Vorgehensweise, d.h. durch ein echtes Mehrgitterverfahren mit entsprechend gewählter Restriktion $R = R_9$ und Prolongation $P = P_9$ sowie einer Galerkin-Approximation RAP als jeweilige Grobgittermatrix der Fehleranteil des glatten Eigenvektors vollständig eliminiert werden kann. Dies ist der Grund dafür, daß der hier dargestellte Mehrgitter-Ansatz für geordnete Systeme besonders effizient ist.

Für ungeordnete Systeme sind, wie bereits in Kapitel 2 angemerkt, klassische Mehrgitterverfahren ungeeignet. Daher wenden wir uns ab jetzt den Schurkomplement-Verfahren zu.

3.4.2 Schurkomplement-Verfahren

Anstelle eines klassischen Mehrgitter-Verfahrens untersuchen wir nun Schurkomplement-Verfahren zur Lösung des originären Systems bei einer kalten Konfiguration. Anschließend werden ausführlich Schurkomplement-Verfahren für realistische und heiße Konfigurationen behandelt. Zu Beginn der Diskussion sehen wir uns das exakte Schurkomplement an:

Das exakte Schurkomplement

Die Länge eines Weges von einem groben Gitterpunkt $x \in \Omega_c$ zu einem anderen Grobgitterpunkt $y \in \Omega_c$ ist bzgl. des Gitters Ω_h stets gerade. Diese erste Überlegung rechtfertigt die explizite Formulierung des folgenden Lemmas.

Lemma 3.22. Für alle $\nu \in \mathbb{N}$ ist

$$D_{cf} D_{ff}^{2\nu+1} D_{fc} = 0$$

und

$$D_{ff}^{2\nu} = \begin{pmatrix} (D_{L_1, L_0} D_{L_0, L_1})^\nu & 0 \\ 0 & (D_{L_0, L_1} D_{L_1, L_0})^\nu \end{pmatrix}.$$

Beweis. Trivial. □

Lemma 3.23. Das exakte Schurkomplement S_A der Matrix A lautet

$$S_A = I - \kappa^2 D_{L_2, L_1} (I - \kappa^2 D_{L_1, L_0} D_{L_0, L_1})^{-1} D_{L_1, L_2}. \quad (3.13)$$

Beweis. Wir rechnen unter Verwendung der Neumann-Reihe

$$S_A = A_{cc} - A_{cf} A_{ff}^{-1} A_{fc} = I - \kappa^2 D_{cf} \sum_{\nu=0}^{\infty} (\kappa D_{ff})^\nu D_{fc}.$$

Mit Hilfe des ersten Teils von Lemma 3.22 folgt

$$S_A = I - \kappa^2 D_{cf} \sum_{\nu=0}^{\infty} (\kappa D_{ff})^{2\nu} D_{fc}.$$

Betrachten wir die beteiligten Matrizen D_{cf} , $D_{ff}^{2\nu}$ und D_{fc} , vgl. (3.9) und Lemma 3.22, so gilt

$$\begin{aligned} S_A &= I - \kappa^2 D_{L_2, L_1} \sum_{\nu=0}^{\infty} (\kappa^2 D_{L_1, L_0} D_{L_0, L_1})^\nu D_{L_1, L_2} \\ &= I - \kappa^2 D_{L_2, L_1} (I - \kappa^2 D_{L_1, L_0} D_{L_0, L_1})^{-1} D_{L_1, L_2}. \end{aligned}$$

Damit ist der Beweis erbracht. \square

Nützlich ist die folgende Identität:

Lemma 3.24.

$$S_A = I - \kappa^2 D_{L_2, L_1} \left(I + \kappa^2 (I - \kappa^2 D_{L_1, L_0} D_{L_0, L_1})^{-1} D_{L_1, L_0} D_{L_0, L_1} \right) D_{L_1, L_2}.$$

Beweis. Wegen

$$\sum_{\nu=0}^{\infty} (\kappa^2 D_{L_1, L_0} D_{L_0, L_1})^\nu = I + \kappa^2 \sum_{\nu=0}^{\infty} (\kappa^2 D_{L_1, L_0} D_{L_0, L_1})^\nu D_{L_1, L_0} D_{L_0, L_1}$$

ist dies klar. \square

Die kalte Konfiguration

Wir diskutieren zunächst Schurkomplement-Verfahren für eine kalte Konfiguration. Ein Gegenüberstellung von Mehrgitter- und Schurkomplement-Verfahren zur Lösung des originären Systems zeigt, daß bei geeigneter Wahl der Parameter die Verfahren vergleichbar effizient sind.

Das exakte Schurkomplement

Die konstante Gitterfunktion $\mathbf{1}$ nicht nur ein Eigenvektor zum kleinsten Eigenwert von A , sondern auch ein Eigenvektor des exakten Schurkomplements:

Lemma 3.25. Für eine kalte Konfiguration gilt

$$S_A \mathbf{1} = \left(1 - \frac{8\kappa^2}{1 - 8\kappa^2} \right) \mathbf{1}.$$

Beweis. Wir betrachten zunächst

$$\begin{aligned} A_{ff}^{-1} D_{fc} \mathbf{1} &= 2A_{ff}^{-1} \begin{pmatrix} \mathbf{1} \\ 0 \end{pmatrix} = 2 \sum_{\nu=0}^{\infty} (\kappa D_{ff})^{2\nu} (I + \kappa D_{ff}) \begin{pmatrix} \mathbf{1} \\ 0 \end{pmatrix} \\ &= 2 \sum_{\nu=0}^{\infty} (\kappa D_{ff})^{2\nu} \begin{pmatrix} \mathbf{1} \\ \star \end{pmatrix} = 2 \sum_{\nu=0}^{\infty} (\kappa D_{ff})^{2\nu} \begin{pmatrix} \mathbf{1} \\ \star \end{pmatrix}. \end{aligned}$$

Nun gilt

$$\sum_{\nu=0}^{\infty} (\kappa^2 D_{L_1, L_0} D_{L_0, L_1})^{\nu} \mathbf{1} = \sum_{\nu=0}^{\infty} (8\kappa^2)^{\nu} \mathbf{1} = \frac{1}{1 - 8\kappa^2} \mathbf{1}.$$

Daher folgt mit

$$S_A \mathbf{1} = \mathbf{1} - \kappa^2 D_{cf} A_{ff}^{-1} D_{fc} \mathbf{1} = \left(1 - \frac{8\kappa^2}{1 - 8\kappa^2}\right) \mathbf{1}$$

die Behauptung. \square

Restriktion und Prolongation

Für eine kalte Konfiguration ist $I + \kappa D_{ff}$ bis auf Skalierung die exakte Inverse für den Eigenvektor $\mathbf{1}$ zum kleinsten Eigenwert der Matrix A :

Lemma 3.26. Es gilt

$$A_{ff}^{-1} \mathbf{1} = \tilde{A}_{ff}^{-1} \mathbf{1}, \quad \tilde{A}_{ff}^{-1} = \frac{1}{1 - 8\kappa^2} (I + \kappa D_{ff}).$$

Beweis. Aus Lemma 3.18 folgt unmittelbar $D_{ff}^2 \mathbf{1} = 8\mathbf{1}$ und daher

$$\begin{aligned} A_{ff}^{-1} \mathbf{1} &= \sum_{\nu=0}^{\infty} (\kappa D_{ff})^{\nu} \mathbf{1} = (I + \kappa D_{ff}) \sum_{\nu=0}^{\infty} (\kappa D_{ff})^{2\nu} \mathbf{1} \\ &= (I + \kappa D_{ff}) \sum_{\nu=0}^{\infty} (8\kappa^2)^{\nu} \mathbf{1} = \frac{1}{1 - 8\kappa^2} (I + \kappa D_{ff}) \mathbf{1} = \tilde{A}_{ff}^{-1} \mathbf{1}, \end{aligned}$$

was zu zeigen war. \square

Lemma 3.27. Es sei die Approximation \tilde{A}_{ff}^{-1} der Inversen der Matrix A_{ff} so wie in Lemma 3.26 gewählt,

$$\tilde{A}_{ff}^{-1} = \frac{1}{1 - 8\kappa^2} (I + \kappa D_{ff}).$$

Dann ist die matrixabhängige Restriktion R gegeben durch

$$\begin{aligned} R &= \begin{pmatrix} -A_{cf}\tilde{A}_{ff}^{-1} & I \end{pmatrix} \\ &= \frac{1}{1-8\kappa} \begin{pmatrix} \kappa D_{L_2,L_1} & \kappa^2 D_{L_2,L_1} D_{L_1,L_0} & (1-8\kappa)I \end{pmatrix}. \end{aligned}$$

Beweis. Ausrechnen. □

Bei der in Lemma 3.27 dargestellten Wahl der approximativen Inversen wirkt die Restriktion auf die konstante Gitterfunktion so wie die „exakte“ Restriktion aus (2.26). Analog erhalten wir eine entsprechende Aussage für $P = R^H$.

Exkurs: Restriktion und Prolongation für die Matrix (1.3)

Das folgende Lemma offenbart, daß die matrixabhängige Restriktion bei geeigneter Parameterwahl mit der bekannten Restriktion R_9 aus (2.24) bis auf einen skalaren Faktor übereinstimmt. Eine entsprechende Aussage gilt auch für die Prolongation.

Lemma 3.28. Für die normalisierte Matrix (1.3) aus Beispiel 1.13,

$$A = I - \kappa D, \quad \kappa = \frac{1}{4}$$

stimmt die matrixabhängige Restriktion

$$R = \begin{pmatrix} -A_{cf}\tilde{A}_{ff}^{-1} & I \end{pmatrix}, \quad \tilde{A}_{ff}^{-1} = 2(I + \kappa D_{ff})$$

bis auf Skalierung mit der 9-Punkt-Restriktion R_9 überein.

Beweis. Ist $\tilde{A}_{ff}^{-1} = \alpha(I + \kappa\beta D_{ff})$, so lautet die Restriktion

$$R = \frac{1}{16} \begin{bmatrix} 2\alpha\beta & 4\alpha & 2\alpha\beta \\ 4\alpha & 16 & 4\alpha \\ 2\alpha\beta & 4\alpha & 2\alpha\beta \end{bmatrix}.$$

Für $\alpha = 2$ und $\beta = 1$ erhalten wir bis auf den skalaren Faktor die Restriktion R_9 . □

Approximation des Schurkomplements

Nachdem mittels Lemma 3.27 die Wirkung der Restriktion und Prolongation auf die konstante Gitterfunktion festgeschrieben wurde, verfolgen wir das Ziel, das exakte Schurkomplement und insbesondere seine Wirkung auf Eigenvektoren zu kleinen Eigenwerten gut zu approximieren. Dazu betrachten wir zwei Ansätze zur Approximation des Schurkomplements.

Der erste Ansatz repräsentiert einen 5-Punkte-Stern auf dem Grobgitter:

Lemma 3.29. Es sei $\tilde{S}_A = I - \omega\kappa^2 D_{L_2, L_1} D_{L_1, L_2}$. Dann gilt $\tilde{S}_A \mathbf{1} = S_A \mathbf{1}$ für $\omega = 1/(1 - 8\kappa^2)$.

Beweis. Offensichtlich ist

$$\tilde{S}_A \mathbf{1} = (I - \omega\kappa^2 D_{L_2, L_1} D_{L_1, L_2}) \mathbf{1} = (1 - 8\omega\kappa^2) \mathbf{1},$$

und ein Vergleich mit Lemma 3.25 liefert das gewünschte Resultat. \square

Eine verbesserte Approximation an das exakte Schurkomplement erwarten wir von einem 9-Punkte-Stern:

Lemma 3.30. Es sei $\tilde{S}_A = I - \kappa^2 D_{L_2, L_1} (I + \omega\kappa^2 D_{L_1, L_0} D_{L_0, L_1}) D_{L_1, L_2}$. Dann gilt $\tilde{S}_A \mathbf{1} = S_A \mathbf{1}$ für $\omega = 1/(1 - 8\kappa^2)$.

Beweis. Wir verifizieren

$$\tilde{S}_A \mathbf{1} = \mathbf{1} - 8\kappa^2 \mathbf{1} + 64\omega\kappa^4 \mathbf{1} = \mathbf{1} - 8\kappa^2 (1 + 8\omega\kappa^2) \mathbf{1}.$$

Für $\omega^{-1} = 1 - 8\kappa^2$ erhalten wir

$$1 + 8\omega\kappa^2 = 1 + \frac{8\kappa^2}{1 - 8\kappa^2} = \frac{1}{1 - 8\kappa^2},$$

also $\tilde{S}_A \mathbf{1} = S_A \mathbf{1}$. \square

Ein Blick auf das exakte Schurkomplement (3.13) offenbart, daß sowohl der 5-Punkte- als auch der 9-Punkte-Stern auf der Approximation

$$(I - \kappa^2 D_{L_1, L_0} D_{L_0, L_1})^{-1} \approx \frac{1}{1 - 8\kappa^2} I$$

beruhen. Dies entspricht einer Approximation des Spektrums durch den größten Eigenwert $1/(1 - 8\kappa^2)$, denn es gilt:

Lemma 3.31. Es ist $\lambda_{\min}(I - \kappa^2 D_{L_1, L_0} D_{L_0, L_1}) = 1 - 8\kappa^2$.

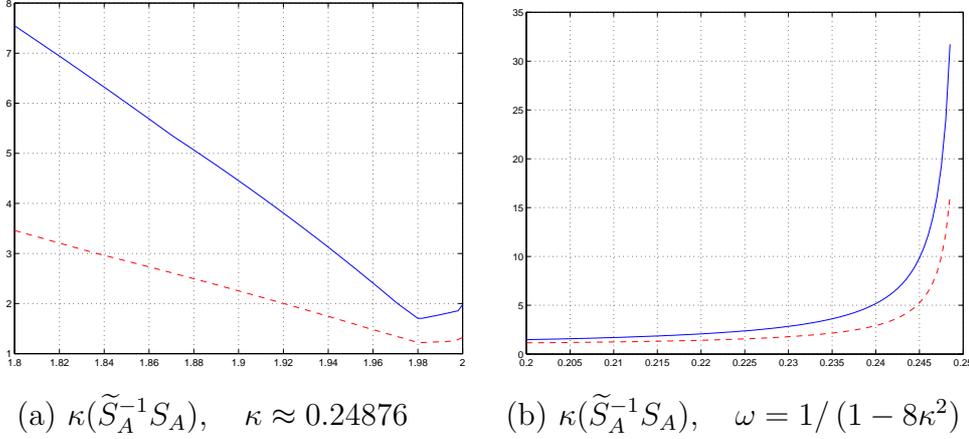


Abbildung 3.9: Die Kondition des präkonditionierten Schurkomplements für verschiedene ω und festes κ sowie für verschiedene κ und optimales ω ($N = 16$). Erwartungsgemäß ist der 9-Punkte-Stern dem 5-Punkte-Stern überlegen, und für hinreichend großes κ beträgt die Verbesserung der Kondition ungefähr einen Faktor 2.

Beweis. Mit Lemma 3.32 und Lemma 3.33 ist dies klar. □

Numerische Experimente zeigen, daß für beide vorgestellten Ansätze zur Approximation des Schurkomplements die Kondition des präkonditionierten Schurkomplements $\tilde{S}_A^{-1}S_A$ bei $\omega = 1/(1 - 8\kappa^2)$ minimal ist, vgl. Abb. 3.9.

Zusammenfassend ist für eine kalte Konfiguration mit dem hier hergeleiteten Schurkomplement-Präkonditionierer ein effizientes Multilevel-Verfahren vorhanden. Der Präkonditionierer besteht aus einer Restriktion und einer Prolongation, welche ähnlich sind zu der bei geometrischen Mehrgitterverfahren verwendeten Restriktion R_9 und Prolongation P_9 . Als Schurkomplement-Approximation bevorzugen wir den vorgeschlagenen 9-Punkte-Stern, da dieser das exakte Schurkomplement besser als der ebenfalls vorgestellte 5-Punkte-Stern approximiert. Die Konstruktion eines echten Multilevel-Verfahrens erfolgt in natürlicher Weise. Für konstante Gitterfunktionen wirkt der Schurkomplement-Präkonditionierer wie die exakte Inverse der Matrix A , und daher ist diese Präkonditionierung für das Dirichlet-Problem so effizient wie die Präkonditionierung durch klassische (geometrische) Mehrgitterverfahren.

Die (realistische und die) heiße Konfiguration

Wir wenden uns ab jetzt dem Fall einer heißen Konfiguration zu und untersuchen Schurkomplement-Verfahren für diese ungeordneten Systeme. Der Fall

einer realistischen Konfiguration wir nicht gesondert behandelt, da dieser sich vom Fall einer heißen Konfiguration (nach Auswertung umfangreicher numerischer Experimente) kaum unterscheidet.

Analog zur Diskussion von Schurkomplement-Verfahren im Falle einer kalten Konfiguration werden wir

- a) lokal wirkende, approximative Inverse der Matrix A_{ff} und
- b) geeignete Schurkomplement-Approximationen

konstruieren. Beweise theoretischer Aussagen, welche es ermöglichen, die Güte der Approximation zu beurteilen, werden wir wiederholt auf den Fall geordneter Systeme zurückspielen, da hierfür entsprechende Aussagen leicht bewiesen werden können.

Die Matrix A_{ff}

Zunächst werden wir zeigen, daß die Matrix $A_{ff} = I - \kappa D_{ff}$ gut konditioniert ist. Dazu benötigen wir einige Aussagen über die Spektren der beteiligten Matrizen.

Lemma 3.32. Es ist $\sigma(D_{L_i, L_1} D_{L_1, L_i}) \subseteq [0, 8]$ für $i = 0, 2$.

Beweis. Für $i = 0, 2$ folgern wir analog zum Beweis von Lemma 3.10, daß $\rho(D_{L_i, L_1} D_{L_1, L_i} - 4I) \leq 4$ ist, also

$$0 \leq \lambda(D_{L_i, L_1} D_{L_1, L_i}) \leq 8$$

gilt. Für eine kalte Konfiguration sind die Abschätzungen scharf. □

Lemma 3.33. Es gilt $\sigma(D_{L_1, L_i} D_{L_i, L_1}) = \sigma(D_{L_i, L_1} D_{L_1, L_i}) \cup \{0\}$ für $i = 0, 2$.

Beweis. [50], Satz 6.4.4. □

Lemma 3.34. Es gilt $\text{diag}(D_{L_1, L_i} D_{L_i, L_1}) = 2$ für $i = 0, 2$.

Beweis. Klar. □

Lemma 3.35. Es gilt $\rho(D_{ff}) \leq 2\sqrt{2}$.

Beweis. Wir kombinieren Lemma 3.22, Lemma 3.32 sowie Lemma 3.33 zu

$$\sigma(D_{ff}^2) = \sigma(D_{L_0, L_1} D_{L_1, L_0}) \cup \sigma(D_{L_1, L_0} D_{L_0, L_1}) \subseteq [0, 8].$$

Es folgt $\sigma(D_{ff}) \subseteq [-2\sqrt{2}, 2\sqrt{2}]$ und $\rho(D_{ff}) \leq 2\sqrt{2}$. □

Lemma 3.36. Mit λ ist auch $-\lambda$ ein Eigenwert der Matrix D_{ff} .

Beweis. Völlig analog zum Beweis von Lemma 3.13. \square

Es ist nun möglich zu beweisen, daß die Matrix A_{ff} gut konditioniert ist:

Satz 3.37. Die Kondition $\kappa(A_{ff})$ der Matrix $A_{ff} = I - \kappa D_{ff}$ genügt der Abschätzung

$$\kappa(A_{ff}) \leq \frac{1 + 2\sqrt{2}\kappa}{1 - 2\sqrt{2}\kappa} < \frac{1 + 2\sqrt{2}\kappa_s}{1 - 2\sqrt{2}\kappa_s} < 12.34, \quad 0 \leq \kappa < \kappa_s.$$

Beweis. Klar, da wir für das Spektrum der Matrix A_{ff} die Abschätzungen

$$\begin{aligned} 0 &\ll 0.15 < 1 - 2\sqrt{2}\kappa_s < 1 - 2\sqrt{2}\kappa \\ &\leq \lambda(A_{ff}) \leq 1 + 2\sqrt{2}\kappa < 1 + 2\sqrt{2}\kappa_s < 1.85 \end{aligned}$$

unmittelbar aus Lemma 3.35 gewinnen. \square

Die approximative Inverse \tilde{A}_{ff}^{-1}

Da die Matrix A_{ff} gut konditioniert ist, kann die approximative Inverse beispielsweise implizit durch einige wenige Schritte eines einfachen Iterationsverfahrens definiert werden. Ist $P - Q$ die das einfache Iterationsverfahren induzierende Zerlegung, so entspricht diese Vorgehensweise nach Lemma 1.42 der Approximation der Inversen durch

$$\tilde{A}_{ff}^{-1} = \sum_{\nu=0}^{k-1} (P^{-1}Q)^\nu P^{-1}. \quad (3.14)$$

Wir formulieren zunächst explizit einige approximative Inverse und diskutieren anschließend die verschiedenen Ansätze.

Jacobi-Präkonditionierer

Lemma 3.38. Die durch die Jacobi-Zerlegung der Matrix A_{ff} induzierte approximative Inverse ist gegeben durch

$$\tilde{A}_{ff}^{-1} = \sum_{\nu=0}^{k-1} (\kappa D_{ff})^\nu.$$

Beweis. Siehe (1.8) mit $\omega = 1$. \square

GS-Präkonditionierer

Lemma 3.39. Die durch die GS-Zerlegung

$$A_{ff} = \begin{pmatrix} I & 0 \\ -\kappa D_{L_0, L_1} & I \end{pmatrix} - \begin{pmatrix} 0 & \kappa D_{L_1, L_0} \\ 0 & 0 \end{pmatrix}$$

der Matrix A_{ff} induzierte approximative Inverse ist gegeben durch

$$\tilde{A}_{ff}^{-1} = \left(I + \sum_{\nu=1}^{k-1} \begin{pmatrix} 0 & \kappa D_{L_1, L_0} (\kappa^2 D_{L_0, L_1} D_{L_1, L_0})^{\nu-1} \\ 0 & (\kappa^2 D_{L_0, L_1} D_{L_1, L_0})^\nu \end{pmatrix} \right) \begin{pmatrix} I & 0 \\ \kappa D_{L_0, L_1} & I \end{pmatrix}.$$

Beweis. Siehe (1.10) mit $\omega = 1$. □

SGS- und ILU-Präkonditionierer

Sowohl die durch die SGS-Zerlegung ($\alpha = 1$) als auch durch die direkte ILU-Zerlegung ($\alpha = 1 - 2\kappa^2$) zum Muster der Nichtnullelemente der Matrix A induzierte approximative Inverse wird durch das folgende Lemma gegeben, vgl. Lemma 3.34.

Lemma 3.40. Es sei $R = (\alpha - 1)I + \kappa^2 D_{L_0, L_1} D_{L_1, L_0}$. Die durch die Zerlegung

$$A_{ff} = \begin{pmatrix} I & 0 \\ -\kappa D_{L_0, L_1} & I \end{pmatrix} \begin{pmatrix} I & -\kappa D_{L_1, L_0} \\ 0 & \alpha I \end{pmatrix} - \begin{pmatrix} 0 & 0 \\ 0 & R \end{pmatrix}$$

der Matrix A_{ff} induzierte approximative Inverse ist gegeben durch

$$\tilde{A}_{ff}^{-1} = \sum_{\nu=0}^{k-1} \begin{pmatrix} 0 & \alpha^{-1} \kappa D_{L_1, L_0} R \\ 0 & \alpha^{-1} R \end{pmatrix}^\nu \begin{pmatrix} I & \alpha^{-1} \kappa D_{L_1, L_0} \\ 0 & \alpha^{-1} I \end{pmatrix} \begin{pmatrix} I & 0 \\ \kappa D_{L_0, L_1} & I \end{pmatrix}.$$

Beweis. Klar. □

Praktikable Präkonditionierer

Die Güte der Approximation der Inversen durch die vorgestellten Präkonditionierer wird durch die Wahl des Parameters k gesteuert. Jedoch ist klar, daß der durch die Präkonditionierung entstehende zusätzliche Aufwand mit k wächst, so daß eine Präkonditionierung für $k > 3$ i.a. nicht mehr praktikabel ist.

Wir wählen den Jacobi-Präkonditionierer aus Lemma 3.38 mit $k = 2$, vernachlässigen Terme höherer Ordnung und bestimmen die Kondition der präkonditionierten Matrix A_{ff} :

Lemma 3.41. Es sei $\tilde{A}_{ff}^{-1} = I + \kappa D_{ff}$. Dann gilt

$$\sigma(\tilde{A}_{ff}^{-1} A_{ff}) = \sigma(I - \kappa^2 D_{L_0, L_1} D_{L_1, L_0}) \cup \sigma(I - \kappa^2 D_{L_1, L_0} D_{L_0, L_1}),$$

und die Kondition der präkonditionierten Matrix $\tilde{A}_{ff}^{-1} A_{ff}$ genügt der Abschätzung

$$\kappa(\tilde{A}_{ff}^{-1} A_{ff}) \leq \frac{1}{1 - 8\kappa^2} < \frac{1}{1 - 8\kappa_s^2} < 3.58, \quad 0 \leq \kappa < \kappa_s.$$

Beweis. Trivial, da $\tilde{A}_{ff}^{-1} A_{ff} = I - \kappa^2 D_{ff}^2$ ist. \square

Mit Lemma 3.56 treten die Eigenwerte der Jacobi-präkonditionierten Matrix A_{ff} mit gerader Vielfachheit auf.

Für die durch die GS-, SGS- und ILU-Zerlegung induzierte approximative Inverse erhalten wir eine zu Lemma 3.41 analoge Aussage, jedoch ist die Eigenwertverteilung in diesen Fällen wesentlich günstiger als bei der durch die Jacobi-Zerlegung induzierten approximativen Inversen:

Lemma 3.42. Es sei

$$\tilde{A}_{ff}^{-1} = \begin{pmatrix} I & 0 \\ \kappa D_{L_0, L_1} & I \end{pmatrix}.$$

Dann gilt

$$\sigma(\tilde{A}_{ff}^{-1} A_{ff}) = \sigma(I - \kappa^2 D_{L_0, L_1} D_{L_1, L_0}) \cup \{1\}, \quad \kappa(\tilde{A}_{ff}^{-1} A_{ff}) \leq \frac{1}{1 - 8\kappa^2}.$$

Beweis. Es gilt

$$\tilde{A}_{ff}^{-1} A_{ff} = \begin{pmatrix} I & -\kappa D_{L_1, L_0} \\ 0 & I - \kappa^2 D_{L_0, L_1} D_{L_1, L_0} \end{pmatrix}.$$

Mit Lemma 3.32 folgt nun die Behauptung. \square

Lemma 3.43. Es sei

$$\tilde{A}_{ff}^{-1} = \begin{pmatrix} I & \alpha^{-1} \kappa D_{L_1, L_0} \\ 0 & \alpha^{-1} I \end{pmatrix} \begin{pmatrix} I & 0 \\ \kappa D_{L_0, L_1} & I \end{pmatrix}.$$

Dann gilt

$$\sigma(\tilde{A}_{ff}^{-1} A_{ff}) = \alpha^{-1} \sigma(I - \kappa^2 D_{L_0, L_1} D_{L_1, L_0}) \cup \{1\}, \quad \kappa(\tilde{A}_{ff}^{-1} A_{ff}) \leq \frac{1}{1 - 8\kappa^2}.$$

Beweis. Es gilt

$$\begin{aligned}\tilde{A}_{ff}^{-1}A_{ff} &= \begin{pmatrix} I & \alpha^{-1}\kappa D_{L_1,L_0} \\ 0 & \alpha^{-1}I \end{pmatrix} \begin{pmatrix} I & -\kappa D_{L_1,L_0} \\ 0 & I - \kappa^2 D_{L_0,L_1} D_{L_1,L_0} \end{pmatrix} \\ &= \begin{pmatrix} I & \star \\ 0 & \alpha^{-1}(I - \kappa^2 D_{L_0,L_1} D_{L_1,L_0}) \end{pmatrix}.\end{aligned}$$

Mit Lemma 3.32 folgt nun die Behauptung. \square

Die präkonditionierte Matrix A_{ff} besitzt den Eigenwert 1 mit Vielfachheit $n/2$, falls der Präkonditionierer aus Lemma 3.42 bzw. Lemma 3.43 verwendet wird.

Da bei diesen Ansätzen Terme höherer Ordnung vernachlässigt werden, erscheint es naheliegend, eine Gewichtung des Nichtdiagonalanteils κD_{ff} einzuführen, um die fehlenden Terme zu subsumieren. Tatsächlich zeigt aber das folgende Lemma, daß der einfache Ansatz jedenfalls bei einer Jacobi-Präkonditionierung bereits optimal ist.

Lemma 3.44. Es sei $\tilde{A}_{ff}^{-1} = I + \omega \kappa D_{ff}$ mit $\omega \in \mathbb{R}$. Dann ist die Kondition der präkonditionierten Matrix $\tilde{A}_{ff}^{-1}A_{ff}$ minimal für $\omega = 1$.

Beweis. Mit $p(t) = (1 + \omega t)(1 - t)$ gilt $\tilde{A}_{ff}^{-1}A_{ff} = p(\kappa D_{ff})$. Nach Lemma 3.36 liegen die Eigenwerte symmetrisch um den Ursprung. Eine einfache Kurvendiskussion offenbart die Minimalität der Kondition für $\omega = 1$: Für $\omega > 1$ ist der rechte Zweig ($t > 0$) der Funktion p größer und der linke ($t < 0$) kleiner als der für $\omega = 1$, vgl. Abb. 3.10, also ist die Kondition größer. Analog für $\omega < 1$. \square

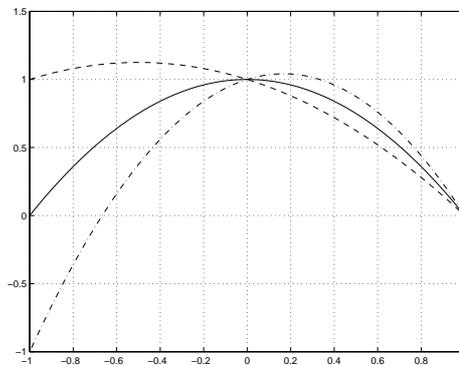


Abbildung 3.10: Die Funktion $p(t) = (1 + \omega t)(1 - t)$ für verschiedene ω .

Ein nächster Schritt wäre die Hinzunahme von Termen höherer Ordnung. Tatsächlich ist die Kondition der präkonditionierten Matrix A_{ff} bereits hinreichend gut, so daß die zu erwartende minimale Verbesserung der Kondition den hierfür zu leistenden Aufwand nicht rechtfertigt.

Restriktion und Prolongation

Da wir bereits eine hinreichend gute approximative Inverse der Matrix A_{ff} bestimmt haben, können wir diese zur Konstruktion der matrixabhängigen Restriktion und Prolongation verwenden. Dabei stellt sich heraus, daß die Restriktion (Prolongation) bei einer durch die SGS-Zerlegung induzierten approximativen Inversen ($k = 2$) identisch ist mit der Restriktion (Prolongation) bei einer durch die Jacobi-Zerlegung induzierten approximativen Inversen ($k = 3$):

Lemma 3.45. Ist \tilde{A}_{ff}^{-1} die durch die SGS-Zerlegung induzierte approximative Inverse, so entspricht die matrixabhängige Restriktion und Prolongation gerade dem Ansatz

$$R = \left(\kappa D_{cf} (I + \kappa D_{ff} + \kappa^2 D_{ff}^2) \quad I \right), \quad P = \begin{pmatrix} \kappa (I + \kappa D_{ff} + \kappa^2 D_{ff}^2) D_{fc} \\ I \end{pmatrix}.$$

Beweis. Da die durch eine approximative Inverse der Gestalt

$$I + \kappa D_{ff} + \kappa^2 \begin{pmatrix} D_{L_1, L_0} D_{L_0, L_1} & 0 \\ 0 & \star \end{pmatrix}$$

beschriebenen Kopplungen zwischen Gitterpunkten aus L_0 weder zur matrixabhängigen Restriktion noch zur matrixabhängigen Prolongation beitragen, sind die Restriktion und die Prolongation bei einer durch die SGS-Zerlegung induzierten approximativen Inversen

$$\tilde{A}_{ff}^{-1} = I + \kappa D_{ff} + \kappa^2 \begin{pmatrix} D_{L_1, L_0} D_{L_0, L_1} & 0 \\ 0 & 0 \end{pmatrix}$$

insbesondere identisch mit der Restriktion und der Prolongation bei einer durch die Jacobi-Zerlegung induzierten approximativen Inversen,

$$I + \kappa D_{ff} + \kappa^2 D_{ff}^2 = I + \kappa D_{ff} + \kappa^2 \begin{pmatrix} D_{L_1, L_0} D_{L_0, L_1} & 0 \\ 0 & D_{L_0, L_1} D_{L_1, L_0} \end{pmatrix},$$

was zu zeigen war. \square

Approximation des Schurkomplements

Der Diskussion der Restriktion und Prolongation folgt nun die Approximation des exakten Schurkomplements

$$S_A = I - \kappa^2 D_{L_2, L_1} (I - \kappa^2 D_{L_1, L_0} D_{L_0, L_1})^{-1} D_{L_1, L_2}.$$

Dazu müssen wir eine geeignete Approximation der Matrix

$$(I - \kappa^2 D_{L_1, L_0} D_{L_0, L_1})^{-1} = \frac{1}{1 - 2\kappa^2} \left(I - \frac{\kappa^2}{1 - 2\kappa^2} (D_{L_1, L_0} D_{L_0, L_1} - 2I) \right)^{-1}$$

konstruieren.

Lemma 3.46. Es gilt

$$\begin{aligned} \sigma \left(I - \frac{\kappa^2}{1 - 2\kappa^2} (D_{L_1, L_0} D_{L_0, L_1} - 2I) \right) &\subseteq \frac{1}{1 - 2\kappa^2} [1 - 8\kappa^2, 1] \\ &\subseteq \frac{1}{1 - 2\kappa_s^2} [1 - 8\kappa_s^2, 1] \subseteq [0.34, 1.22], \quad 0 \leq \kappa < \kappa_s. \end{aligned}$$

Beweis. Siehe Lemma 3.33 und Lemma 3.32. □

Wir folgern, daß die Matrix $(I - \kappa^2 D_{L_1, L_0} D_{L_0, L_1})^{-1}$ gut konditioniert ist. Es bietet sich daher an, zur Approximation den Ansatz

$$\frac{1}{1 - 2\kappa^2} \left(\omega_1 I + \omega_2 \frac{\kappa^2}{1 - 2\kappa^2} (D_{L_1, L_0} D_{L_0, L_1} - 2I) \right) \quad (3.15)$$

zu verwenden. Die Matrix (3.15) beschreibt einen 7-Punkte-Stern und die zugehörige Schurkomplement-Approximation

$$\tilde{S}_A = I - \frac{\kappa^2}{1 - 2\kappa^2} D_{L_2, L_1} \left(\omega_1 I + \omega_2 \frac{\kappa^2}{1 - 2\kappa^2} (D_{L_1, L_0} D_{L_0, L_1} - 2I) \right) D_{L_1, L_2} \quad (3.16)$$

einen 9-Punkte-Stern, vgl. Abb. 3.11.

Für $\omega_1 \neq 0$ und $\omega_2 = 0$ reduziert sich die Schurkomplement-Approximation auf eine Nächste-Nachbar-Kopplung auf den Gitterpunkten aus L_2 ,

$$\tilde{S}_A = I - \omega \frac{\kappa^2}{1 - 2\kappa^2} D_{L_2, L_1} D_{L_1, L_2}, \quad (3.17)$$

bei der die Nichtdiagonalanteile durch einen Relaxationsparameter ω gewichtet werden.

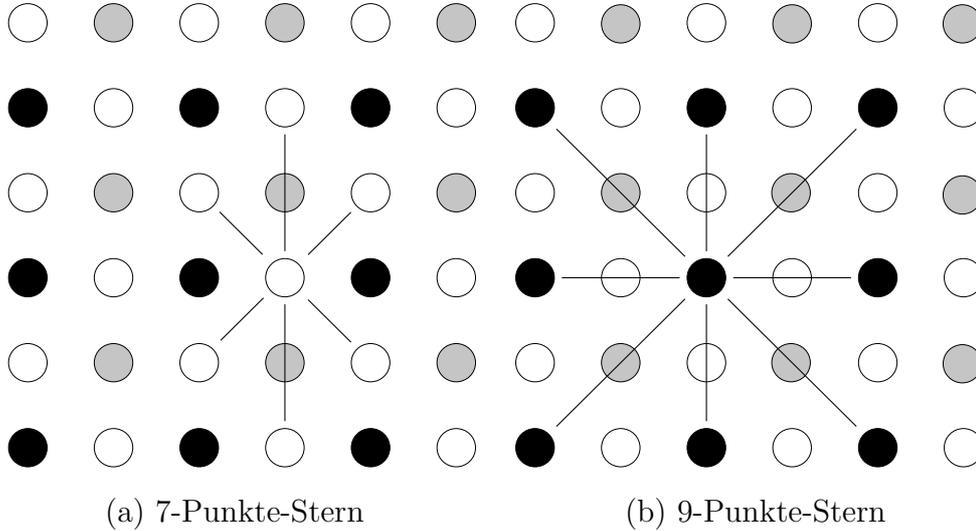


Abbildung 3.11: Der 7-Punkte-Stern der Matrix (3.15) und der 9-Punkte-Stern der Schurkomplement-Approximation \tilde{S}_A .

Eine mögliche und naheliegende Wahl in (3.15) bzw. in (3.16) ist $\omega_1 = 1$. Um die Approximation zu verbessern, bestimmen wir eine erste Näherung der Diagonale der Matrix

$$\left(I - \frac{\kappa^2}{1 - 2\kappa^2} (D_{L_1, L_0} D_{L_0, L_1} - 2I) \right)^{-1},$$

indem wir die entsprechende Neumann-Reihe trunkieren,

$$\omega_1 = 1 + 6 \left(\frac{\kappa^2}{1 - 2\kappa^2} \right)^2 + 12 \left(\frac{\kappa^2}{1 - 2\kappa^2} \right)^3 \leq 1.09, \quad 0 \leq \kappa < \kappa_s. \quad (3.18)$$

Den optimalen Wert ω_2 , so daß die Kondition des präkonditionierten Schurkomplements minimal wird, ermitteln wir konfigurationsabhängig. Numerische Experimente zeigen aber, daß im allgemeinen $\omega_1 = 1$ und $\omega_2 = 2$ eine gute Parameterwahl ist, vgl. Abb. 3.12. Für ω_1 aus (3.18) erscheint $\omega_2 \approx 1.65$ sinnvoll, und die Kondition des präkonditionierten Schurkomplements ist gegenüber $\omega_1 = 1$ und $\omega_2 = 2$ kleiner. In der Abbildung ist außerdem deutlich zu erkennen, daß der Übergang von einer Nächsten-Nachbar-Kopplung hin zu einem 9-Punkte-Stern zu einer Verbesserung der Kondition des präkonditionierten Schurkomplements um einen Faktor 5 – 6 führt. Aus diesem Grund ist der 9-Punkte-Stern dem 5-Punkte-Stern vorzuziehen.

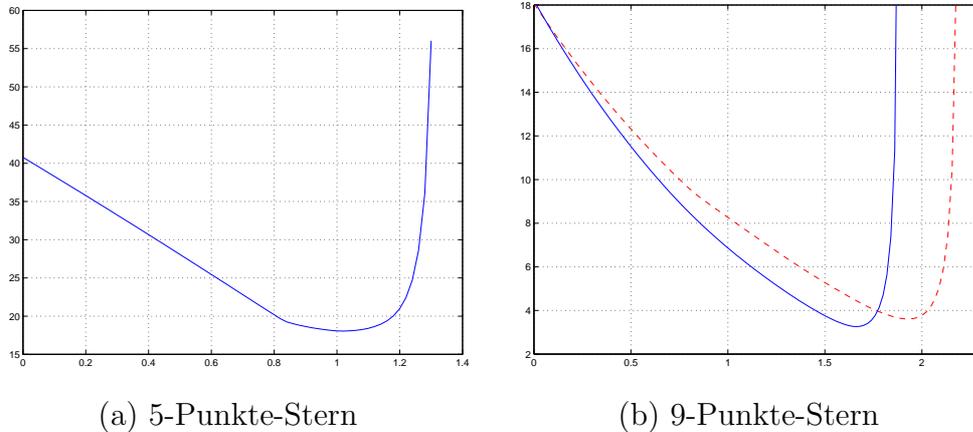


Abbildung 3.12: Die Kondition des präkonditionierten Schurkomplements in Abhängigkeit des Relaxationsparameters für eine gegebene heiße Konfiguration. Im Falle des 9-Punkte-Sterns ist die Kondition des präkonditionierten Systems für ω_1 aus (3.18) und $\omega_1 = 1$ (gestrichelte Linie) aufgetragen.

3.4.3 Schurkomplement-Verfahren zur Lösung des odd-even reduzierten Systems

Wir werden im folgenden darlegen, daß die vorgestellten Ideen zur Konstruktion von Schurkomplement-Präkonditionierern auch für das odd-even reduzierte System erfolgreich angewendet werden können. Wir werden beweisen, daß die Schurkomplement-Verfahren zur Lösung des odd-even reduzierten Systems denen zur Lösung des originären Systems deutlich überlegen sind.

Ausgehend vom odd-even reduzierten System stellen wir in diesem Abschnitt fest, daß es im Gegensatz zum ursprünglichen, nicht reduzierten System nun möglich ist, die Kopplungen des odd-even reduzierten Systems einzuteilen in schwache und starke. Trotz der auf dieser algebraischen Charakterisierung basierenden Auswahl der groben Gitterpunkte kann eine Hierarchie von Gittern definiert werden, welche identisch ist mit der auf rein geometrischen Kriterien beruhenden Auswahl.

F/C-Anordnung der geraden Gitterpunkte

Betrachten wir eine F/C-Anordnung der geraden Gitterpunkte mit $\Omega_f = L_0$ und $\Omega_c = L_2$, so sind die Matrizen D_{eo} und D_{oe} gegeben durch

$$D_{eo} = \begin{pmatrix} D_{L_0, L_1} \\ D_{L_2, L_1} \end{pmatrix}, \quad D_{oe} = (D_{L_1, L_0} \quad D_{L_1, L_2}),$$

und die odd-even reduzierte Matrix $A^{(e)} = A_e$ besitzt die Gestalt

$$A^{(e)} = \begin{pmatrix} A_{ff}^{(e)} & A_{fc}^{(e)} \\ A_{cf}^{(e)} & A_{cc}^{(e)} \end{pmatrix} = I - \kappa^2 \begin{pmatrix} D_{ff}^{(e)} & D_{fc}^{(e)} \\ D_{cf}^{(e)} & D_{cc}^{(e)} \end{pmatrix} \quad (3.19)$$

mit

$$\begin{aligned} D_{ff}^{(e)} &= D_{L_0, L_1} D_{L_1, L_0}, & D_{fc}^{(e)} &= D_{L_0, L_1} D_{L_1, L_2}, \\ D_{cf}^{(e)} &= D_{L_2, L_1} D_{L_1, L_0}, & D_{cc}^{(e)} &= D_{L_2, L_1} D_{L_1, L_2}. \end{aligned}$$

Wir bestimmen im folgenden die Nichtnullblöcke der beteiligten Teilmatrizen.

Besetztheitsmuster der Blockdiagonalmatrizen

Lemma 3.47. Für $i = 0, 2$ ist $\text{diag}(D_{L_i, L_1} D_{L_1, L_i}) = 4I$.

Beweis. Für $i = 0, 2$ gilt

$$(D_{L_i, L_1} D_{L_1, L_i})_{x, x} = \sum_{\mu=1}^2 U_\mu(x) U_\mu^H(x) + U_\mu(x - e_\mu) U_\mu^H(x - e_\mu).$$

Da $|U_\mu(x)| = 1$ ist, erhalten wir die Behauptung unmittelbar. \square

Die Matrizen $A_{cc}^{(e)}$ und $A_{ff}^{(e)}$ repräsentieren eine Nächste-Nachbar-Kopplung auf den groben bzw. den feinen Gitterpunkten:

Lemma 3.48. Für $i = 0, 2$ besitzt die Matrix $D_{L_i, L_1} D_{L_1, L_i}$ genau 5 Nicht-nullelemente an den Positionen ($\mu = 1, 2$)

$$\begin{aligned} (x, x) &: 4 \\ (x, x + 2e_\mu) &: U_\mu(x) U_\mu(x + e_\mu) \\ (x, x - 2e_\mu) &: U_\mu^H(x - e_\mu) U_\mu^H(x - 2e_\mu). \end{aligned}$$

Beweis. Klar. \square

Besetztheitsmuster der Nichtdiagonalblöcke

Zwischen benachbarten geraden Gitterpunkten $x \in L_i(\Omega_h)$ und $y \in L_j(\Omega_h)$, $i \neq j$, gibt es genau zwei Wege der Länge 2 bzgl. des ursprünglichen Gitters Ω_h . Diese zwei Wege sind für Systeme mit konstanten Gewichten identisch, für ungeordneter Systeme hingegen sind diese unterschiedlich gewichtet.

Lemma 3.49. Für $i, j = 0, 2$ mit $i \neq j$ besitzt die Matrix $D_{L_i, L_1} D_{L_1, L_j}$ genau 4 Nichtnullelemente an den Positionen (x, y) ,

$$\begin{aligned} y = x + e_\mu + e_\nu & : U_\mu(x)U_\nu(x + e_\mu) + U_\nu(x)U_\mu(x + e_\nu) \\ y = x + e_\mu - e_\nu & : U_\mu(x)U_\nu^H(y) + U_\nu^H(x - e_\nu)U_\mu(x - e_\nu) \\ y = x - e_\mu + e_\nu & : U_\mu^H(x - e_\mu)U_\nu(x - e_\mu) + U_\nu(x)U_\mu^H(y) \\ y = x - e_\mu - e_\nu & : U_\mu^H(x - e_\mu)U_\nu^H(y) + U_\nu^H(x - e_\nu)U_\mu^H(y), \end{aligned}$$

wobei $\mu \neq \nu$ ist.

Beweis. Nachrechnen. □

Diagonaldominanz

Lemma 3.50. Die Matrizen $A_{ff}^{(e)}$ und $A_{cc}^{(e)}$ sind diagonaldominant.

Beweis. Es gilt

$$1 - 4\kappa^2 = |(A_{ff}^{(e)})_{ii}| > \sum_{j \neq i} |(A_{ff}^{(e)})_{ij}| = 4\kappa^2 \text{ für alle } |\kappa| < \frac{1}{2\sqrt{2}},$$

insbesondere also wegen $1/(2\sqrt{2}) \approx 0.35$ für $0 \leq \kappa < \kappa_s$. □

Die Eigenschaft 'streng diagonaldominant' rechtfertigt eine Approximation der Inversen der Diagonalblöcke des odd-even reduzierten Systems durch eine lokal wirkende Matrix.

Nachdem wir uns die einzelnen Teilmatrizen der odd-even reduzierten Matrix genauer angesehen haben, werden wir nun mit Hilfe einer Einteilung der Kopplungen in schwache und starke erklären, warum die gewählte F/C-Anordnung mit $\Omega_f = L_0$ und $\Omega_c = L_2$ eine sinnvolle Anordnung der Gitterpunkte ist.

Schwache und starke Kopplungen

Durch die folgenden Überlegungen wird eine Klassifizierung der Kopplungen in starke und schwache im Sinne der Definition 2.18 mit $\beta = 0.5$ nahegelegt.

Lemma 3.51. Für $i, j = 0, 2$ seien $x \in L_i$ und $y \in L_j$ zwei Gitterpunkte mit $x \neq y$. Im Falle einer kalten Konfiguration ist die Kopplung zwischen x und y für $i = j$ schwach, für $i \neq j$ hingegen stark.

Beweis. Für $i = j$ erhalten wir mit Lemma 3.48 sofort $|A_{x,y}^{(e)}| = \kappa^2$, für $i \neq j$ hingegen $|A_{x,y}^{(e)}| = 2\kappa^2$ mit Lemma 3.49. \square

Dieses Resultat läßt sich auf den Fall einer heißen Konfiguration übertragen, indem wir eine analoge Aussage für den Erwartungswert $E(\cdot)$ beweisen.

Satz 3.52. Für $i, j = 0, 2$ seien $x \in L_i$ und $y \in L_j$ zwei Gitterpunkte mit $x \neq y$. Dann gilt

$$|E(A_{x,y}^{(e)})| = \begin{cases} \kappa^2 & : i = j \\ 4\kappa^2/\pi & : i \neq j \end{cases}$$

Beweis. Siehe Lemma A.1 und Satz A.2 im Anhang. \square

Für realistische Konfigurationen läßt sich die Frage nach dem Erwartungswert nicht ohne weiteres beantworten, da in diesem Fall die Gewichte $U_\mu(x)$ nicht mehr gleichverteilt sind.

Verallgemeinern wir die Bedingung (2.21) zu

$$|E(A_{x,y}^{(e)})| > \beta \max_{y \neq x} |E(A_{x,y}^{(e)})|, \quad (3.20)$$

so gilt:

Lemma 3.53. Für $i, j = 0, 2$ seien $x \in L_i$ und $y \in L_j$ zwei Gitterpunkte mit $x \neq y$. Im Falle einer heißen Konfiguration und $\beta = \pi/4$ ist die Kopplung zwischen x und y für $i = j$ im Sinne der verallgemeinerten Bedingung (3.20) schwach.

Beweis. Wegen $|A_{x,y}^{(e)}| = \kappa^2$ für $i = j$ folgt die Behauptung sofort mit Satz 3.52. \square

Für $i \neq j$ ist a priori nicht klar, ob die Kopplung zwischen x und y tatsächlich eine starke oder eine schwache im Sinne der Bedingung (3.20) ist, jedoch erwarten wir aufgrund des Satzes 3.52 eine starke.

Auswahl der Grobgitterpunkte

Wir bezeichnen die Kantenmenge E_S der starken Kopplungen für den Fall einer kalten Konfiguration auch mit E_S^* . Für eine realistische bzw. eine heiße Konfiguration untersuchen wir anstelle des reduzierten Graphen $G_{A^{(e)}}(V, E_S)$ den (evtl. um schwache Kopplungen repräsentierende Kanten erweiterten) Graphen $G_{A^{(e)}}(V, E_S^*)$, vgl. Abb. 3.13.

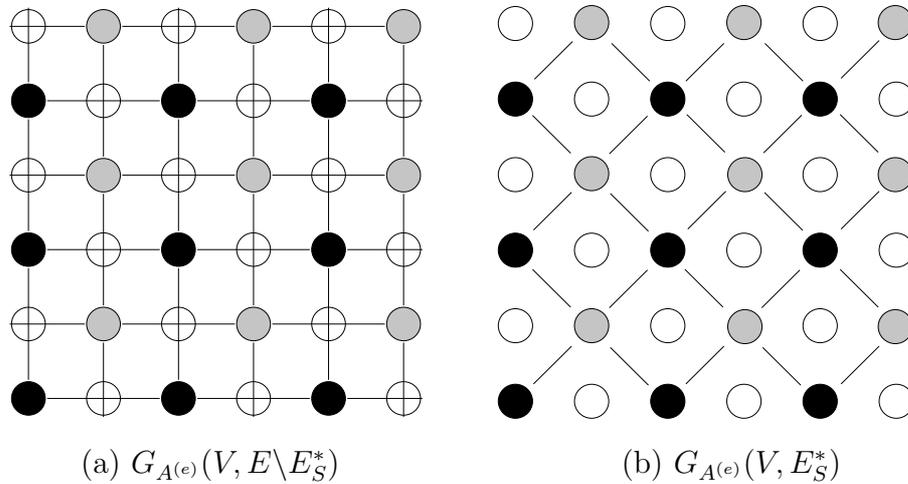


Abbildung 3.13: Die Zerlegung des Graphen $G_{A^{(e)}}(V, E)$ in die Graphen $G_{A^{(e)}}(V, E \setminus E_S^*)$ und $G_{A^{(e)}}(V, E_S^*)$.

Trivialerweise ist $\Omega_f = L_0(\Omega_h)$ eine maximal unabhängige Menge des Graphen $G_{A^{(e)}}(V, E_S^*)$ bzgl. der Kantenmenge E_S^* . Obwohl für realistische und heiße Konfigurationen die Kantenmenge E_S^* nicht notwendig ausschließlich aus starken Kopplungen besteht, ist die durch diese Vorgehensweise motivierte Auswahl der feinen Gitterpunkte offensichtlich ähnlich zu der in rein algebraischen Ansätzen. Die verbleibenden Gitterpunkte bilden das Grobgitter $\Omega_c = L_2(\Omega_h)$. Bemerkenswert bei diesem Ansatz ist es, daß die resultierenden Gitter a priori bekannt sind, so daß aufwendige Auswahlverfahren (engl.: coarsening strategies) entfallen. Insbesondere erscheint durch diese Überlegungen die F/C-Einteilung der geraden Gitterpunkte im Ansatz (3.19) gerechtfertigt.

Bevor wir uns der Konstruktion einer geeigneten Approximation des Schurkomplements zuwenden, untersuchen wir die Eigenschaften der Matrix $A_{ff}^{(e)}$, definieren lokal wirkende Präkonditionierer und analysieren deren Effizienz.

Die Matrix $A_{ff}^{(e)}$

Sehen wir uns die Matrix $A_{ff}^{(e)} = I - \kappa^2 D_{ff}^{(e)}$ genauer an:

Lemma 3.54. Eine obere Schranke der Kondition der Matrix $A_{ff}^{(e)}$ ist gegeben durch

$$\kappa \left(A_{ff}^{(e)} \right) \leq \frac{1}{1 - 8\kappa^2} < \frac{1}{1 - 8\kappa_s^2} < 3.58, \quad 0 \leq \kappa < \kappa_s.$$

Beweis. Die Matrix $A_{ff}^{(e)}$ kann geschrieben werden als

$$A_{ff}^{(e)} = (1 - 4\kappa^2) \left(I - \frac{\kappa^2}{1 - 4\kappa^2} (D_{ff}^{(e)} - 4I) \right).$$

Es sei ρ der Spektralradius der die Nächste-Nachbar-Kopplung auf den feinen Gitterpunkten repräsentierenden Matrix $D_{ff}^{(e)} - 4I$,

$$\rho = \rho \left(D_{ff}^{(e)} - 4I \right).$$

Mit Lemma 3.32 erhalten wir $\rho \leq 4$, womit nun

$$\kappa \left(A_{ff}^{(e)} \right) \leq \left(1 + \rho \frac{\kappa^2}{1 - 4\kappa^2} \right) / \left(1 - \rho \frac{\kappa^2}{1 - 4\kappa^2} \right) \leq \frac{1}{1 - 8\kappa^2}$$

unmittelbar folgt. \square

Trunkierte Neumann-Reihen

Basierend auf einer trunkeierten Neumann-Reihe können wir nun einen ersten, einfachen Prädiktionierer der Matrix $A_{ff}^{(e)}$ einführen:

$$\left(\tilde{A}_{ff}^{(e)} \right)^{-1} = \frac{1}{1 - 4\kappa^2} \left(I + \omega \frac{\kappa^2}{1 - 4\kappa^2} (D_{ff}^{(e)} - 4I) \right). \quad (3.21)$$

Dies ist der Jacobi-Prädiktionierer mit einer zusätzlichen Gewichtung des Nichtdiagonalanteils, welche durch den Parameter $\omega \in \mathbb{R}$ gesteuert wird.

Lemma 3.55. Es sei eine approximative Inverse der Matrix $A_{ff}^{(e)}$ gegeben durch (3.21). Die Kondition der prädiktionierten Matrix ist minimal für $\omega = 1$, und in diesem Fall gilt die Abschätzung

$$\begin{aligned} \kappa \left(\left(\tilde{A}_{ff}^{(e)} \right)^{-1} A_{ff}^{(e)} \right) &\leq (1 - 4\kappa^2)^2 \frac{1}{1 - 8\kappa^2} \\ &< (1 - 4\kappa_s^2)^2 \frac{1}{1 - 8\kappa_s^2} < 1.47, \quad 0 \leq \kappa < \kappa_s. \end{aligned}$$

Beweis. Völlig analog zum Beweis von Lemma 3.44 erhalten wir die Optimalität für $\omega = 1$, vgl. auch Abb. 3.10, und für diese Wahl von ω , also

$$\left(\tilde{A}_{ff}^{(e)}\right)^{-1} A_{ff}^{(e)} = I - \left(\frac{\kappa^2}{1 - 4\kappa^2} \left(D_{ff}^{(e)} - 4I\right)\right)^2, \quad (3.22)$$

gilt die Abschätzung

$$\kappa \left(\left(\tilde{A}_{ff}^{(e)}\right)^{-1} A_{ff}^{(e)}\right) \leq \left(1 - \left(\rho \frac{\kappa^2}{1 - 4\kappa^2}\right)^2\right)^{-1} \leq (1 - 4\kappa^2)^2 \frac{1}{1 - 8\kappa^2},$$

wobei wie zuvor $\rho = \rho \left(D_{ff}^{(e)} - 4I\right)$ ist. \square

Odd-even Anordnung der feinen Gitterpunkte

Das Ziel ist es nun, genauere Aussagen über die Eigenwertverteilung der präkonditionierten Matrix (3.22) herzuleiten. Dazu setzen wir im folgenden eine odd-even Anordnung der feinen (L_0 -)Gitterpunkte voraus:

Ist $\hat{D}_{ff}^{(e)} = D_{ff}^{(e)} - 4I$, so kann die Matrix $A_{ff}^{(e)}$ in der Form

$$A_{ff}^{(e)} = (1 - 4\kappa^2) \left(I - \hat{\kappa} \hat{D}_{ff}^{(e)}\right), \quad \hat{\kappa} = \frac{\kappa^2}{1 - 4\kappa^2}, \quad \hat{D}_{ff}^{(e)} = \begin{pmatrix} 0 & \hat{D}_{oe}^{(e)} \\ \hat{D}_{eo}^{(e)} & 0 \end{pmatrix}$$

mit $\hat{D}_{oe}^{(e)} = D_{oe}^{(e)}$ und $\hat{D}_{eo}^{(e)} = D_{eo}^{(e)}$ geschrieben werden.

Spektrum der Jacobi-präkonditionierten Matrix $A_{ff}^{(e)}$

Lemma 3.56. Es gilt $\sigma(\hat{D}_{eo}^{(e)} \hat{D}_{oe}^{(e)}) = \sigma(\hat{D}_{oe}^{(e)} \hat{D}_{eo}^{(e)})$.

Beweis. [50], Satz 6.4.4. \square

Als eine unmittelbare Folgerung halten wir fest, daß die Eigenwerte der Matrix (3.22) mit gerader Vielfachheit auftreten:

Satz 3.57. Es gilt

$$\sigma \left(\left(\tilde{A}_{ff}^{(e)}\right)^{-1} A_{ff}^{(e)}\right) = (1 - 4\hat{\kappa}^2) \sigma \left(I - \frac{\hat{\kappa}}{1 - 4\hat{\kappa}^2} \left(\hat{D}_{eo}^{(e)} \hat{D}_{oe}^{(e)} - 4I\right)\right).$$

Beweis. Aus der Gleichung

$$\begin{aligned} \left(\tilde{A}_{ff}^{(e)}\right)^{-1} A_{ff}^{(e)} &= I - \hat{\kappa}^2 \left(\hat{D}_{ff}^{(e)}\right)^2 = I - \hat{\kappa}^2 \begin{pmatrix} \hat{D}_{oe}^{(e)} \hat{D}_{eo}^{(e)} & 0 \\ 0 & \hat{D}_{eo}^{(e)} \hat{D}_{oe}^{(e)} \end{pmatrix} \\ &= (1 - 4\hat{\kappa}^2) \left(I - \frac{\hat{\kappa}^2}{1 - 4\hat{\kappa}^2} \begin{pmatrix} \hat{D}_{oe}^{(e)} \hat{D}_{eo}^{(e)} - 4I & 0 \\ 0 & \hat{D}_{eo}^{(e)} \hat{D}_{oe}^{(e)} - 4I \end{pmatrix} \right) \end{aligned}$$

folgt mit Hilfe des Lemma 3.56 sofort die Behauptung. \square

Unvollständige Dreieckszerlegungen

Neben der trunkierten Jacobi-Entwicklung bietet die unvollständige Dreieckszerlegung eine weitere Methode zur Konstruktion effizienter Prädiktionierer:

Lemma 3.58. Es sei $1 - 4\hat{\kappa}^2 \neq 0$. Die Matrizen L , U und R seien definiert durch

$$L = \begin{pmatrix} I & 0 \\ -\hat{\kappa} \hat{D}_{eo}^{(e)} & I \end{pmatrix}, \quad U = \begin{pmatrix} I & -\hat{\kappa} \hat{D}_{oe}^{(e)} \\ 0 & (1 - 4\hat{\kappa}^2)I \end{pmatrix}, \quad R = \hat{\kappa}^2 \begin{pmatrix} 0 & 0 \\ 0 & \hat{D}_{eo}^{(e)} \hat{D}_{oe}^{(e)} - 4I \end{pmatrix}.$$

Dann ist $LU - R$ die direkte ILU-Zerlegung der Matrix $(1 - 4\hat{\kappa}^2)^{-1} A_{ff}^{(e)}$ zum Muster der Nichtnullelemente.

Beweis. Wir rechnen nach:

$$\begin{pmatrix} I & 0 \\ -\hat{\kappa} \hat{D}_{eo}^{(e)} & I \end{pmatrix} \begin{pmatrix} I & -\hat{\kappa} \hat{D}_{oe}^{(e)} \\ 0 & (1 - 4\hat{\kappa}^2)I \end{pmatrix} = \begin{pmatrix} I & -\hat{\kappa} \hat{D}_{oe}^{(e)} \\ -\hat{\kappa} \hat{D}_{eo}^{(e)} & (1 - 4\hat{\kappa}^2)I + \hat{\kappa}^2 \hat{D}_{eo}^{(e)} \hat{D}_{oe}^{(e)} \end{pmatrix}.$$

Aus $\text{diag} \left(\hat{D}_{eo}^{(e)} \hat{D}_{oe}^{(e)} \right) = 4I$ folgt $\text{diag} \left((1 - 4\hat{\kappa}^2)I + \hat{\kappa}^2 \hat{D}_{eo}^{(e)} \hat{D}_{oe}^{(e)} \right) = I$. \square

Die Existenz der direkten ILU-Zerlegung klärt das folgende Lemma:

Lemma 3.59. Die direkte ILU-Zerlegung der Matrix $A_{ff}^{(e)}$ existiert für alle $0 \leq \kappa < \kappa_s$.

Beweis. Aus $1 - 4\hat{\kappa}^2 = 0$ folgt $\hat{\kappa} = 0.5$, also $1 - 4\kappa^2 = 2\kappa^2$. Letzteres gilt, falls $\kappa = 1/\sqrt{6} \approx 0.4$ ist. Es gilt aber stets $0 \leq \kappa < \kappa_s \ll 0.4$. \square

Ohne Mühe können wir nun auch die entsprechende direkte unvollständige LDU-Zerlegung angeben:

Lemma 3.60. Die Matrizen L , D und U seien definiert durch

$$L = \begin{pmatrix} I & 0 \\ -\hat{\kappa}\hat{D}_{eo}^{(e)} & I \end{pmatrix}, \quad D = \begin{pmatrix} I & 0 \\ 0 & (1 - 4\hat{\kappa}^2)I \end{pmatrix}, \quad U = \begin{pmatrix} I & -\hat{\kappa}\hat{D}_{oe}^{(e)} \\ 0 & I \end{pmatrix}.$$

Dann ist $LDU - R$ mit R aus Lemma 3.58 die direkte unvollständige LDU-Zerlegung der Matrix $(1 - 4\kappa^2)^{-1}A_{ff}^{(e)}$ zum Muster der Nichtnullelemente.

Beweis. Offensichtlich. \square

Spektrum der ILU-präkonditionierten Matrix $A_{ff}^{(e)}$

Satz 3.61. Mit den Bezeichnungen aus Lemma 3.60 gilt

$$\sigma\left(\left((1 - 4\kappa^2)LDU\right)^{-1}A_{ff}^{(e)}\right) = \{1\} \cup \sigma\left(I - \frac{\hat{\kappa}^2}{1 - 4\hat{\kappa}^2}\left(\hat{D}_{eo}^{(e)}\hat{D}_{oe}^{(e)} - 4I\right)\right).$$

Beweis. Es gilt

$$\begin{aligned} \sigma\left(\left((1 - 4\kappa^2)LDU\right)^{-1}A_{ff}^{(e)}\right) &= \sigma\left(I - (LDU)^{-1}R\right) \\ &= \sigma\left(I - U^{-1}D^{-1}L^{-1}R\right) = \sigma\left(I - D^{-1}L^{-1}RU^{-1}\right). \end{aligned}$$

Da aber $L^{-1}RU^{-1} = R$ ist, folgt

$$\sigma\left(\left((1 - 4\kappa^2)LDU\right)^{-1}A_{ff}^{(e)}\right) = \sigma\left(I - D^{-1}R\right) = \sigma\left(I - \frac{1}{1 - 4\hat{\kappa}^2}R\right).$$

Insgesamt erhalten wir das gewünschte Resultat. \square

Vergleich der Präkonditionierer

Nachdem wir nun zwei verschiedene Arten der Präkonditionierung zur Konstruktion einer approximativen Inversen der Matrix $A_{ff}^{(e)}$ untersucht haben, stellen wir diese vergleichend gegenüber:

Satz 3.62. Es sei $(1 - 4\kappa^2)LU$ der durch die direkte ILU-Zerlegung von $A_{ff}^{(e)}$ zum Muster der Nichtnullelemente induzierte ILU-Präkonditionierer und $\left(\tilde{A}_{ff}^{(e)}\right)^{-1}$ der Jacobi-Präkonditionierer aus (3.21) mit $\omega = 1$. Dann gilt

$$\sigma\left(\left((1 - 4\kappa^2)LU\right)^{-1}A_{ff}^{(e)}\right) = \{1\} \cup \frac{1}{1 - 4\hat{\kappa}^2} \sigma\left(\left(\tilde{A}_{ff}^{(e)}\right)^{-1}A_{ff}^{(e)}\right). \quad (3.23)$$

Beweis. Satz 3.57 und Satz 3.61. \square

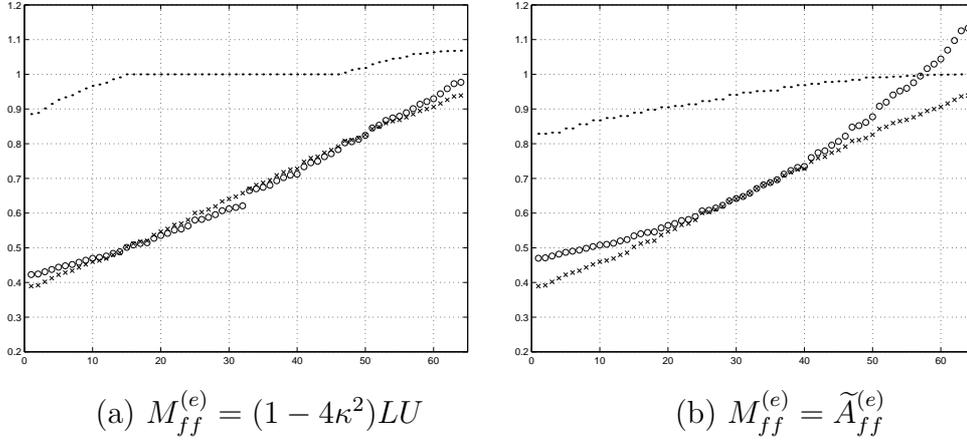


Abbildung 3.14: Spektrum der Matrix $A_{ff}^{(e)}$ (\times), des Prädiktionierers $M_{ff}^{(e)}$ (\circ) sowie der präkonditionierten Matrix $(M_{ff}^{(e)})^{-1} A_{ff}^{(e)}$ (\cdot).

Abbildung 3.14 zeigt eine typische Eigenwertverteilung der beteiligten Matrizen.

Offensichtlich ist der ILU-Prädiktionierer dem Jacobi-Prädiktionierer aufgrund der günstigeren Eigenwertverteilung vorzuziehen. Wir zeigen nun, daß auch die Effizienz des ILU-Prädiktionierers durch eine Relaxation nicht verbessert werden kann, vgl. Lemma 3.55.

Relaxierte ILU-Prädiktionierung

Ausgehend von der direkten ILU-Zerlegung ist eine Verallgemeinerung hin zu relaxierten ILU-Zerlegungen einfach.

Lemma 3.63. Ist $W^{(e)}$ die Diagonalmatrix mit Diagonaleinträgen

$$W_{ii}^{(e)} = \omega \hat{\kappa}^2 \sum_j \left(\hat{D}_{eo}^{(e)} \hat{D}_{oe}^{(e)} - 4I \right)_{ij},$$

so ist die relaxierte ILU-Zerlegung $LU - R$ der Matrix $(1 - 4\kappa^2)^{-1} A_{ff}^{(e)}$ zum Muster der Nichtnullelemente gegeben durch

$$L = \begin{pmatrix} I & 0 \\ -\hat{\kappa} \hat{D}_{eo}^{(e)} & I \end{pmatrix}, \quad U = \begin{pmatrix} I & -\hat{\kappa} \hat{D}_{oe}^{(e)} \\ 0 & (1 - 4\hat{\kappa}^2)I + W^{(e)} \end{pmatrix}.$$

Beweis. Als Restmatrix R erhalten wir

$$R = \begin{pmatrix} 0 & 0 \\ 0 & W^{(e)} + \hat{\kappa}^2 \left(\hat{D}_{eo}^{(e)} \hat{D}_{oe}^{(e)} - 4I \right) \end{pmatrix}.$$

Per Definition von $W^{(e)}$ ist klar, daß die Gleichung

$$R_{ii} = \omega \sum_{j \neq i} R_{ij}, \quad i = 1, \dots, n$$

erfüllt ist. □

Durch eine Argumentation ähnlich zu der des Beweises von Satz 3.61 beweisen wir:

Lemma 3.64. Sei $A_{ff}^{(e)} = LDU - R$ eine Zerlegung der Matrix $(1 - 4\kappa^2)^{-1} A_{ff}^{(e)}$ mit

$$L = \begin{pmatrix} I & 0 \\ -\hat{\kappa} \hat{D}_{eo}^{(e)} & I \end{pmatrix}, \quad D = \begin{pmatrix} I & 0 \\ 0 & M^{(e)} \end{pmatrix}, \quad U = \begin{pmatrix} I & -\hat{\kappa} \hat{D}_{oe}^{(e)} \\ 0 & I \end{pmatrix}.$$

Dann gilt

$$\sigma \left(((1 - 4\kappa^2) LDU)^{-1} A_{ff}^{(e)} \right) = \{1\} \cup \\ (1 - 4\hat{\kappa}^2) \sigma \left((M^{(e)})^{-1} \left(I - \frac{\hat{\kappa}^2}{1 - 4\hat{\kappa}^2} \left(\hat{D}_{eo}^{(e)} \hat{D}_{oe}^{(e)} - 4I \right) \right) \right).$$

Beweis. Analog zum Beweis von Satz 3.61. □

Mit den Bezeichnungen aus Lemma 3.64 erhalten wir für

$$M^{(e)} = (1 - 4\hat{\kappa}^2)I + W^{(e)}$$

die relaxierte ILU-Zerlegung der Matrix $(1 - 4\kappa^2)^{-1} A_{ff}^{(e)}$.

Für eine kalte Konfiguration ist $W^{(e)} = 12\omega\hat{\kappa}^2 I$ konstant, so daß eine relaxierte ILU-Präkonditionierung gegenüber der nicht-relaxierten Variante zu keiner Verbesserung der Kondition führt. Das gleiche Resultat erhalten wir aufgrund numerischer Experimente für eine heiße Konfiguration, vgl. Abb. 3.15.

Wir wenden uns jetzt der Approximation des exakten Schurkomplements zu:

Das exakte Schurkomplement

Lemma 3.65. Es gilt

$$S_{A^{(e)}} = I - \kappa^2 D_{L_2, L_1} \left(I - \kappa^2 D_{L_1, L_0} D_{L_0, L_1} \right)^{-1} D_{L_1, L_2}.$$

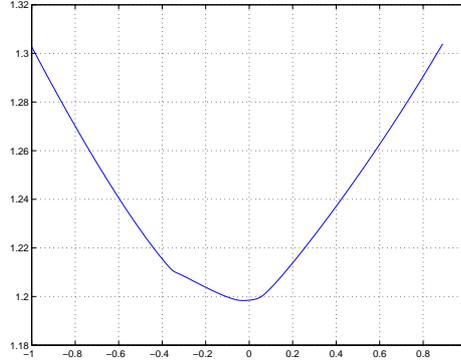


Abbildung 3.15: Die relaxierte ILU-Präkonditionierung der Matrix $A_{ff}^{(e)}$. Für eine heiße Konfiguration ($N = 16$) wurde die Kondition der präkonditionierten Matrix $A_{ff}^{(e)}$ in Abhängigkeit von ω bestimmt.

Beweis. Anfänglich ist

$$S_{A^{(e)}} = I - \kappa^2 D_{L_2, L_1} \left(I + \kappa^2 D_{L_1, L_0} (I - \kappa^2 D_{L_0, L_1} D_{L_1, L_0})^{-1} D_{L_0, L_1} \right) D_{L_1, L_2}.$$

Aus $\rho(\kappa^2 D_{L_0, L_1} D_{L_1, L_0}) \leq 8\kappa^2 < 1$ für alle $\kappa < \kappa_c$ folgt

$$\begin{aligned} S_{A^{(e)}} &= I - \kappa^2 D_{L_2, L_1} \left(I + \kappa^2 D_{L_1, L_0} \sum_{\nu=0}^{\infty} (\kappa^2 D_{L_0, L_1} D_{L_1, L_0})^{\nu} D_{L_0, L_1} \right) D_{L_1, L_2} \\ &= I - \kappa^2 D_{L_2, L_1} \left(I + \kappa^2 \sum_{\nu=1}^{\infty} (\kappa^2 D_{L_1, L_0} D_{L_0, L_1})^{\nu} \right) D_{L_1, L_2} \\ &= I - \kappa^2 D_{L_2, L_1} \sum_{\nu=0}^{\infty} (\kappa^2 D_{L_1, L_0} D_{L_0, L_1})^{\nu} D_{L_1, L_2}, \end{aligned}$$

also $S_{A^{(e)}} = I - \kappa^2 D_{L_2, L_1} (I - \kappa^2 D_{L_1, L_0} D_{L_0, L_1})^{-1} D_{L_1, L_2}$. \square

Approximation des Schurkomplements

Ein Vergleich des exakten Schurkomplements der odd-even reduzierten Matrix $A^{(e)}$ mit dem der originären Matrix A zeigt, daß diese identisch sind, vgl. Lemma 3.23. Insofern ist die Konstruktion einer geeigneten Approximation des Schurkomplements eine bereits bekannte Aufgabe, siehe Seite 66f. Im Gegensatz zum originären Problem kommt es jedoch im Falle des odd-even reduzierten Systems bei der Approximation des Schurkomplements zu keiner Auffüllung der Grobmatrix. Somit ist eine rekursive Vorgehensweise ohne Probleme möglich.

Restriktion und Prolongation

Zur Konstruktion von matrixabhängiger Restriktion und Prolongation können einerseits die bereits hergeleiteten approximativen Inversen der Matrix $A_{ff}^{(e)}$ verwendet werden. Andererseits ist es denkbar, aufgrund der Identität

$$\begin{aligned} \begin{pmatrix} -A_{cf}A_{ff}^{-1} & I \end{pmatrix} &= \begin{pmatrix} \kappa^2 D_{L_2,L_1} D_{L_1,L_0} (I - \kappa^2 D_{L_0,L_1} D_{L_1,L_0})^{-1} & I \end{pmatrix} \\ &= \begin{pmatrix} \kappa^2 D_{L_2,L_1} (I - \kappa^2 D_{L_1,L_0} D_{L_0,L_1})^{-1} D_{L_1,L_0} & I \end{pmatrix} \end{aligned}$$

bzw.

$$\begin{aligned} \begin{pmatrix} -A_{ff}^{-1}A_{fc} \\ I \end{pmatrix} &= \begin{pmatrix} \kappa^2 (I - \kappa^2 D_{L_0,L_1} D_{L_1,L_0})^{-1} D_{L_0,L_1} D_{L_1,L_2} \\ I \end{pmatrix} \\ &= \begin{pmatrix} \kappa^2 D_{L_0,L_1} (I - \kappa^2 D_{L_1,L_0} D_{L_0,L_1})^{-1} D_{L_1,L_2} \\ I \end{pmatrix} \end{aligned}$$

eine geeignete approximative Inverse der Matrix $I - \kappa^2 D_{L_1,L_0} D_{L_0,L_1}$ zu verwenden, vgl. (3.15) und die anschließende Diskussion der Parameterwahl.

Aufwand

Um den durch die Präkonditionierung entstehenden zusätzlichen Rechenaufwand pro Iterationsschritt zu bewerten, bestimmen wir den Aufwand für eine Matrix-Vektor-Multiplikation (MVM) mit dem Schurkomplement-Präkonditionierer im Vergleich zu dem Aufwand einer MVM mit der odd-even reduzierten Matrix. Als Maß k verwenden wir dabei die Anzahl der Kopplungen, die zu berechnen sind:

- a) Der Aufwand einer MVM mit D_{L_i,L_j} ist $k(D_{L_i,L_j}) = n$.
- b) Eine MVM mit D_{eo} (D_{oe}) besteht aus einer MVM mit D_{L_0,L_1} (D_{L_1,L_0}) und einer mit D_{L_2,L_1} (D_{L_1,L_2}), also gilt $k(D_{eo}) = k(D_{oe}) = 2n$.
- c) Sowohl für eine MVM mit A als auch für eine mit der odd-even reduzierten Matrix $A^{(e)}$ folgt $k(A) = 4n$ respektive $k(A^{(e)}) = 4n$.
- d) Der Aufwand einer MVM mit beiden vorgestellten approximativen Inversen (die Matrix aus (3.21) und die durch die direkte ILU-Zerlegung induzierte aus Satz 3.61) der Matrix $A_{ff}^{(e)}$ und der Aufwand einer MVM mit $A_{ff}^{(e)}$ ist gleich. Bei einer naiven Realisierung der MVM beträgt dieser $k(D_{L_0,L_1}) + k(D_{L_1,L_0}) = 2n$. Da die Matrix $A_{ff}^{(e)}$ eine Nächste-Nachbar-Kopplung auf den feinen Gitterpunkten beschreibt, kann der Aufwand jedoch in einfacher Weise halbiert werden, $k(A_{ff}^{(e)}) = n$.

- e) Wegen $k(D_{L_2,L_1}D_{L_2,L_0}) = k(D_{L_0,L_1}D_{L_1,L_2}) = 2n$ beträgt der Aufwand für die matrixabhängige Restriktion und Prolongation mit der approximativen Inversen aus (3.21) bzw. der durch die direkte ILU-Zerlegung induzierten approximativen Inversen, s. Satz 3.61, insgesamt jeweils $3n$. Werden die Kopplungen $D_{L_2,L_1}D_{L_2,L_0}$ und $D_{L_0,L_1}D_{L_1,L_2}$ in einer Vorverarbeitungsphase explizit bestimmt, reduziert sich der Aufwand auf $k(D_{L_2,L_1}D_{L_2,L_0}) = k(D_{L_0,L_1}D_{L_1,L_2}) = n$. Diese Reduktion bedeutet für die Restriktion und Prolongation jeweils einen Aufwand von $2n$.

Vernachlässigen wir den Aufwand zur Lösung des Grobgittersystems, so sind der Aufwand für eine MVM mit dem Schurkomplement-Präkonditionierer und der für eine MVM mit der Matrix A bei einer geschickten Implementierung und einer Vorverarbeitung zur Berechnung der effektiven Kopplungen identisch.

3.5 Numerische Ergebnisse

Die numerischen Ergebnisse wurden mit Hilfe von MATLAB Version 5.3 erzeugt [1]. Für verschiedene Gittergrößen $N = 16$, $N = 32$ und $N = 64$ wurden jeweils 5 heiße Konfigurationen zufällig generiert und anschließend die odd-even reduzierte Matrix $A^{(e)}$ initialisiert. Berechnet wurde

- a) Die Kondition der Matrix $A_{ff}^{(e)}$ und der durch die Jacobi- bzw. der ILU-Zerlegung induzierten approximativen Inversen, vgl. Satz 3.62.
- b) Die Kondition des exakten Schurkomplements $S_A^{(e)}$ und die des durch (3.16) präkonditionierten Schurkomplements mit ω_1 aus (3.18). Der Wert ω_2 wurde so gewählt, daß die Kondition des präkonditionierten Schurkomplements minimal ist.
- c) Die Kondition des odd-even reduzierten Systems $A^{(e)}$ sowie des Schurkomplement-präkonditionierten odd-even reduzierten Systems. Dabei stehen in der ersten Zeile die erzielten Resultate für die ILU-präkonditionierte Matrix $A_{ff}^{(e)}$ und in der zweiten Zeile die Resultate für die Jacobi-Präkonditionierung.
- d) Die asymptotische Konvergenzgeschwindigkeit

$$A_{CG} \propto \frac{1}{\sqrt{\kappa(A)}}$$

des CG-Verfahrens für das odd-even reduzierten Systems $A^{(e)}$ und das Schurkomplement-präkonditionierte odd-even reduzierte System. Wie

zuvor stehen in der ersten Zeile die erzielten Resultate für die ILU-präkonditionierte Matrix $A_{ff}^{(e)}$ und in der zweiten Zeile die Resultate für die Jacobi-Präkonditionierung.

In den Tabellen 3.2, 3.3 und 3.4 sind die einzelnen numerischen Ergebnisse, in Tabelle 3.5 die Mittelwerte und Standardabweichungen dargestellt. Dabei wurde der Parameter κ so gewählt, daß gilt

$$m = \frac{1}{2} \left(\frac{1}{\kappa} - \frac{1}{\kappa_c} \right) = 0.01. \quad (3.24)$$

Eine Auswertung der numerischen Ergebnisse ergibt eine um einen Faktor 5–6 höhere asymptotische Konvergenzgeschwindigkeit des CG-Verfahrens für das Schurkomplement-präkonditionierte odd-even reduzierte System im Vergleich zum unpräkonditionierten System. Dies entspricht einer Verbesserung der asymptotischen Konvergenzgeschwindigkeit des CG-Verfahrens für das originäre System um einen Faktor 10 – 12. Die asymptotische Konvergenzgeschwindigkeit des CG-Verfahrens ist für die durch die Jacobi-Zerlegung induzierte approximative Inverse erwartungsgemäß etwas schlechter als die durch die ILU-Zerlegung induzierte, und das optimale konfigurationsabhängige Gewicht variiert geringfügig.

Des weiteren wurden numerische Messungen durchgeführt für die durch die ILU-Zerlegung induzierte approximative Inverse bei

$$\kappa(A) = \left(1 + \frac{\kappa}{\kappa_c} \right) / \left(1 - \frac{\kappa}{\kappa_c} \right) = N^2 - 1. \quad (3.25)$$

Die Ergebnisse zeigen mit wachsender Gitterdimension N eine Verbesserung der asymptotischen Konvergenzgeschwindigkeit des CG-Verfahrens für das Schurkomplement-präkonditionierte odd-even reduzierte System im Vergleich zur asymptotischen Konvergenzgeschwindigkeit des CG-Verfahrens für das unpräkonditionierte System, vgl. Tab. 3.6, Tab. 3.7 und Tab. 3.8. Diese Aussage ist auch für eine kalte Konfiguration richtig, vgl. Tab. 3.9.

Konfiguration	conf16a	conf16b	conf16c	conf16d	conf16e
$\kappa(A_{ff}^{(e)})$	2.3878	2.5721	2.5451	2.3530	2.3001
$\kappa(\left(\tilde{A}_{ff}^{(e)}\right)^{-1} A_{ff}^{(e)})$	1.2015	1.2401	1.2342	1.1945	1.1837
$\kappa(S_{A^{(e)}})$	40.5566	40.7711	41.1354	43.1539	39.1483
ω_2	1.62	1.66	1.72	1.56	1.7
$\kappa(\tilde{S}_{A^{(e)}}^{-1} S_{A^{(e)}})$	2.8927	3.2617	2.7714	2.4923	3.6004
$\kappa(A^{(e)})$	86.9595	86.2654	86.3168	87.5484	87.4731
$\kappa(M^{-1}A^{(e)})$	3.2356	3.6685	3.0878	2.7547	4.0140
	4.1594	4.5823	3.9828	3.4627	5.0927
$A_{cg}(A^{(e)})$	0.1072	0.1076	0.1076	0.1068	0.1069
$A_{cg}(M^{-1}A^{(e)})$	0.5559	0.5221	0.5690	0.6025	0.4991
	0.4903	0.4671	0.5010	0.5374	0.4431

Tabelle 3.2: Numerische Ergebnisse für die Schurkomplement-präkonditionierte odd-even reduzierte Laplace Matrix auf Ω_h mit $N = 16$ und $m = 0.01$.

Konfiguration	conf32a	conf32b	conf32c	conf32d	conf32e
$\kappa(A_{ff}^{(e)})$	2.5006	2.5084	2.4812	2.4284	2.3956
$\kappa(\left(\tilde{A}_{ff}^{(e)}\right)^{-1} A_{ff}^{(e)})$	1.2251	1.2268	1.2210	1.2100	1.2033
$\kappa(S_{A^{(e)}})$	42.0053	40.3408	40.3210	41.9699	41.5442
ω_2	1.68	1.48	1.58	1.62	1.7
$\kappa(\tilde{S}_{A^{(e)}}^{-1} S_{A^{(e)}})$	3.0404	4.5696	3.8590	2.9750	2.4077
$\kappa(A^{(e)})$	87.3258	86.8460	86.7904	87.4749	87.9807
$\kappa(M^{-1}A^{(e)})$	3.3807	4.9403	4.2127	3.3027	2.7025
	4.3318	5.6930	5.2370	4.2042	3.4580
$A_{cg}(A^{(e)})$	0.1070	0.1073	0.1073	0.1069	0.1066
$A_{cg}(M^{-1}A^{(e)})$	0.5438	0.4499	0.4872	0.5502	0.6083
	0.4804	0.4191	0.4369	0.4877	0.5377

Tabelle 3.3: Numerische Ergebnisse für die Schurkomplement-präkonditionierte odd-even reduzierte Laplace Matrix auf Ω_h mit $N = 32$ und $m = 0.01$.

Konfiguration	conf64a	conf64b	conf64c	conf64d	conf64e
$\kappa(A_{ff}^{(e)})$	2.4301	2.3442	2.3983	2.4452	2.4473
$\kappa\left(\left(\tilde{A}_{ff}^{(e)}\right)^{-1} A_{ff}^{(e)}\right)$	1.2104	1.1927	1.2038	1.2135	1.2140
$\kappa(S_{A^{(e)}})$	43.6641	42.5844	42.3731	41.2609	41.349
ω_2	1.67	1.7	1.69	1.73	1.66
$\kappa(\tilde{S}_{A^{(e)}}^{-1} S_{A^{(e)}})$	2.4386	2.4325	1.9482	2.7938	3.3437
$\kappa(A^{(e)})$	88.0057	88.8251	88.4891	87.6063	87.5839
$\kappa(M^{-1}A^{(e)})$	2.7427	2.8249	2.2100	3.2510	3.7214
	3.5587	3.8357	3.1346	4.6097	4.8103
$A_{cg}(A^{(e)})$	0.1066	0.1061	0.1063	0.1068	0.1068
$A_{cg}(M^{-1}A^{(e)})$	0.6038	0.5949	0.6726	0.5546	0.5183
	0.5300	0.5105	0.5648	0.4657	0.4559

Tabelle 3.4: Numerische Ergebnisse für die Schurkomplement-präkonditionierte odd-even reduzierte Laplace Matrix auf Ω_h mit $N = 64$ und $m = 0.01$.

Konfiguration	conf16	conf32	conf64
$\kappa(A_{ff}^{(e)})$	2.4316 (0.1204)	2.4628 (0.0489)	2.4130 (0.0432)
$\kappa\left(\left(\tilde{A}_{ff}^{(e)}\right)^{-1} A_{ff}^{(e)}\right)$	1.2108 (0.0250)	1.2172 (0.0102)	1.2069 (0.0089)
$\kappa(S_{A^{(e)}})$	40.9531 (1.4426)	41.2362 (0.8462)	42.2463 (0.9895)
ω_2	1.6520 (0.0642)	1.6120 (0.0879)	1.6900 (0.0274)
$\kappa(\tilde{S}_{A^{(e)}}^{-1} S_{A^{(e)}})$	3.0037 (0.4331)	3.3703 (0.8469)	2.5914 (0.5170)
$\kappa(A^{(e)})$	86.9126 (0.6112)	87.2836 (0.4897)	88.1020 (0.5468)
$\kappa(M^{-1}A^{(e)})$	3.3521 (0.4946)	3.7078 (0.8740)	2.9500 (0.5683)
	4.2560 (0.6164)	4.5848 (0.8849)	3.9898 (0.7068)
$A_{cg}(A^{(e)})$	0.1073 (0.0004)	0.1070 (0.0003)	0.1065 (0.0003)
$A_{cg}(M^{-1}A^{(e)})$	0.5498 (0.0404)	0.5279 (0.0611)	0.5889 (0.0579)
	0.4878 (0.0356)	0.4724 (0.0465)	0.5054 (0.0452)

Tabelle 3.5: Mittelwert und Standardabweichung für verschiedene Gittergrößen.

Konfiguration	conf16a	conf16b	conf16c	conf16d	conf16e
$\kappa(A_{ff}^{(e)})$	2.3734	2.5552	2.5285	2.3389	2.2868
$\kappa\left(\left(\tilde{A}_{ff}^{(e)}\right)^{-1} A_{ff}^{(e)}\right)$	1.1985	1.2365	1.2307	1.1916	1.1810
$\kappa(S_{A^{(e)}})$	30.1218	30.5336	30.7768	31.8265	28.9235
ω_2	1.61	1.65	1.70	1.56	1.68
$\kappa(\tilde{S}_{A^{(e)}}^{-1} S_{A^{(e)}})$	2.4709	2.7579	2.4206	2.1671	2.9518
$\kappa(A^{(e)})$	64.2488	64.2494	64.2496	64.2474	64.2501
$\kappa(M^{-1}A^{(e)})$	2.7318	3.0708	2.6663	2.3802	3.2521
$A_{cg}(A)$	0.06262	0.06262	0.06262	0.06262	0.06262
$A_{cg}(A^{(e)})$	0.12476	0.12476	0.12476	0.12476	0.12476
$A_{cg}(M^{-1}A^{(e)})$	0.60502	0.57065	0.61242	0.64818	0.55452

Tabelle 3.6: Numerische Ergebnisse für die Schurkomplement-präkonditionierte odd-even reduzierte Laplace Matrix auf Ω_h mit $N = 16$ und $\kappa(A) = 255$.

Konfiguration	conf32a	conf32b	conf32c	conf32d	conf32e
$\kappa(A_{ff}^{(e)})$	2.5309	2.5394	2.5114	2.4565	2.4224
$\kappa\left(\left(\tilde{A}_{ff}^{(e)}\right)^{-1} A_{ff}^{(e)}\right)$	1.2315	1.2333	1.2274	1.2159	1.2088
$\kappa(S_{A^{(e)}})$	122.0795	117.7743	117.8172	121.7541	119.8830
ω_2	1.69	1.46	1.57	1.64	1.71
$\kappa(\tilde{S}_{A^{(e)}}^{-1} S_{A^{(e)}})$	6.3782	11.6993	9.5515	6.4028	3.3336
$\kappa(A^{(e)})$	256.2501	256.2501	256.2500	256.2501	256.2501
$\kappa(M^{-1}A^{(e)})$	7.2825	12.6838	10.4471	7.2848	4.1384
$A_{cg}(A)$	0.03126	0.03126	0.03126	0.03126	0.03126
$A_{cg}(A^{(e)})$	0.06246	0.06246	0.06246	0.06246	0.06246
$A_{cg}(M^{-1}A^{(e)})$	0.37056	0.28079	0.30939	0.37050	0.49157

Tabelle 3.7: Numerische Ergebnisse für die Schurkomplement-präkonditionierte odd-even reduzierte Laplace Matrix auf Ω_h mit $N = 32$ und $\kappa(A) = 1023$.

Konfiguration	conf64a	conf64b	conf64c	conf64d	conf64e
$\kappa(A_{ff}^{(e)})$	2.4689	2.3790	2.4354	2.485	2.4871
$\kappa\left(\begin{pmatrix} \tilde{A}_{ff}^{(e)} & \\ & A_{ff}^{(e)} \end{pmatrix}^{-1}\right)$	1.2185	1.1998	1.2115	1.2218	1.2223
$\kappa(S_{A^{(e)}})$	501.9592	484.8792	484.4884	475.9954	476.7222
ω_2	1.69	1.74	1.72	1.70	1.65
$\kappa(\tilde{S}_{A^{(e)}}^{-1} S_{A^{(e)}})$	12.6197	13.5279	2.4807	13.3761	27.0091
$\kappa(A^{(e)})$	1024.248	1024.249	1024.249	1024.249	1024.249
$\kappa(M^{-1}A^{(e)})$	15.2301	18.0533	3.2643	17.5445	30.8162
$A_{cg}(A)$	0.01562	0.01562	0.01562	0.01562	0.01562
$A_{cg}(A^{(e)})$	0.03124	0.03124	0.03124	0.03124	0.03124
$A_{cg}(M^{-1}A^{(e)})$	0.25624	0.23535	0.55348	0.23874	0.18014

Tabelle 3.8: Numerische Ergebnisse für die Schurkomplement-präkonditionierte odd-even reduzierte Laplace Matrix auf Ω_h mit $N = 64$ und $\kappa(A) = 4095$.

Konfiguration	conf16	conf32	conf64
$\kappa(A_{ff}^{(e)})$	1.9693	1.9922	1.998
$\kappa\left(\begin{pmatrix} \tilde{A}_{ff}^{(e)} & \\ & A_{ff}^{(e)} \end{pmatrix}^{-1}\right)$	1.1193	1.1235	1.1246
$\kappa(S_{A^{(e)}})$	32.6255	128.6251	512.6250
ω_2	1.65	1.67	1.67
$\kappa(\tilde{S}_{A^{(e)}}^{-1} S_{A^{(e)}})$	1.2091	1.3205	1.6476
$\kappa(A^{(e)})$	64.2510	256.2502	1024.2501
$\kappa(M^{-1}A^{(e)})$	1.7357	3.4133	11.7903
$A_{cg}(A)$	0.06262	0.03126	0.01562
$A_{cg}(A^{(e)})$	0.12476	0.06246	0.03124
$A_{cg}(M^{-1}A^{(e)})$	0.75905	0.54127	0.29123

Tabelle 3.9: Numerische Ergebnisse für die Schurkomplement-präkonditionierte odd-even reduzierte Laplace Matrix auf Ω_h mit $N = 16$, $N = 32$, $N = 64$ für eine kalte Konfiguration und $\kappa(A) = N^2 - 1$.

3.6 Zusammenfassung

Für eine kalte Konfiguration ist die konstante Gitterfunktion ein Eigenvektor zum kleinsten Eigenwert der Laplace Matrix. Daher wurden matrixabhängig lokal wirkende Restriktionen und Prolongationen angegeben, welche konstante Gitterfunktionen bis auf Skalierung exakt von einem Gitter auf das andere transformieren. Weiter wurden zwei Schurkomplement-Ansätze vorgestellt und gezeigt, daß bei geeigneter Parameterwahl die Wirkung der Schurkomplement-Approximation auf konstante Gitterfunktionen mit der des exakten Schurkomplements übereinstimmt. Erwartungsgemäß fällt die Kondition des präkonditionierten Schurkomplements bei einer 9-Punkte-Approximation gegenüber einer 5-Punkte-Approximation deutlich günstiger aus.

Für eine heiße Konfiguration wurde gezeigt, daß bei einem Standard-Grobgitter und einer entsprechenden Partition der Laplace Matrix Aussagen über die Spektren der beteiligten Teilmatrizen A_{ff} und A_{cc} getroffen werden können. Insbesondere führte dies zu einer oberen Schranke der Kondition der Matrix A_{ff} , so daß diese Matrix für alle $0 \leq \kappa < \kappa_c$ gut konditioniert ist. Diese Beobachtung rechtfertigte eine Approximation der Inversen von A_{ff} durch eine lokal wirkende Matrix. Genauer untersucht wurden ein Jacobi- und ein ILU-Präkonditionierer. Für die präkonditionierte Matrix A_{ff} wurde jeweils die Eigenwertverteilung bestimmt und eine obere Schranke für die Kondition angegeben. Ein Vergleich der Spektren offenbarte die Überlegenheit des ILU- gegenüber dem Jacobi-Präkonditionierer. Da zudem auch bei der matrixabhängigen Restriktion und Prolongation bei gleichem Aufwand mehr Gitterpunkte berücksichtigt werden, ist die ILU- der Jacobi-Präkonditionierung vorzuziehen. Eine relaxierte ILU-Zerlegung bringt keine weitere Verbesserung. Es wurde des weiteren motiviert, warum es sich nicht lohnt, aufwendigere Präkonditionierer der Matrix A_{ff} zu untersuchen.

Die verwendeten Schurkomplement-Approximationen basieren auf einer approximativen Inversen eines gut konditionierten 7-Punkte-Sterns. Eine konfigurationsabhängige Gewichtung des Nichtdiagonalanteils des 7-Punkte-Sterns führt zu einer wesentlichen Verbesserung der Kondition des präkonditionierten Schurkomplements.

Es wurde festgestellt, daß die odd-even reduzierte Laplace Matrix im Gegensatz zur originären Matrix eine in rein algebraischen Multilevel-Verfahren übliche Charakterisierung der Kopplungen erlaubt, welche zu a priori bekannten und einfach zu konstruierenden Grobgittern führt. Ebenso wie für die originäre Matrix wurde auch für die odd-even reduzierte Laplace Matrix

in der resultierenden F/C-Anordnung der Gitterpunkte die Kondition der (nicht-)präkonditionierten Teilmatrix $A_{ff}^{(e)}$ ausführlich diskutiert und jeweils obere Schranken ermittelt. Da das resultierende Schurkomplement identisch ist mit dem im Falle der originären Laplace Matrix, entfiel eine gesonderte Betrachtung der Schurkomplement-Approximation.

Die Präkonditionierung des odd-even reduzierten Systems ist der des originären Systems vorzuziehen: Zum einen führt allein der Übergang vom originären System zum odd-even reduzierten System zu einer Verbesserung der Kondition, zum anderen ist die obere Schranke der Kondition der präkonditionierten Matrix $A_{ff}^{(e)}$ kleiner als die der präkonditionierten Matrix A_{ff} , so daß wir bei gleicher Schurkomplement-Approximation von dem Präkonditionierer der odd-even reduzierten Matrix bessere Approximationseigenschaften erwarten.

Die numerischen Resultate offenbaren eine um einen Faktor 5 – 6 höhere asymptotische Konvergenzgeschwindigkeit des CG-Verfahrens für das Schurkomplement-präkonditionierte odd-even reduzierte System im Vergleich zum unpräkonditionierten System. Dies entspricht einer Verbesserung der asymptotischen Konvergenzgeschwindigkeit des CG-Verfahrens für das originäre System um einen Faktor 10 – 12. Die asymptotische Konvergenzgeschwindigkeit des CG-Verfahrens ist für die durch die Jacobi-Zerlegung induzierte approximative Inverse erwartungsgemäß etwas schlechter als die durch die ILU-Zerlegung induzierte, und das optimale konfigurationsabhängige Gewicht variiert geringfügig.

Mit proportional zur Gitterdimension N wachsender Kondition zeigen die numerischen Ergebnisse sowohl für eine heiße als auch für eine kalte Konfiguration eine Verbesserung der asymptotischen Konvergenzgeschwindigkeit des CG-Verfahrens für das Schurkomplement-präkonditionierte odd-even reduzierte System im Vergleich zu der des unpräkonditionierten Systems.

Kapitel 4

Das Schwinger-Modell

Das Schwinger Modell ist ein einfaches physikalisches Modell, das bereits einige interessante, ebenfalls in der Quantenchromodynamik zu beobachtende Phänomene aufweist. Nach einer kurzen Einführung des Schwinger Modells und der Darstellung wichtiger elementarer Eigenschaften der Schwinger Matrix werden wir zeigen, daß sich die in Kapitel 3 hergeleiteten Schurkomplement-Verfahren erfolgreich zur Lösung des Schwinger Modells adaptieren lassen.

4.1 Einführung

Das Schwinger Modell repräsentiert wiederum eine Nächste-Nachbar-Kopplung auf einem zweidimensionalen Gitter. Im Gegensatz zu dem in Kapitel 3 vorgestellten Modellproblem existieren jedoch zwei Freiheitsgrade auf jedem Gitterpunkt.

Um das Schwinger Modell zu formulieren, werden einige Definitionen benötigt:

Definition 4.1. Die Matrizen $\gamma_i \in \mathbb{C}^{2 \times 2}$ für $i = 1, 2$ sind wie folgt definiert:

$$\gamma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \gamma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}.$$

Definition 4.2. Für $A = (a_{ij}) \in \mathbb{C}^{m \times n}$ und $B = (b_{ij}) \in \mathbb{C}^{p \times q}$ ist das Kroneckerprodukt $A \otimes B \in \mathbb{C}^{mp \times nq}$ definiert durch

$$A \otimes B = \begin{pmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ \vdots & & & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{pmatrix}.$$

Lemma 4.3. Es gilt $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$, wann immer die auftretenden Produkte definiert sind.

Beweis. Nachrechnen. □

Lemma 4.4. Es seien $A \in \mathbb{C}^{m \times n}$ und $B, C \in \mathbb{C}^{p \times q}$. Dann gilt

- a) $(A \otimes B)^H = A^H \otimes B^H$
- b) $A \otimes (B + C) = (A \otimes B) + (A \otimes C)$.

Beweis. Nachrechnen. □

So wie in Kapitel 3 definieren wir Konfigurationen und unterscheiden diese:

Definition 4.5. Eine Konfiguration $U = \{U_\mu(x)\}$ ist eine Sammlung von $2n_1n_2$ unitären Zahlen $U_\mu(x)$,

$$U_\mu(x) = \exp(-2\pi i \phi(x, \mu)), \quad \phi(x, \mu) \in [0, 1),$$

wobei x Gitterpunkt und $\mu \in \{1, 2\}$ ist.

Definition 4.6. Eine Konfiguration heißt kalt, falls alle $U_\mu(x) = 1$ sind. Sie heißt heiß, falls alle $U_\mu(x)$ gleichverteilt sind. Eine Konfiguration, die weder kalt noch heiß ist, heißt realistisch.

Nun formulieren wir das Gleichungssystem des Schwinger Modells:

Definition 4.7. Das lineare Gleichungssystem des Schwinger Modells für eine gegebene Konfiguration $\{U_\mu(x)\}$ und den reellen Parameter $\kappa \geq 0$ ist gegeben durch

$$A\psi = \varphi, \quad A = I - \kappa D, \tag{4.1}$$

wobei die Matrix $D = (D_{x,y})$ gegeben ist durch ($\mu = 1, 2$)

$$D_{x,y} = ((I - \gamma_\mu) \otimes U_\mu(x)) \delta_{x,y-e_\mu} + ((I + \gamma_\mu) \otimes U_\mu^H(x - e_\mu)) \delta_{x,y+e_\mu}. \tag{4.2}$$

Ein Vergleich der Schwinger Matrix mit der Laplace Matrix offenbart, daß diese sich lediglich in der γ_μ -Struktur unterscheiden.

Definition 4.8. Der kritische Wert κ_c ist definiert als $\kappa_c = 1 / \max_{\lambda \in \sigma(D)} \operatorname{Re} \lambda$.

Physikalisch ist $\kappa \approx \kappa_c$ besonders interessant, da dies kleinen Gittermassen m_q auf physikalischen Gittern mit Gitterabstand a ,

$$m_q \propto \frac{1}{2a} \left(\frac{1}{\kappa} - \frac{1}{\kappa_c} \right), \quad (4.3)$$

entspricht, vgl. auch (3.24).

4.2 Eigenschaften

Analog zur Diskussion der Eigenschaften der Laplace Matrix halten wir elementare Eigenschaften der Schwinger Matrix fest. Dazu nehmen wir im folgenden stets an, daß

$$n_i = 2^{N_i}, \quad N_i \geq 2, \quad i = 1, 2$$

erfüllt ist.

Das Besetzungsmuster

Lemma 4.9. Die Matrix D besitzt genau vier Nichtnulleinträge pro Zeile, nämlich an den Positionen $(\mu = 1, 2)$

$$\begin{aligned} (x, x + e_\mu) &: (I - \gamma_\mu) \otimes U_\mu(x) \\ (x, x - e_\mu) &: (I + \gamma_\mu) \otimes U_\mu^H(x - e_\mu). \end{aligned}$$

Beweis. Trivial. □

Die γ_μ -Matrizen

Die γ_μ -Matrizen antikommutieren.

Lemma 4.10. Es gilt $\gamma_\mu^2 = I$ und $\gamma_\mu \gamma_\nu + \gamma_\nu \gamma_\mu = 0$ für $\mu \neq \nu$.

Beweis. Klar. □

Lemma 4.11. Es gilt $(I + \gamma_\mu)(I - \gamma_\mu) = (I - \gamma_\mu)(I + \gamma_\mu) = 0$ und $(I \pm \gamma_\mu)^2 = 2(I \pm \gamma_\mu)$.

Beweis. Folgt unmittelbar aus $\gamma_\mu^2 = I$. □

Symmetrieeigenschaften

Die Schwinger Matrix besitzt zwei wichtige Symmetrieeigenschaften. Die erste ist die bereits bekannte odd-even Symmetrie, die zweite die sogenannte γ_5 -Symmetrie.

Definition 4.12. Die Matrix $\gamma_5 \in \mathbb{C}^{2 \times 2}$ ist definiert durch

$$\gamma_5 = i\gamma_1\gamma_2 = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Lemma 4.13. Es gilt $\gamma_5 = \gamma_5^T = \gamma_5^H = \gamma_5^{-1}$.

Beweis. Offensichtlich. □

Lemma 4.14. Es gilt $\gamma_5\gamma_\mu\gamma_5 = -\gamma_\mu$ für alle $\mu = 1, 2$.

Beweis. Nachrechnen. □

Definition 4.15. Die Matrix $\Gamma_5 \in \mathbb{C}^{n \times n}$ ist definiert durch $\Gamma_5 = I_{\frac{n}{2}} \otimes \gamma_5$.

Satz 4.16. Die Matrix D ist γ_5 -symmetrisch, d.h. $D\Gamma_5 = \Gamma_5 D^H$.

Beweis. Wir zeigen, daß $\Gamma_5 D\Gamma_5 = D^H$ ist. Dazu betrachten wir die Blockzeile x der Matrix $\Gamma_5 D\Gamma_5$. Dort gibt es 4 Nichtnullblöcke der Größe 2×2 , nämlich an den Positionen ($\mu = 1, 2$)

$$\begin{aligned} (x, x - e_\mu) &: (I + \gamma_5\gamma_\mu\gamma_5) \otimes U_\mu^H(x - e_\mu) \\ (x, x + e_\mu) &: (I - \gamma_5\gamma_\mu\gamma_5) \otimes U_\mu(x). \end{aligned}$$

Entsprechend ergeben sich für die x -te Blockzeile von D^H die Blöcke ($\mu = 1, 2$)

$$\begin{aligned} (x, x - e_\mu) &: (I - \gamma_\mu) \otimes U_\mu^H(x - e_\mu) \\ (x, x + e_\mu) &: (I + \gamma_\mu) \otimes U_\mu(x). \end{aligned}$$

Da die γ_μ -Matrizen nach Lemma 4.14 antikommutieren, erhalten wir die Behauptung. □

Die γ_5 -Symmetrie der Matrix D überträgt sich auf die Schwinger Matrix A .

Satz 4.17. Die Matrix D ist i.a. nicht normal, d.h. $D^H D \neq D D^H$.

Beweis. Es sei $y = x + e_\mu + e_\nu$ mit $\mu \neq \nu$. Dann gilt

$$\begin{aligned}
(D^H D)_{x,y} &= ((I + \gamma_\mu) \otimes U_\mu(x)) ((I - \gamma_\nu) \otimes U_\nu(x + e_\mu)) \\
&\quad + ((I + \gamma_\nu) \otimes U_\nu(x)) ((I - \gamma_\mu) \otimes U_\mu(x + e_\nu)) \\
&= ((I + \gamma_\mu)(I - \gamma_\nu) \otimes U_\mu(x)U_\nu(x + e_\mu)) \\
&\quad + ((I + \gamma_\nu)(I - \gamma_\mu) \otimes U_\nu(x)U_\mu(x + e_\nu)) \\
&= ((I - \gamma_\nu + \gamma_\mu - \gamma_\mu\gamma_\nu) \otimes U_\mu(x)U_\nu(x + e_\mu)) \\
&\quad + ((I - \gamma_\mu + \gamma_\nu - \gamma_\nu\gamma_\mu) \otimes U_\nu(x)U_\mu(x + e_\nu))
\end{aligned}$$

und

$$\begin{aligned}
(DD^H)_{x,y} &= ((I - \gamma_\mu) \otimes U_\mu(x)) ((I + \gamma_\nu) \otimes U_\nu(x + e_\mu)) \\
&\quad + ((I - \gamma_\nu) \otimes U_\nu(x)) ((I + \gamma_\mu) \otimes U_\mu(x + e_\nu)) \\
&= ((I - \gamma_\mu)(I + \gamma_\nu) \otimes U_\mu(x)U_\nu(x + e_\mu)) \\
&\quad + ((I - \gamma_\nu)(I + \gamma_\mu) \otimes U_\nu(x)U_\mu(x + e_\nu)) \\
&= ((I + \gamma_\nu - \gamma_\mu - \gamma_\mu\gamma_\nu) \otimes U_\mu(x)U_\nu(x + e_\mu)) \\
&\quad + ((I + \gamma_\mu - \gamma_\nu - \gamma_\nu\gamma_\mu) \otimes U_\nu(x)U_\mu(x + e_\nu)).
\end{aligned}$$

Wegen $\gamma_\mu \neq \gamma_\nu$ liefert ein Koeffizientenvergleich $(D^H D)_{x,y} \neq (DD^H)_{x,y}$. \square

Wir schließen, daß auch die Schwinger Matrix A i.a. nicht normal ist.

Das Spektrum

Die Eigenwerte von D kommen in konjugiert komplexen Paaren vor:

Lemma 4.18. Ist v ein rechtsseitiger Eigenvektor zum Eigenwert λ der Matrix D , so ist $v^H \Gamma_5$ ein linksseitiger Eigenvektor zum Eigenwert $\bar{\lambda}$.

Beweis. Aus $\lambda v = Dv = \Gamma_5 D^H \Gamma_5 v$ folgt $(\lambda \Gamma_5 v)^H = (D^H \Gamma_5 v)^H$ und daher $\bar{\lambda} v^H \Gamma_5 = v^H \Gamma_5 D$. \square

Zusammen mit der odd-even Symmetrie ist klar, daß die Eigenwerte der Matrix D in Quadrupel $\{\lambda, -\lambda, \bar{\lambda}, -\bar{\lambda}\}$ auftreten, vgl. Lemma 3.13.

Lemma 4.19. Für $\kappa < \kappa_c$ ist die Matrix $A = I - \kappa D$ positiv reell, d.h. alle Eigenwerte liegen in der rechten Halbebene.

Beweis. Klar, s. Def. 4.8. \square

Aus numerischen Experimenten ist bekannt, daß der Spektralradius einer heißen stets kleiner ist als der einer kalten Konfiguration. Für eine kalte Konfiguration läßt sich wegen $D\mathbf{1} = 4\mathbf{1}$ leicht zeigen, daß $\rho(D) \geq 4$ ist. Tatsächlich gilt sogar $\rho(D) = 4$, was mittels einer diskreten Fouriertransformation auch bewiesen werden kann. Abbildung 4.1 zeigt eine typische Eigenwertverteilung einer kalten und einer heißen Konfiguration.

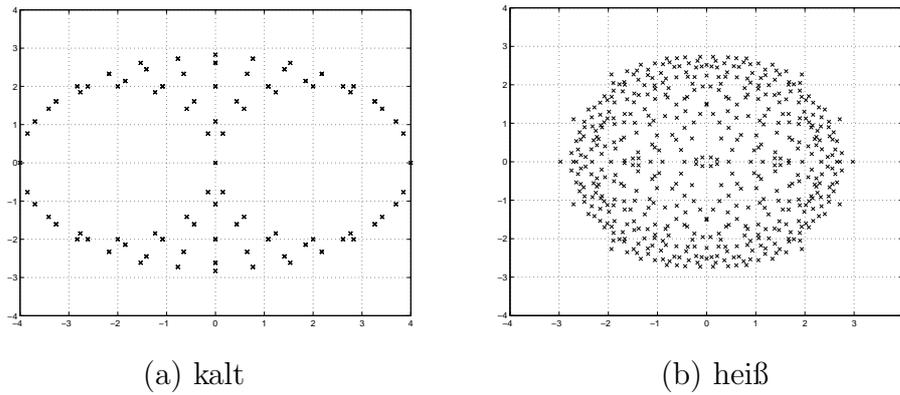


Abbildung 4.1: Das Spektrum $\sigma(D)$ der Matrix D im Falle einer kalten und einer heißen Konfiguration.

In vielen Bereichen der Gittereichtheorie ist eine Eichfixierung der Konfiguration auf dem Gitter hilfreich. Hierzu stehen verschiedene nicht-triviale, iterative Verfahren zur Verfügung [57], [34]. Allen diesen Verfahren ist gemein, daß sie auf einer Folge von (lokalen) Eichtransformationen

$$U_\mu(x) \rightarrow G^H(x)U_\mu(x)G(x + e_\mu), \quad G(x) \in U(n) \quad (G(x) \in SU(n))$$

beruhen, so daß die anfängliche Konfiguration in eine finale Konfiguration überführt wird, welche beispielsweise die Landau-Bedingung,

$$R(U) = \operatorname{Re} \sum_{\mu, x} 1 - \frac{1}{n} \operatorname{tr}(U_\mu(x)),$$

minimiert. Dies führt dazu, daß die resultierenden Systeme geordneter und die Eigenvektoren zu betragskleinen Eigenwerten geometrisch glatter sind.

Eine einfachere Möglichkeit, die sich auch als (Eich-)Fixierung der Konfiguration interpretieren läßt, liefert der folgende Satz, welcher sicherstellt, daß der zum betragskleinsten Eigenwert zugehörige Eigenvektor sowohl geometrisch glatt als auch reell ist.

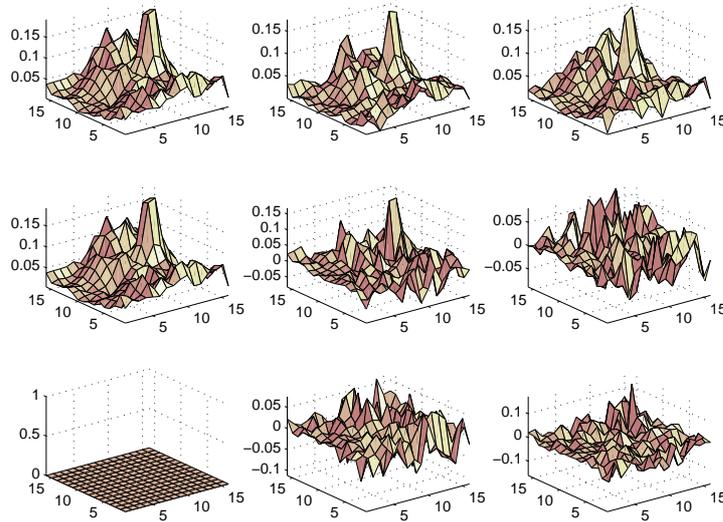
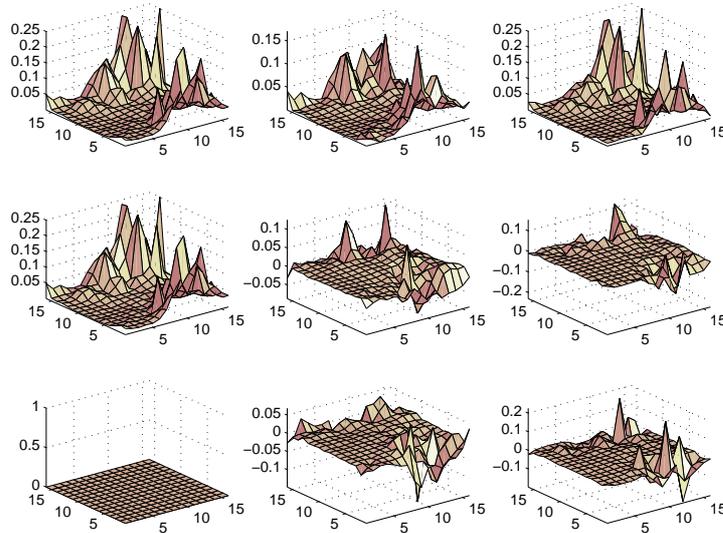
(a) Eigenvektor x (b) Eigenvektor $V^{-1}x$

Abbildung 4.2: Ein Eigenvektor x zum Eigenwert λ_2 der Matrix A und der korrespondierende Eigenvektor $V^{-1}x$ der Matrix $V^{-1}AV$ als Gitterfunktion. In den Zeilen steht der Betrag, der Real- und der Imaginärteil, in den Spalten die Norm, die erste Komponente und die zweite Komponente auf dem jeweiligen Gitterpunkt.

Satz 4.20. Es sei v ein Eigenvektor zum Eigenwert λ der Matrix $A = I - \kappa D$. Dann gilt $\tilde{A}\mathbf{1} = \lambda\mathbf{1}$ mit $\tilde{A} = V^{-1}AV$, wobei $V = \text{diag}(v)$ ist.

Beweis. Es ist $V^{-1}AV\mathbf{1} = V^{-1}Av = \lambda V^{-1}v = \mathbf{1}$. □

Das Spektrum der Matrix $V^{-1}AV$ bleibt gegenüber dem der Matrix A unverändert. Betrachten wir einen Eigenvektor zum betragsmäßig zweitkleinsten Eigenwert als Gitterfunktion aufgetragen, vgl. Abb. 4.2, so erfolgt eine Glättung der Komponenten.

Eine ähnliche Idee ist in [70] zu finden, wobei dort zur Eichfixierung die Eigenvektoren zu kleinen Eigenwerten der kovarianten Laplace Matrix benutzt werden.

4.3 Schurkomplement-Verfahren

Im diesem Abschnitt werden wir zeigen, daß die bereits in Kapitel 3 hergeleiteten Schurkomplement-Verfahren für die Laplace Matrix auch im Fall der Schwinger Matrix erfolgreich angewendet werden können. Wir beschränken uns hier auf die Darstellung der Schurkomplement-Verfahren zur Präkonditionierung der odd-even reduzierten Schwinger Matrix für eine heiße Konfiguration.

Die odd-even reduzierte Schwinger Matrix

Bei einer F/C-Anordnung der (geraden) Gitterpunkte mit $\Omega_f = L_0$ und $\Omega_c = L_2$ besitzt die odd-even reduzierte Schwinger Matrix $A^{(e)}$ die Blockstruktur

$$A^{(e)} = \begin{pmatrix} A_{ff}^{(e)} & A_{fc}^{(e)} \\ A_{cf}^{(e)} & A_{cc}^{(e)} \end{pmatrix} = I - \kappa^2 \begin{pmatrix} D_{ff}^{(e)} & D_{fc}^{(e)} \\ D_{cf}^{(e)} & D_{cc}^{(e)} \end{pmatrix}$$

mit

$$\begin{aligned} D_{ff}^{(e)} &= D_{L_0, L_1} D_{L_1, L_0}, & D_{fc}^{(e)} &= D_{L_0, L_1} D_{L_1, L_2}, \\ D_{cf}^{(e)} &= D_{L_2, L_1} D_{L_1, L_0}, & D_{cc}^{(e)} &= D_{L_2, L_1} D_{L_1, L_2}. \end{aligned}$$

Diese Anordnung der Gitterpunkte entspricht der in Kapitel 3 gewählten Anordnung, vgl. (3.19).

Wir geben die Nichtnullblöcke der beteiligten Teilmatrizen an und beweisen, daß die γ_5 -Symmetrie beim Übergang auf das odd-even reduzierte System bestehen bleibt:

Besetztheitsmuster der Blockdiagonalmatrizen

Lemma 4.21. Für $i = 0, 2$ gilt $\text{diag}(D_{L_i, L_1} D_{L_1, L_i}) = 0$.

Beweis. Wir zeigen, daß sogar die Blockdiagonale verschwindet:

$$\begin{aligned}
(x, x) &: \sum_{\mu=1}^2 ((I - \gamma_\mu) \otimes U_\mu(x)) ((I + \gamma_\mu) \otimes U_\mu^H(x)) \\
&\quad + \sum_{\mu=1}^2 ((I + \gamma_\mu) \otimes U_\mu^H(x - e_\mu)) ((I - \gamma_\mu) \otimes U_\mu(x - e_\mu)) \\
&= \sum_{\mu=1}^2 (I - \gamma_\mu)(I + \gamma_\mu) \otimes U_\mu(x) U_\mu^H(x) \\
&\quad + \sum_{\mu=1}^2 (I + \gamma_\mu)(I - \gamma_\mu) \otimes U_\mu^H(x - e_\mu) U_\mu(x - e_\mu) \\
&= 2 \sum_{\mu=1}^2 (I - \gamma_\mu^2) \otimes I = 0
\end{aligned}$$

da $\gamma_\mu^2 = I$ ist. □

Lemma 4.22. Für $i = 0, 2$ besitzt die Matrix $D_{L_i, L_1} D_{L_1, L_i}$ genau 4 Nicht-nullblöcke in der Blockzeile x ($\mu = 1, 2$),

$$\begin{aligned}
(x, x + 2e_\mu) &: 2(I - \gamma_\mu) \otimes U_\mu(x) U_\mu(x + e_\mu) \\
(x, x - 2e_\mu) &: 2(I + \gamma_\mu) \otimes U_\mu^H(x - e_\mu) U_\mu^H(x - 2e_\mu).
\end{aligned}$$

Beweis. Es gilt

$$\begin{aligned}
(D_{L_i, L_1} D_{L_1, L_i})_{x, x+2e_\mu} &= (D_{L_i, L_1})_{x, x+e_\mu} (D_{L_1, L_i})_{x+e_\mu, x+2e_\mu} \\
&= ((I - \gamma_\mu) \otimes U_\mu(x)) ((I - \gamma_\mu) \otimes U_\mu(x + e_\mu)) \\
&= (I - \gamma_\mu)(I - \gamma_\mu) \otimes U_\mu(x) U_\mu(x + e_\mu).
\end{aligned}$$

Analog erhalten wir

$$\begin{aligned}
(D_{L_i, L_1} D_{L_1, L_i})_{x, x-2e_\mu} &= (D_{L_i, L_1})_{x, x-e_\mu} (D_{L_1, L_i})_{x-e_\mu, x-2e_\mu} \\
&= ((I + \gamma_\mu) \otimes U_\mu^H(x - e_\mu)) ((I + \gamma_\mu) \otimes U_\mu^H(x - 2e_\mu)) \\
&= (I + \gamma_\mu)(I + \gamma_\mu) \otimes U_\mu^H(x - e_\mu) U_\mu^H(x - 2e_\mu).
\end{aligned}$$

Mit Lemma 4.11 folgt nun die Behauptung. □

Besetztheitsmuster der Nichtdiagonalblöcke

Lemma 4.23. Für $i, j = 0, 2$ mit $i \neq j$ besitzt die Matrix $D_{L_i, L_1} D_{L_1, L_j}$ genau 4 Nichtnullblöcke in der Blockzeile x ,

$$\begin{aligned}
(x, x + e_\mu + e_\nu) &: (I - \gamma_\mu)(I - \gamma_\nu) \otimes U_\mu(x)U_\nu(x + e_\mu) \\
&\quad + (I - \gamma_\nu)(I - \gamma_\mu) \otimes U_\nu(x)U_\mu(x + e_\nu) \\
(x, x + e_\mu - e_\nu) &: (I - \gamma_\mu)(I + \gamma_\nu) \otimes U_\mu(x)U_\nu^H(x + e_\mu - e_\nu) \\
&\quad + (I + \gamma_\nu)(I - \gamma_\mu) \otimes U_\nu^H(x - e_\nu)U_\mu(x - e_\nu) \\
(x, x - e_\mu + e_\nu) &: (I + \gamma_\mu)(I - \gamma_\nu) \otimes U_\mu^H(x - e_\mu)U_\nu(x - e_\mu) \\
&\quad + (I - \gamma_\nu)(I + \gamma_\mu) \otimes U_\nu(x)U_\mu^H(x - e_\mu + e_\nu) \\
(x, x - e_\mu - e_\nu) &: (I + \gamma_\mu)(I + \gamma_\nu) \otimes U_\mu^H(x - e_\mu)U_\nu^H(x - e_\mu - e_\nu) \\
&\quad + (I + \gamma_\nu)(I + \gamma_\mu) \otimes U_\nu^H(x - e_\nu)U_\mu^H(x - e_\mu - e_\nu),
\end{aligned}$$

wobei $\mu \neq \nu$ ist.

Beweis. Nachrechnen (wie im Beweis von Lemma 4.22). \square

Wir geben die Einträge der Matrizen $(I \pm \gamma_\mu)(I \pm \gamma_\nu)$ explizit an:

Lemma 4.24. Es gilt

$$\begin{aligned}
(I - \gamma_1)(I - \gamma_2) &= \begin{pmatrix} 1+i & -1+i \\ -1-i & 1-i \end{pmatrix}, \quad (I + \gamma_1)(I - \gamma_2) = \begin{pmatrix} 1-i & 1+i \\ 1-i & 1+i \end{pmatrix}, \\
(I - \gamma_2)(I - \gamma_1) &= \begin{pmatrix} 1-i & -1+i \\ -1-i & 1+i \end{pmatrix}, \quad (I - \gamma_2)(I + \gamma_1) = \begin{pmatrix} 1+i & 1+i \\ 1-i & 1-i \end{pmatrix}, \\
(I - \gamma_1)(I + \gamma_2) &= \begin{pmatrix} 1-i & -1-i \\ -1+i & 1+i \end{pmatrix}, \quad (I + \gamma_1)(I + \gamma_2) = \begin{pmatrix} 1+i & 1-i \\ 1+i & 1-i \end{pmatrix}, \\
(I + \gamma_2)(I - \gamma_1) &= \begin{pmatrix} 1+i & -1-i \\ -1+i & 1-i \end{pmatrix}, \quad (I + \gamma_2)(I + \gamma_1) = \begin{pmatrix} 1-i & 1-i \\ 1+i & 1+i \end{pmatrix}.
\end{aligned}$$

Beweis. Ausrechnen. \square

Lemma 4.25. Die odd-even reduzierte Schwinger Matrix $A^{(e)}$ ist γ_5 -symmetrisch.

Beweis. Da $A = I - \kappa D$ eine γ_5 -symmetrische Matrix ist, erfüllen die Teilmatrizen D_{eo} und D_{oe} die Gleichungen

$$\Gamma_5 D_{eo} \Gamma_5 = D_{oe}^H, \quad \Gamma_5 D_{oe} \Gamma_5 = D_{eo}^H. \quad (4.4)$$

Die Matrix $D_{eo} D_{oe}$ (und $D_{oe} D_{eo}$) ist γ_5 -symmetrisch,

$$\Gamma_5 (D_{eo} D_{oe}) \Gamma_5 = \Gamma_5 D_{eo} \Gamma_5 \Gamma_5 D_{oe} \Gamma_5 = D_{oe}^H D_{eo}^H = (D_{eo} D_{oe})^H. \quad (4.5)$$

Der Nachweis der γ_5 -Symmetrie von $A^{(e)} = I - \kappa^2 D_{eo} D_{oe}$ ist nun trivial. \square

Starke und schwache Kopplungen

Die in Kapitel 3 eingeführte Einteilung in starke und schwache Kopplungen kann für eine kalte Konfiguration in natürlicher Weise auf die Blöcke übertragen werden (vgl. Lemma 3.51):

Lemma 4.26. Für $i, j = 0, 2$ seien $x \in L_i$ und $y \in L_j$ zwei Gitterpunkte mit $x \neq y$. Im Falle einer kalten Konfiguration ist die Kopplung zwischen x und y für $i = j$ schwach, für $i \neq j$ hingegen stark.

Beweis. Vergleichen wir die Betragsmatrizen der die Kopplungen zwischen den Gitterpunkten x und y beschreibenden Nichtnullblöcke der Matrix $D_{L_i, L_1} D_{L_1, L_j}$, so erhalten wir für $i = j$ mit Lemma 4.22 Nichtnullelemente an den Positionen ($\mu = 1, 2$)

$$\begin{aligned} (x, x + 2e_\mu) &: |2(I - \gamma_\mu)| = 2|I - \gamma_\mu| = 2 \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \\ (x, x - 2e_\mu) &: |2(I + \gamma_\mu)| = 2|I + \gamma_\mu| = 2 \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}. \end{aligned}$$

Für $i \neq j$ folgt mit Lemma 4.23 unter Beachtung von Lemma 4.10

$$\begin{aligned} (x, x + e_\mu + e_\nu) &: |(I - \gamma_\mu)(I - \gamma_\nu) + (I - \gamma_\nu)(I - \gamma_\mu)| \\ &= |2(I - \gamma_\mu - \gamma_\nu) + (\gamma_\mu\gamma_\nu + \gamma_\nu\gamma_\mu)| = 2|I - \gamma_\mu - \gamma_\nu| \\ (x, x + e_\mu - e_\nu) &: |(I - \gamma_\mu)(I + \gamma_\nu) + (I + \gamma_\nu)(I - \gamma_\mu)| \\ &= |2(I - \gamma_\mu + \gamma_\nu) - (\gamma_\mu\gamma_\nu + \gamma_\nu\gamma_\mu)| = 2|I - \gamma_\mu + \gamma_\nu| \\ (x, x - e_\mu + e_\nu) &: |(I + \gamma_\mu)(I - \gamma_\nu) + (I - \gamma_\nu)(I + \gamma_\mu)| \\ &= |2(I + \gamma_\mu - \gamma_\nu) - (\gamma_\mu\gamma_\nu + \gamma_\nu\gamma_\mu)| = 2|I + \gamma_\mu - \gamma_\nu| \\ (x, x - e_\mu - e_\nu) &: |(I + \gamma_\mu)(I + \gamma_\nu) + (I + \gamma_\nu)(I + \gamma_\mu)| \\ &= |2(I + \gamma_\mu + \gamma_\nu) + (\gamma_\mu\gamma_\nu + \gamma_\nu\gamma_\mu)| = 2|I + \gamma_\mu + \gamma_\nu|, \end{aligned}$$

wobei $\mu \neq \nu$ ist. Des weiteren gilt

$$\begin{aligned} |I + \gamma_\mu + \gamma_\nu| &= |I + \gamma_\mu - \gamma_\nu| = |I - \gamma_\mu + \gamma_\nu| \\ &= |I - \gamma_\mu - \gamma_\nu| = \begin{pmatrix} 1 & \sqrt{2} \\ \sqrt{2} & 1 \end{pmatrix}. \end{aligned}$$

Ein Koeffizientenvergleich offenbart

$$2\kappa^2 \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \leq 2\kappa^2 \begin{pmatrix} 1 & \sqrt{2} \\ \sqrt{2} & 1 \end{pmatrix},$$

also sind die Kopplungen für $i = j$ (innerhalb eines Blockes komponentenweise) schwächer als die für $i \neq j$. \square

Für eine heiße Konfiguration hingegen tritt eine andere Situation ein:

Lemma 4.27. Für $i, j = 0, 2$ mit $i \neq j$ ist der Erwartungswert E der Komponenten des die Kopplung zwischen $x \in L_i$ und $y \in L_j$ beschreibenden Nichtnullblockes im Falle einer heißen Konfiguration gegeben durch

$$4\sqrt{2}\kappa^2/\pi \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

Beweis. Es sei $y = x + e_1 + e_2$. Des weiteren sei

$$a = U_1(x)U_2(x + e_1), \quad b = U_2(x)U_1(x + e_2).$$

Dann gilt mit Lemma 4.24

$$\begin{aligned} & (I - \gamma_1)(I - \gamma_2) \otimes a + (I - \gamma_2)(I - \gamma_1) \otimes b \\ &= \begin{pmatrix} (a + b) + i(a - b) & -(a + b) + i(a + b) \\ -(a + b) + i(a + b) & (a + b) - i(a - b) \end{pmatrix}. \end{aligned}$$

Mit Satz A.3 erhalten wir

$$\begin{pmatrix} E(|(a + b) + i(a - b)|) & E(|-(a + b) + i(a + b)|) \\ E(|-(a + b) + i(a + b)|) & E(|(a + b) - i(a - b)|) \end{pmatrix} = \frac{4\sqrt{2}\kappa^2}{\pi} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

Die verbleibenden Fälle werden analog bewiesen. \square

Obwohl für eine heiße Konfiguration nach dem obigen Lemma 4.27 der Erwartungswert der Kopplungen für $i = j$ größer ist als der für $i \neq j$,

$$2\kappa^2 \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} > \frac{4\sqrt{2}\kappa^2}{\pi} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \approx 1.8\kappa^2 \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix},$$

rechtfertigen numerische Experimente eine zu Lemma 4.26 analoge Aussage, vgl. Abb. 4.3, so daß ebenso wie für die Laplace Matrix auch im Fall des Schwinger Modells die gewählte Partitionierung der odd-even reduzierten Matrix gerechtfertigt erscheint. Eine Erklärung hierfür liefert die folgende Beobachtung:

In Abbildung 4.3 ist insbesondere zu sehen, daß es in jedem Nichtnullblock (mindestens) eine Kopplung gibt, die betragsmäßig $\geq 2\kappa^2$ ist und somit auch betragsmäßig \geq als jede Kopplung zwischen zwei Gitterpunkten $x, y \in L_i$, $x \neq y$. (In diesem Sinne ist für eine heiße Konfiguration und $i, j = 0, 2$ die Kopplung zwischen $x \in L_i$ und $y \in L_j$ mit $x \neq y$ für $i \neq j$ stark und für $i = j$ schwach.)

Diese Beobachtung läßt sich wie folgt beweisen:

Satz 4.28. Für $i, j = 0, 2$ mit $i \neq j$ ist mindestens ein Diagonalelement des die Kopplung zwischen $x \in L_i$ und $y \in L_j$ beschreibenden Nichtnullblockes betragsmäßig $\geq 2\kappa^2$.

Beweis. Ein Blick auf Lemma 4.24 zeigt, daß die Diagonalelemente des die Kopplung zwischen $x \in L_i$ und $y \in L_j$ beschreibenden Nichtnullblockes von der Gestalt $-\kappa^2(a + b \pm i(a - b))$ mit $a = \exp(2\pi is)$, $b = \exp(2\pi it)$ und $s, t \in [0, 1)$ sind. Mit Lemma A.4 folgt nun die Behauptung. \square

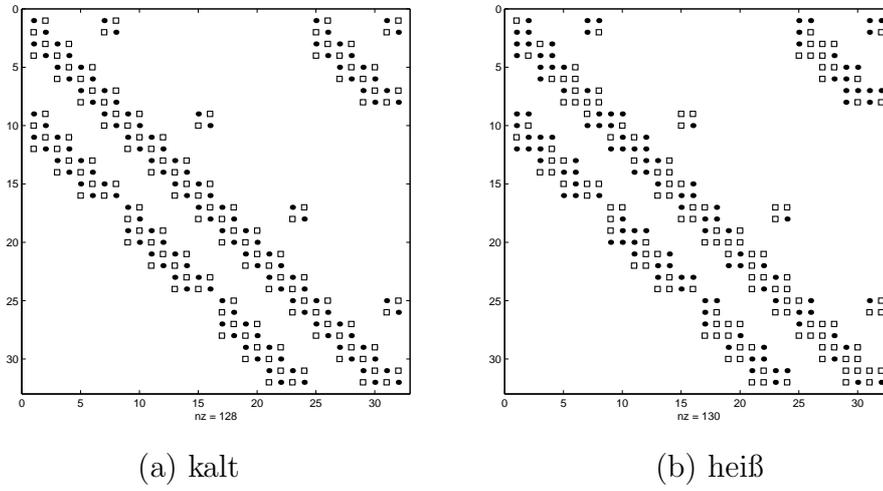


Abbildung 4.3: Das Besetzungsmuster der Matrix $A_{fc}^{(e)}$. Die Matrixeinträge an den durch \square gekennzeichneten Positionen sind betragsmäßig $\geq 2\kappa^2$, und der zugehörige Nichtnullblock repräsentiert eine starke Kopplung.

Die Matrix $A_{ff}^{(e)}$

Mit Hilfe des Lemma 4.22 ist es einfach zu zeigen, daß die Matrix $A_{ff}^{(e)}$ interpretiert werden kann als eine neue Schwinger Matrix auf den Gitterpunkten aus L_0 ,

$$\hat{A} = I - \hat{\kappa}\hat{D}, \quad \hat{\kappa} = 2\kappa^2, \quad \hat{D} = \hat{D}(\hat{U}), \quad (4.6)$$

wobei die Matrix \hat{D} aus sogenannten Produktlinks besteht,

$$\hat{U} = \left\{ \hat{U}_\mu(x) \right\}, \quad \hat{U}_\mu(x) = U_\mu(x)U_\mu(x + e_\mu). \quad (4.7)$$

Insbesondere ist $A_{ff}^{(e)}$ eine γ_5 -symmetrische Matrix, und falls $4|N$ gilt, so ist auch die odd-even-Symmetrie gegeben.

Für eine gegebene kalte (heiße) Konfiguration U ist die resultierende neue Konfiguration \widehat{U} wiederum kalt (heiß). Insbesondere gilt $\kappa_c = \widehat{\kappa}_c$. Numerische Experimente zeigen, daß für realistische Konfigurationen U die Konfiguration \widehat{U} „heißer“ wird, d.h. es gilt $\kappa_c \leq \widehat{\kappa}_c$.

Da $\rho(\widehat{D}) \leq \rho(D)$ und $\widehat{\kappa} \ll \kappa$ für alle $0 \leq \kappa < \kappa_c$ ist, ist die Matrix $A_{ff}^{(e)}$ deutlich besser konditioniert als die ursprüngliche Matrix A , vgl. Abb. 4.4.

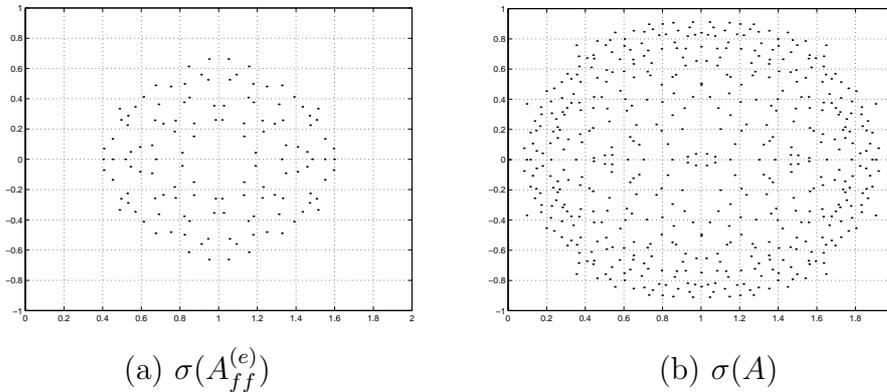


Abbildung 4.4: Die Spektren der Matrizen A und $A_{ff}^{(e)}$ für eine heiße Konfiguration.

Diese Beobachtung motiviert eine Approximation der Inversen durch einen lokal wirkenden Präkonditionierer:

Approximative Inverse der Matrix $A_{ff}^{(e)}$

Völlig analog zu den in Kapitel 3 diskutierten Präkonditionierern lassen sich entsprechende Präkonditionierer für die Matrix $A_{ff}^{(e)}$ herleiten. Ein wichtiges Hilfsmittel hierzu ist das folgende Lemma:

Lemma 4.29 (Zarantonello). Es sei $K(0, \rho)$ der Kreis mit Radius ρ um den Ursprung und $z_0 \in \mathbb{C} \setminus K(0, \rho)$. Dann gilt

$$\min_{p \in P_k, p(z_0)=1} \max_{z \in K(0, \rho)} |p(z)| = \left(\frac{\rho}{|z_0|} \right)^k,$$

und das Minimum wird angenommen für das Polynom $(z/z_0)^k$.

Beweis. Siehe [56].

□

Im Falle einer heißen Konfiguration gilt für das Spektrum der Matrix $A_{ff}^{(e)}$ die Inklusion

$$\sigma\left(A_{ff}^{(e)}\right) \subseteq K(1, \rho\kappa^2), \quad \rho = \rho(D_{L_0, L_1} D_{L_1, L_0}).$$

Mit Hilfe des Lemmas 4.29 können wir daher für eine heiße Konfiguration eine polynomiale Approximation der Inversen der Matrix $A_{ff}^{(e)}$ konstruieren, so daß das Spektrum der präkonditionierten Matrix bestmöglich um 1 geclustert wird:

Lemma 4.30. Es sei eine approximative Inverse der Matrix $A_{ff}^{(e)}$ definiert als

$$\left(\tilde{A}_{ff}^{(e)}\right)^{-1} = I + \omega\kappa^2 D_{L_0, L_1} D_{L_1, L_0}.$$

Dann wird für die Parameterwahl $\omega = 1$ das Spektrum der präkonditionierten Matrix bestmöglich um 1 geclustert, und es gilt

$$\sigma\left(\left(\tilde{A}_{ff}^{(e)}\right)^{-1} A_{ff}^{(e)}\right) \subseteq K\left(1, (\rho\kappa^2)^2\right), \quad \rho = \rho(D_{L_0, L_1} D_{L_1, L_0}).$$

Beweis. Mit $p_\omega(z) = \omega z^2 - (\omega - 1)z$ ist $I - p_\omega(\kappa D_{L_0, L_1} D_{L_1, L_0})$ die präkonditionierte Matrix. Wir wenden das Lemma 4.29 an und erhalten

$$\min_{p_\omega \in \mathcal{P}_2, p_\omega(1)=1} \max_{z \in K(0, \rho\kappa^2)} |p_\omega(z)| = (\rho\kappa^2)^2.$$

Das Minimum wird für $p_1(z) = z^2$, also für $\omega = 1$ angenommen. \square

Aufgrund der odd-even Symmetrie besitzt die Matrix $A_{ff}^{(e)}$ bei einer odd-even Anordnung der feinen Gitterpunkte die Gestalt

$$A_{ff}^{(e)} = I - \kappa^2 \begin{pmatrix} 0 & D_{oe}^{(e)} \\ D_{eo}^{(e)} & 0 \end{pmatrix}.$$

Vergleichen wir den durch die direkte ILU-Zerlegung zum Muster der Nicht-nullelemente induzierten Präkonditionierer

$$M_{\text{ILU}} = \begin{pmatrix} I & 0 \\ -\kappa^2 D_{eo}^{(e)} & I \end{pmatrix} \begin{pmatrix} I & -\kappa^2 D_{oe}^{(e)} \\ 0 & I \end{pmatrix}$$

mit dem Jacobi-Präkonditionierer

$$M_{\text{J}}^{-1} = I + \kappa^2 \begin{pmatrix} 0 & D_{oe}^{(e)} \\ D_{eo}^{(e)} & 0 \end{pmatrix},$$

so ist der Aufwand in beiden Fällen identisch. Auch die Eigenwertverteilung ist im wesentlichen die gleiche,

$$\begin{aligned}\sigma\left(M_{ILU}^{-1}A_{ff}^{(e)}\right) &= \{1\} \cup \sigma\left(I - \kappa^4 D_{eo}^{(e)} D_{oe}^{(e)}\right) \\ \sigma\left(M_J^{-1}A_{ff}^{(e)}\right) &= \sigma\left(I - \kappa^4 D_{oe}^{(e)} D_{eo}^{(e)}\right) \cup \sigma\left(I - \kappa^4 D_{eo}^{(e)} D_{oe}^{(e)}\right).\end{aligned}$$

Der Eigenwert 1 der ILU-präkonditionierten Matrix tritt $n/2$ -fach auf. Die Eigenwerte der Jacobi-präkonditionierten Matrix treten doppelt auf. Aus diesem Grund ist die ILU- der Jacobi-Präkonditionierung vorzuziehen. Außerdem trägt die durch

$$M_{ILU}^{-1} = \begin{pmatrix} I & \kappa^2 D_{oe}^{(e)} \\ 0 & I \end{pmatrix} \begin{pmatrix} I & 0 \\ \kappa^2 D_{eo}^{(e)} & I \end{pmatrix} = M_J^{-1} + \kappa^4 \begin{pmatrix} D_{oe}^{(e)} D_{eo}^{(e)} & 0 \\ 0 & 0 \end{pmatrix} \quad (4.8)$$

gegebene approximative Inverse über zusätzliche Kopplungen zur Restriktion und Prolongation bei.

Beide Präkonditionierer M_{ILU}^{-1} und M_J^{-1} sind γ_5 -symmetrisch, reflektieren also die in Lemma 4.25 bewiesene γ_5 -Symmetrie der odd-even reduzierten Matrix.

Schurkomplement-Approximation

Zur Konstruktion einer geeigneten Approximation des exakten Schurkomplements der odd-even reduzierten Schwinger Matrix,

$$S_{A^{(e)}} = I - \kappa^2 D_{L_2, L_1} \left(I - \kappa^2 D_{L_1, L_0} D_{L_0, L_1} \right)^{-1} D_{L_1, L_2},$$

bestimmen wir in einem ersten Schritt eine trunkeerte Neumann-Reihe als polynomiale Approximation der Inversen der Matrix $I - \kappa^2 D_{L_1, L_0} D_{L_0, L_1}$ und verwenden anschließend als eigentliche Schurkomplement-Approximation den Ansatz

$$\tilde{S}_{A^{(e)}} = I - \kappa^2 D_{L_2, L_1} \sum_{\nu=0}^k \left(\kappa^2 D_{L_1, L_0} D_{L_0, L_1} \right)^\nu D_{L_1, L_2}. \quad (4.9)$$

Wegen $\Gamma_5 D_{L_i, L_j} \Gamma_5 = D_{L_j, L_i}^H$ sind sowohl das exakte Schurkomplement als auch die obige Schurkomplement-Approximation γ_5 -symmetrische Matrizen.

Da wir einer rekursive Vorgehensweise ermöglichen wollen, fordern wir, daß die Schurkomplement-Approximation ebenso wie die odd-even reduzierte Schwinger Matrix einen 9-Punkte-Stern beschreibt. Diese Forderung führt

dazu, daß wir die Schurkomplement-Approximation (4.9) für $k > 1$ auf das entsprechende Muster trunkieren müssen.

Bereits bei der Konstruktion einer Schurkomplement-Approximation für die Laplace Matrix wurde festgestellt, daß der 9-Punkte-Stern ($k = 1$) dem 5-Punkte-Stern ($k = 0$) vorzuziehen ist, s. Seite 66f. Daher verwenden wir als Schurkomplement-Approximation (4.9) mit $k = 1$, und erhalten als ersten einfachen Ansatz die Approximation

$$\tilde{S}_{A^{(e)}} = I - \kappa^2 D_{L_2, L_1} (I + \kappa^2 D_{L_1, L_0} D_{L_0, L_1}) D_{L_1, L_2}. \quad (4.10)$$

Aufgrund von Lemma 3.33 sind die Spektren der Matrizen $I - \kappa^2 D_{L_1, L_0} D_{L_0, L_1}$ und $I - \kappa^2 D_{L_0, L_1} D_{L_1, L_0}$ im wesentlichen identisch, so daß für eine heiße Konfiguration $I + \kappa^2 D_{L_1, L_0} D_{L_0, L_1}$ die im min-max-Sinne beste polynomiale Approximation der Inversen der Matrix $I - \kappa^2 D_{L_1, L_0} D_{L_0, L_1}$ ist, vgl. Lemma 4.29. Natürlich muß sich diese Optimalität nicht auf die Schurkomplement-Approximation übertragen, so daß eine durch eine Relaxation modifizierte Matrix

$$\tilde{S}_{A^{(e)}} = I - \kappa^2 D_{L_2, L_1} (I + \omega \kappa^2 D_{L_1, L_0} D_{L_0, L_1}) D_{L_1, L_2} \quad (4.11)$$

durchaus bessere Approximationseigenschaften aufweisen kann. Da es aber unmöglich erscheint, a priori zu einer gegebenen Konfiguration einen optimalen Relaxationsparameter anzugeben, verwenden wir statt dessen mit $k = 2$ einen anderen Zugang,

$$\tilde{S}_{A^{(e)}} = I - \kappa^2 D_{L_2, L_1} \sum_{\nu=0}^2 (\kappa^2 D_{L_1, L_0} D_{L_0, L_1})^\nu D_{L_1, L_2}, \quad (4.12)$$

und trunkieren die resultierende Schurkomplement-Approximation, einen 21-Punkte-Stern, so daß wir den üblichen 9-Punkte-Stern erhalten.

Interessant ist dieser Ansatz insbesondere deswegen, weil mit Lemma 4.31 die Graphen der Matrizen $D_{L_1, L_0} D_{L_0, L_1}$ und $(D_{L_1, L_0} D_{L_0, L_1})^2$ disjunkte Kantenmengen besitzen. Daher liefern die durch die Matrix

$$D_{L_2, L_1} (D_{L_1, L_0} D_{L_0, L_1})^2 D_{L_1, L_2}$$

repräsentierten Kopplungen gegenüber dem ersten Ansatz (4.10) echte neue Beiträge, welche auch nicht durch eine Relaxation der Gestalt (4.11) berücksichtigt und in einfacher Weise aufaddiert werden können.

Lemma 4.31. Es sei $1 \leq k \leq 3$. Dann gilt $\text{diag} \left((D_{L_1, L_0} D_{L_0, L_1})^k \right) = 0$.

Beweis. Betrachten wir die Abbildung 4.5, so sind alle durch $(D_{L_1,L_0}D_{L_0,L_1})^k$ repräsentierten Wege mit Anfangs- und Endpunkt x von der Gestalt

$$(x = v_1, \dots, v_i, v_{i+1}, v_{i+2}, \dots, v_{2k+1} = x), \quad v_{i+2} = v_i, \quad i \in \{1, \dots, 2k - 1\}.$$

Aufgrund der γ_μ -Struktur der Matrizen, vgl. Lemma 4.11, tragen diese Wege jedoch nicht zum Diagonaleintrag

$$\left((D_{L_1,L_0}D_{L_0,L_1})^k \right)_{x,x}$$

bei. (Die Situation ändert sich erstmalig für $k = 4$, s. Abb. 4.5). □

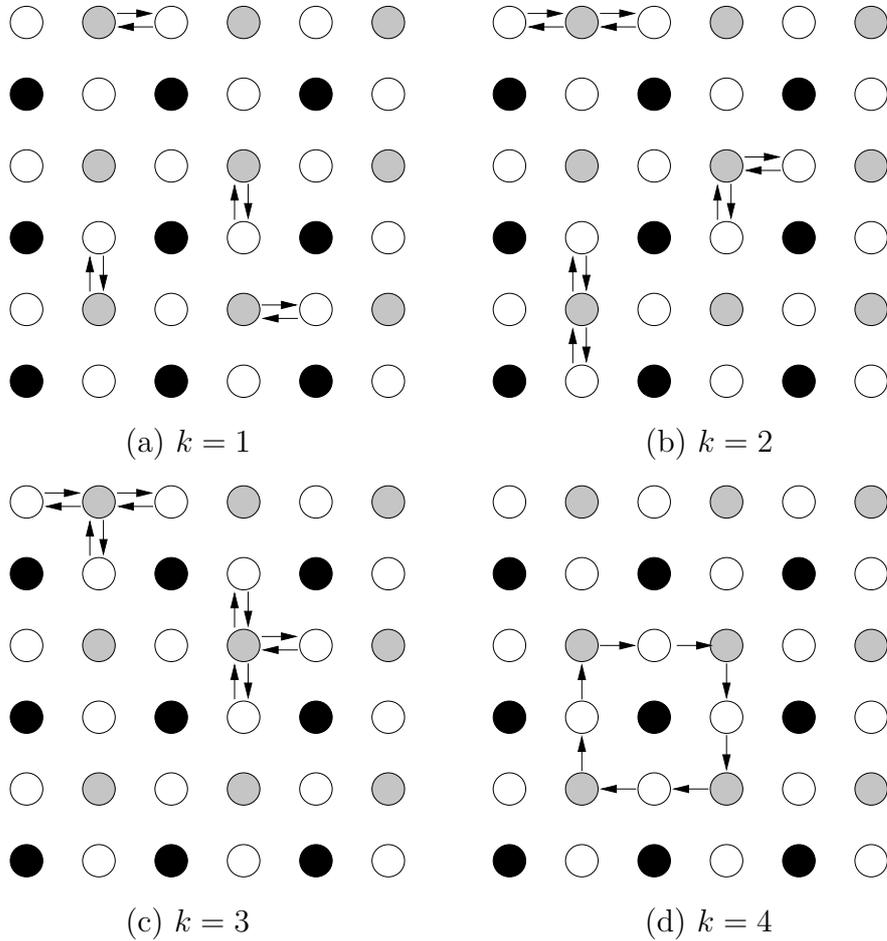


Abbildung 4.5: Die Diagonale der Matrix $(D_{L_1,L_0}D_{L_0,L_1})^k$ für $1 \leq k \leq 4$.

Da in beiden Ansätzen die Kopplungen zu den nächsten Nachbarn von der Ordnung $\mathcal{O}(\kappa^2)$, die Kopplungen zu den verbleibenden Nachbarn hingegen von der Ordnung $\mathcal{O}(\kappa^4)$ sind, siehe Abb. 4.6(b), ist eine Einteilung

der Kantenmenge in starke und schwache Kanten offensichtlich. Diese Einteilung der Kantenmenge korrespondiert zu einer odd-even Anordnung der Schurkomplement-Approximation. Dazu partitionieren wir die Gitterpunkte aus L_2 in eine Menge von geraden und eine Menge von ungeraden Gitterpunkten. Die Kopplungen der geraden (ungeraden) Gitterpunkte mit geraden (ungeraden) Gitterpunkten sind schwach, die Kopplungen zwischen geraden und ungeraden Gitterpunkten sind stark. Eine rekursive Vorgehensweise ist nun naheliegend, so daß echte Multilevel-Verfahren in einfacher Weise konstruiert werden können, siehe Abb. 4.6(c) und Abb. 4.6(d).

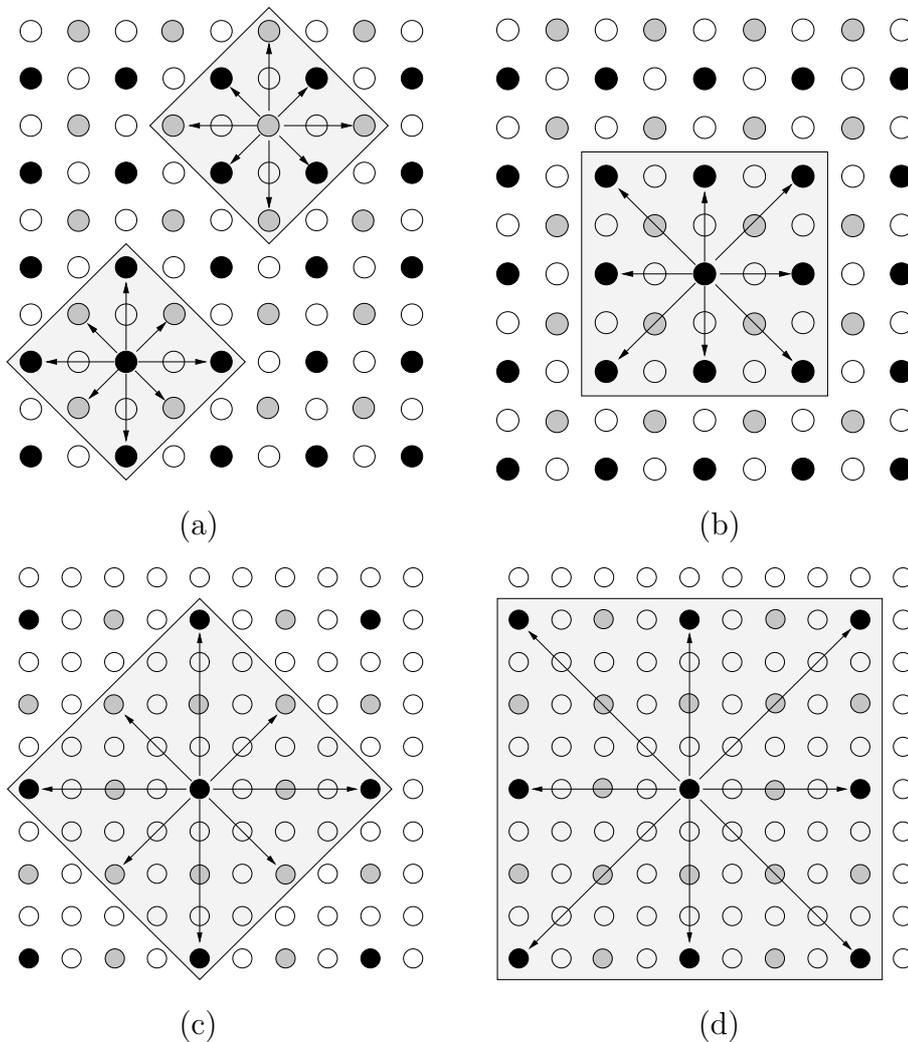


Abbildung 4.6: Die 9-Punkte-Sterne der ersten vier (Grobgitter-)Matrizen. Ausgangspunkt ist der 9-Punkte-Stern des odd-even reduzierten Systems (a).

4.4 Numerische Ergebnisse

Die numerischen Ergebnisse wurden mit MATLAB 5.3 erzeugt. Als Löser diente das BiCGStab-Verfahren in Form der MATLAB-Funktion `bicgstab`. Die Konvergenz wurde erreicht, falls das relative Residuum der Bedingung

$$\|r^{(k)}\|_2 \leq \epsilon \|\varphi\|_2, \quad \epsilon = 1.0E - 9$$

genügte. Zu lösen war

- a) Das originäre System $A\psi = \varphi$,
- b) Das odd-even reduzierte System $A_e\psi_e = \tilde{\varphi}_e$,
- c) Das präkonditionierte odd-even reduzierte System,

jeweils auf Gittern verschiedener Größe und für 5 zufällig erzeugte heiße Konfigurationen bei einer Masse $m_q = 0.01$ ($a = 1$). Die rechte Seite φ wurde zufällig erzeugt, $\tilde{\varphi}_e$ wie in Definition 3.15 initialisiert. Der Startwert der Iteration war stets der Nullvektor.

Es wurden zwei Schurkomplement-Präkonditionierer untersucht, die sich lediglich in der Auswahl der Schurkomplement-Approximation unterschieden. Bei der ersten Methode der Präkonditionierung wurde der Ansatz (4.10), bei der zweiten hingegen (4.12) verwendet. In beiden Fällen kam der ILU-Präkonditionierer (4.8) als approximative Inverse zum Einsatz.

Abbildung 4.7 zeigt das typische Konvergenzverhalten für die drei oben angegebenen Systeme. In der Tabelle 4.1 ist die Anzahl der benötigten BiCGStab-Iterationsschritte bis zum Erreichen des vorgegebenen Abbruchkriteriums (Konvergenzkriteriums) aufgelistet.

Eine Bewertung der numerischen Ergebnisse erfolgt in Abschnitt 4.6.

4.5 Die hermitesche Schwinger Matrix

In diesem Abschnitt führen wir die hermitesche Schwinger Matrix ein und zeigen kurz auf, wie im wesentlichen ohne zusätzlichen Aufwand ein Schurkomplement-Verfahren für die hermitesche Schwinger Matrix aus einem existierenden Schurkomplement-Verfahren für die originäre Schwinger Matrix abgeleitet werden kann. Der Schwerpunkt dieses Abschnitts liegt jedoch im Nachweis der Glättungseigenschaft (2.29). Für die hermitesche Schwinger Matrix ist deshalb eine Kombination von Schurkomplement-Präkonditionierern mit Glättungsiterationen, wie sie beispielsweise in [49] vorgeschlagen wurde, realisierbar.

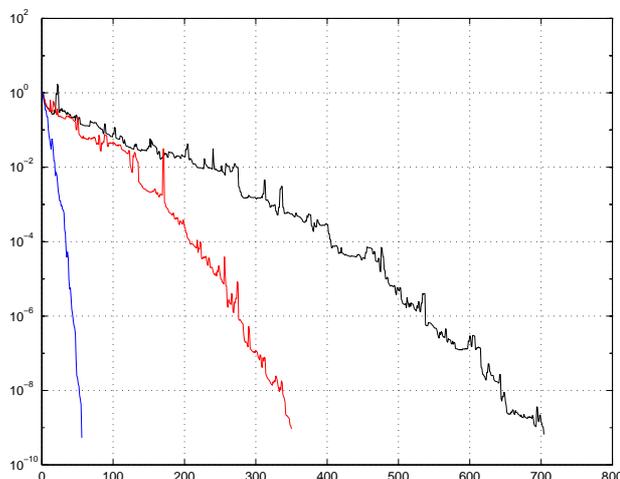


Abbildung 4.7: Das typische Konvergenzverhalten des BiCGStab-Verfahrens für die drei äquivalenten Systeme. In der Abbildung sind die Resultate für die Konfiguration `conf32c` zu sehen ($N = 32$). Aufgetragen wurden die Norm des relativen Residuums über die BiCGStab-Halbschritte. Das unpräkonditionierte BiCGStab-Verfahren zur Lösung von $A\psi = \varphi$ benötigt ungefähr 700 Halbschritte, das unpräkonditionierte BiCGStab-Verfahren zur Lösung des odd-even reduzierten Systems $A_e\psi_e = \tilde{\varphi}_e$ etwa 350 Halbschritte. Das via (4.12) präkonditionierte odd-even reduzierte System verdeutlicht mit 55 Halbschritten die Effizienz des Präkonditionierers.

Lemma 4.32. Es gilt $(\Gamma_5 A)^H = \Gamma_5 A$.

Beweis. $Q^H = (\Gamma_5 A)^H = A^H \Gamma_5 = (\Gamma_5 A \Gamma_5) \Gamma_5 = \Gamma_5 A$. \square

Definition 4.33. Die Matrix $Q = \Gamma_5 A$ heißt hermitesche Schwinger Matrix.

Ist $0 < \kappa_c(Q)$ der Wert von κ , bei dem die Matrix Q zum ersten Mal singularär wird, so gilt:

Lemma 4.34. Die hermitesche Schwinger Matrix Q ist maximal indefinit für $0 \leq \kappa < \kappa_c(Q)$.

Beweis. Für $\kappa = 0$ besitzt die Matrix Q die Eigenwerte ± 1 mit Vielfachheit $n/2$. Also ist Q maximal indefinit. Für $0 < \kappa < \kappa_c(Q)$ folgt die Behauptung durch ein einfaches Homotopieargument. \square

4.5.1 Schurkomplement-Verfahren

Die Herleitung von Schurkomplement-Verfahren für Q kann zurückgeführt werden auf die Konstruktion geeigneter Verfahren für A . Ausgangspunkt ist eine exakte Schurkomplement-Zerlegung der Matrix A .

Konfiguration	N_1	N_2 (N_1/N_2)	N_3 (N_2/N_3)	N_4 (N_2/N_4)
conf16a	173 0	88.5 (1.9548)	20.0 (4.4250)	18.5 (4.7838)
conf16b	185.5	93.5 (1.9840)	22.5 (4.1556)	20.5 (4.5610)
conf16c	184.5	92.5 (1.9946)	21.5 (4.3023)	18.0 (5.1389)
conf16d	159.5	78.5 (2.0318)	20.0 (3.9250)	19.5 (4.0256)
conf16e	159.5	77.5 (2.0581)	18.0 (4.3056)	17.5 (4.4286)
Mittelwert	172.4	86.1 (2.0023)	20.4 (4.2206)	18.8 (4.5798)
conf32a	369.5	186.5 (1.9812)	35.5 (5.2535)	31.5 (5.9206)
conf32b	399.5	200.0 (1.9975)	41.5 (4.8193)	37.5 (5.3333)
conf32c	351.5	174.5 (2.0143)	31.0 (5.6290)	27.5 (6.3455)
conf32d	322.0	159.5 (2.0188)	29.0 (5.5000)	25.5 (6.2549)
conf32e	372.0	180.5 (2.0609)	36.5 (4.9452)	33.5 (5.3881)
Mittelwert	362.9	180.2 (2.0139)	34.7 (5.1931)	31.1 (5.7942)
conf64a	388.0	204.5 (1.8973)	39.0 (5.2436)	32.0 (6.3906)
conf64b	277.0	143.5 (1.9303)	26.5 (5.4151)	23.5 (6.1064)
conf64c	405.5	203.5 (1.9926)	35.5 (5.7324)	31.5 (6.4603)
conf64d	284.5	143.0 (1.9895)	26.0 (5.5000)	21.0 (6.8095)
conf64e	226.0	113.5 (1.9912)	19.5 (5.8205)	17.5 (6.4857)
Mittelwert	316.2	161.6 (1.9567)	29.3 (5.5154)	25.1 (6.4382)

Tabelle 4.1: Anzahl benötigter BiCGStab-Iterationen bis zum Erreichen des Abbruchkriteriums. Dabei kennzeichnet N_1 die Iterationszahlen für das originäre, N_2 die für das odd-even reduzierte sowie N_3 und N_4 die für das via (4.10) bzw. (4.12) präkonditionierte odd-even reduzierte System.

Lemma 4.35. Es gilt

$$Q = \begin{pmatrix} I & 0 \\ \Gamma_5 A_{cf} A_{ff}^{-1} \Gamma_5 & I \end{pmatrix} \begin{pmatrix} \Gamma_5 A_{ff} & 0 \\ 0 & \Gamma_5 S_A \end{pmatrix} \begin{pmatrix} I & A_{ff}^{-1} A_{fc} \\ 0 & I \end{pmatrix}.$$

Beweis. Wir kennen bereits die exakte Schurkomplement-Zerlegung (2.13) von A . Daher folgt

$$\begin{aligned} Q &= \Gamma_5 \begin{pmatrix} I & 0 \\ A_{cf} A_{ff}^{-1} & I \end{pmatrix} \begin{pmatrix} A_{ff} & 0 \\ 0 & S_A \end{pmatrix} \begin{pmatrix} I & A_{ff}^{-1} A_{fc} \\ 0 & I \end{pmatrix} \\ &= \Gamma_5 \begin{pmatrix} I & 0 \\ A_{cf} A_{ff}^{-1} & I \end{pmatrix} \Gamma_5 \Gamma_5 \begin{pmatrix} A_{ff} & 0 \\ 0 & S_A \end{pmatrix} \begin{pmatrix} I & A_{ff}^{-1} A_{fc} \\ 0 & I \end{pmatrix} \\ &= \begin{pmatrix} I & 0 \\ \Gamma_5 A_{cf} A_{ff}^{-1} \Gamma_5 & I \end{pmatrix} \begin{pmatrix} \Gamma_5 A_{ff} & 0 \\ 0 & \Gamma_5 S_A \end{pmatrix} \begin{pmatrix} I & A_{ff}^{-1} A_{fc} \\ 0 & I \end{pmatrix}, \end{aligned}$$

wobei Γ_5 entsprechend partitioniert sei. \square

Ein Schurkomplement-Präkonditionierer der Matrix Q ist gegeben durch

$$M_Q^{-1} = \begin{pmatrix} I & -\tilde{A}_{ff}^{-1}A_{fc} \\ 0 & I \end{pmatrix} \begin{pmatrix} \tilde{A}_{ff}^{-1}\Gamma_5 & 0 \\ 0 & \tilde{S}_A^{-1}\Gamma_5 \end{pmatrix} \begin{pmatrix} I & 0 \\ -\Gamma_5 A_{cf} \tilde{A}_{ff}^{-1} \Gamma_5 & I \end{pmatrix}, \quad (4.13)$$

wobei die approximative Inverse der Matrix A_{ff} und die Schurkomplement-Approximation bereits bekannt seien aus der Diskussion geeigneter Schurkomplement-Präkonditionierer der Matrix A .

Abschließend weisen wir darauf hin, daß die matrixabhängige Restriktion und die zugehörige Prolongation

$$R_Q = \begin{pmatrix} -\Gamma_5 A_{cf} \tilde{A}_{ff}^{-1} \Gamma_5 & I \end{pmatrix}, \quad P_Q = \begin{pmatrix} -\tilde{A}_{ff}^{-1} A_{fc} \\ I \end{pmatrix}, \quad (4.14)$$

zueinander adjungiert sind, sofern eine γ_5 -symmetrische approximative Inverse der Matrix A_{ff} gewählt wird,

$$\Gamma_5 \tilde{A}_{ff}^{-1} \Gamma_5 = \tilde{A}_{ff}^{-H} \implies R_Q = P_Q^H. \quad (4.15)$$

4.5.2 Glättungseigenschaft

Da die Matrix Q hermitesch ist, können wir die Glättungseigenschaft (2.29) für eine Reihe von Iterationsverfahren beweisen. Wir werden den Nachweis der Glättungseigenschaft zunächst explizit für den modifizierten Jacobi-Glätter führen und anschließend einen optimalen semiiterativen Glätter herleiten. Dies ist bereits als eigenständiges Resultat interessant. Aber auch eine Kombination von Schurkomplement-Präkonditionierern mit Glättungsiterationen [49] ist dadurch für die hermitesche Schwinger Matrix vorstellbar.

Der (modifizierte) Jacobi-Glätter

Wir untersuchen zunächst einen einfachen Glättungsansatz. Der vorgestellte Glätter entspricht dem Jacobi-Glätter im Falle von Q^2 .

Lemma 4.36. Es sei $S = (I - \theta Q)(I + \theta Q)$ mit $\theta = 1/\|Q\|_2$. Der Glätter S besitzt die Glättungseigenschaft, denn es ist

$$\|QS^\nu\|_2 \leq \frac{1}{\sqrt{2\nu+1}} \left(\frac{2\nu}{2\nu+1} \right)^\nu \|Q\|_2.$$

Beweis. Es gilt $\|QS^\nu\|_2 = \theta^{-1}\|\theta Q(I - \theta^2 Q^2)^\nu\|_2 = \|p(\theta Q)\|_2 \|Q\|_2$, wobei das Polynom p durch $p(t) = t(1 - t^2)^\nu$ definiert ist. Wegen $\sigma(\theta Q) \subseteq [-1, 1]$ genügt es, das Maximum von p im Intervall $[-1, 1]$ durch eine einfache Kurvendiskussion zu bestimmen. Aus

$$p'(t) = 0 \iff t = \pm 1, \quad t = \pm \frac{1}{\sqrt{2\nu + 1}},$$

$$p(\pm 1) = 0, \quad p\left(\pm \frac{1}{\sqrt{2\nu + 1}}\right) = \pm \frac{1}{\sqrt{2\nu + 1}} \left(\frac{2\nu}{2\nu + 1}\right)^\nu$$

folgt die Behauptung. \square

Lemma 4.37. Die Glättungiteration wird durch die Zerlegung $Q = M - N$ der Matrix Q mit $M^{-1} = \theta^2 Q$ induziert.

Beweis. Es gilt $M^{-1}N = M^{-1}(M - Q) = I - M^{-1}Q = I - \theta^2 Q^2 = S$. \square

Die Iterationsvorschrift lautet

$$x^{(k+1)} = (I - M^{-1}Q)x^{(k)} + M^{-1}b = x^{(k)} + M^{-1}r^{(k)},$$

und für das Residuum erhalten wir

$$r^{(k+1)} = b - Qx^{(k+1)} = b - Q(x^{(k)} + M^{-1}r^{(k)}) = r^{(k)} - QM^{-1}r^{(k)}.$$

Insgesamt erhalten wir Algorithmus 4.1.

Algorithmus 4.1 Jacobi-Glätter für Q

- 1: wähle $x^{(0)}$
 - 2: $r^{(0)} = b - Qx^{(0)}$
 - 3: **for** $k = 0, \dots, \nu - 1$ **do**
 - 4: $p^{(k)} = \theta^2 Qr^{(k)}$
 - 5: $x^{(k+1)} = x^{(k)} + p^{(k)}$
 - 6: $r^{(k+1)} = r^{(k)} - Qp^{(k)}$
 - 7: **end for**
-

Semiiterative Glätter

Verallgemeinern wir den Ansatz einer modifizierten Jacobi-Glättung, indem wir ν Glättungsschritte mit einem noch näher zu bestimmenden Glätter S durchführen,

$$S^\nu = s(\theta Q), \quad s(t) = 1 - t^2 q(t^2), \quad \theta = 1/\|Q\|_2, \quad (4.16)$$

so ist zum Beweis der Glättungseigenschaft wegen

$$\|QS^\nu\|_2 = \frac{1}{\theta} \|\theta Qs(\theta Q)\|_2 = \|p(\theta Q)\|_2 \|Q\|_2, \quad p(t) = ts(t) \quad (4.17)$$

ein geeignetes Polynom p zu konstruieren und die Norm des Polynoms p über dem Spektrum von θQ durch eine Nullfolge $\eta(\nu)$ abzuschätzen.

Aufgrund des Ansatzes ist klar, daß das gesuchte Polynom p den Bedingungen

$$p(0) = 0, \quad p'(0) = 1 \quad (4.18)$$

genügen muß.

Zur Konstruktion eines geeigneten Polynoms p (s. Satz 4.39) werden wir die Chebyshev-Polynome T_m ,

$$T_m(t) = \cos(m \arccos(t)), \quad |t| \leq 1, \quad (4.19)$$

verwenden. Daher erinnern wir zunächst an einige wichtige Eigenschaften der Chebyshev-Polynome:

Satz 4.38. Die Chebyshev-Polynome T_m genügen der Rekursionsvorschrift

$$T_0(t) \equiv 1, \quad T_1(t) = t, \quad T_{m+1}(t) = 2T_m(t) - T_{m-1}(t), \quad m \geq 1. \quad (4.20)$$

Es gilt $|T_m(t)| \leq 1$ für alle $|t| \leq 1$, wobei die angegebene Schranke scharf ist,

$$T_m(\sigma_i) = (-1)^i, \quad \sigma_i = \cos\left(\frac{i\pi}{m}\right), \quad i = 0, \dots, m. \quad (4.21)$$

Des weiteren ist $T'_{2m+1}(0) = (-1)^m(2m+1)$.

Beweis. [38]. □

Wir konstruieren das Polynom p als Lösung der Minimierungsaufgabe (4.22):

Satz 4.39. Das Minimierungsproblem

$$\min_{p \in P_{2m+1}, p(0)=0, p'(0)=1} \max_{-1 \leq t \leq 1} |p(t)| \quad (4.22)$$

hat die eindeutige Lösung

$$p(t) = \frac{(-1)^m}{2m+1} T_{2m+1}(t), \quad |p(t)| \leq \frac{1}{2m+1}, \quad (4.23)$$

wobei T_m die Chebyshev-Polynome sind.

Beweis. Per Konstruktion erfüllt das normalisierte Chebyshev-Polynom

$$p(t) = \frac{(-1)^m}{2m+1} T_{2m+1}(t)$$

die Bedingung 4.18, und es ist durch

$$|p(t)| = \frac{1}{2m+1} |T_{2m+1}(t)| \leq \frac{1}{2m+1}$$

eine geeignete obere Schranke $\eta(\nu)$ gegeben, welche nach (4.21) auch angenommen wird,

$$p(\sigma_i) = (-1)^m \frac{1}{2m+1}, \quad \sigma_i = \cos\left(\frac{i\pi}{2m+1}\right), \quad i = 0, \dots, 2m+1.$$

Es bleibt zu zeigen, daß für jedes andere Polynom p^* vom Grad $\deg(p^*) \leq 2m+1$ das Maximum in (4.22) größer als $1/(2m+1)$ ausfällt. Aus

$$|p^*(\sigma_i)| \leq \frac{1}{2m+1} = |p(\sigma_i)|, \quad i = 0, \dots, 2m+1$$

schließt man für die Differenz $r = p - p^*$ unmittelbar

$$r(\sigma_i) = p(\sigma_i) - p(\sigma_i)^* = \begin{cases} \geq 0 & : \quad i \text{ gerade} \\ \leq 0 & : \quad i \text{ ungerade} \end{cases}, \quad i = 0, \dots, 2m+1,$$

so daß r der Vielfachheit nach $2m+1$ Nullstellen hat. Wegen $r(0) = 0$ hat r eine weitere $2m+2$ -te Nullstelle, und aus $\deg(r) \leq 2m+1$ folgt $r \equiv 0$, also $p = p^*$. \square

In Tabelle 4.2 werden die Glättungsfaktoren η_1 des (modifizierten) Jacobi- und η_2 des semiiterativen Glätters,

$$\eta_1(\nu) = \frac{1}{\sqrt{2\nu+1}} \left(\frac{2\nu}{2\nu+1}\right)^\nu, \quad \eta_2(\nu) = \frac{1}{2\nu+1},$$

einander gegenübergestellt.

Lemma 4.40. Ist p die Lösung der Minimierungsaufgabe (4.22), so erhalten wir als Glätter s

$$s(t) = \frac{1}{t} \frac{(-1)^m}{2m+1} T_{2m+1}(t). \quad (4.24)$$

Beweis. $p(t) = ts(t)$. \square

ν	1	2	3	4	5	6	7	8
$\eta_1(\nu)$	0.385	0.286	0.238	0.208	0.187	0.176	0.159	0.149
$\eta_2(\nu)$	0.333	0.2	0.143	0.111	0.091	0.077	0.067	0.059

Tabelle 4.2: Der Glättungsfaktor im Falle eines modifizierten Jacobi-Glätters η_1 und bei einer semi-iterativen Glättung η_2 .

Die Darstellung (4.24) des Glätters ist aus praktischen Gründen nicht sehr sinnvoll. Statt dessen wäre eine kurze Rekursionsvorschrift ähnlich zu der der Chebyshev-Polynome (4.20) wünschenswert.

Satz 4.41. Die Funktionen s_n ,

$$s_n(t) = \frac{1}{t} \frac{(-1)^n}{2n+1} T_{2n+1}(t),$$

genügen der Rekursionsvorschrift

$$s_n(t) = (\alpha_{n-1}t^2 + \beta_{n-1})s_{n-1}(t) + (1 - \beta_{n-1})s_{n-2}(t), \quad n \geq 1 \quad (4.25)$$

mit

$$\alpha_{n-1} = -4 \frac{2n-1}{2n+1} \quad (n \geq 1), \quad \beta_{n-1} = 2 \frac{2n-1}{2n+1} \quad (n \geq 2),$$

$\beta_0 = 1$ und $s_{-1}(t) = s_0(t) = 1$.

Beweis. Aus der Rekursionsvorschrift der Chebyshev-Polynome folgt

$$\begin{aligned} T_m &= 2tT_{m-1} - T_{m-2} = 2t(2tT_{m-2} - T_{m-3}) - T_{m-2} \\ &= (4t^2 - 1)T_{m-2} - 2tT_{m-3} = (4t^2 - 1)T_{m-2} - (T_{m-2} + T_{m-4}) \\ &= (4t^2 - 2)T_{m-2} - T_{m-4}. \end{aligned}$$

Für $m = 2n + 1$ erhalten wir

$$T_{2n+1} = (4t^2 - 2)T_{2n-1} - T_{2n-3} = (4t^2 - 2)T_{2(n-1)+1} - T_{2(n-2)+1}.$$

Gehen wir mit dieser Identität ein in die Darstellung der Funktion s_n , so gilt:

$$\begin{aligned} \frac{(-1)^n T_{2n+1}}{(2n+1)t} &= \frac{(-1)^n}{(2n+1)t} ((4t^2 - 2)T_{2(n-1)+1} - T_{2(n-2)+1}) \\ &= \left(\frac{(-1)^n (4t^2 - 2)}{(2n+1)t} \frac{2(n-1)+1}{(-1)^{n-1}} \right) \frac{(-1)^{n-1} T_{2(n-1)+1}}{2(n-1)+1} \\ &\quad - \left(\frac{(-1)^n}{(2n+1)t} \frac{2(n-2)+1}{(-1)^{n-2}} \right) \frac{(-1)^{n-2} T_{2(n-2)+1}}{2(n-2)+1}, \end{aligned}$$

also erhalten wir die Rekursionsformel

$$s_n = -\frac{2n-1}{2n+1}(4t^2-2)s_{n-1} - \frac{2n-3}{2n+1}s_{n-2}.$$

Durch Nachrechnen bestätigen wir $s_0 = 1$, $s_1 = 1 - 4/3t^2$. Es ist

$$1 - \beta_{n-1} = 1 - 2\frac{2n-1}{2n+1} = \frac{2n+1 - (4n-2)}{2n+1} = \frac{-2n+3}{2n+1} = -\frac{2n-3}{2n+1}.$$

Die Definition der α_{n-1} ist offensichtlich. Die Wahl $\beta_0 = 1$ ist trivial. \square

Lemma 4.42. Es gilt

$$\begin{aligned} s_2 &= 1 - 4x^2 + \frac{16}{5}x^4 \\ s_3 &= 1 - 8x^2 + 16x^4 - \frac{64}{7}x^6 \\ s_4 &= 1 - \frac{40}{3}x^2 + 48x^4 - 64x^6 + \frac{256}{9}x^8. \end{aligned}$$

Beweis. Anwenden der Rekursionsformel. \square

Satz 4.43. Erfüllen die Funktionen s_n eine Drei-Term-Rekursion der Gestalt (4.25), so gilt für die Iterierten $x^{(n)}$ der zugehörigen Glättungsiteration

$$x^{(n)} = \beta_{n-1}x^{(n-1)} - \alpha_{n-1}\theta^2Q(b - Qx^{(n-1)}) + (1 - \beta_{n-1})x^{(n-2)}. \quad (4.26)$$

Beweis. Es sei $\theta = 1/\|Q\|_2$. Wegen $r^{(n)} = s_n(\theta Q)r^{(0)}$ erhalten wir

$$\begin{aligned} r^{(n)} &= (\alpha_{n-1}\theta^2Q^2 + \beta_{n-1})s_{n-1}(\theta Q)r^{(0)} + (1 - \beta_{n-1})s_{n-2}(\theta Q)r^{(0)} \\ &= (\alpha_{n-1}\theta^2Q^2 + \beta_{n-1})r^{(n-1)} + (1 - \beta_{n-1})r^{(n-2)} \\ &= (\alpha_{n-1}\theta^2Q^2 + \beta_{n-1})(b - Qx^{(n-1)}) + (1 - \beta_{n-1})(b - Qx^{(n-2)}). \end{aligned}$$

Durch Zusammenfassen der Terme folgt

$$\begin{aligned} -Qx^{(n)} &= \alpha_{n-1}\theta^2Q^2b - \alpha_{n-1}\theta^2Q^2Qx^{(n-1)} \\ &\quad - \beta_{n-1}Qx^{(n-1)} - Qx^{(n-2)} + \beta_{n-1}Qx^{(n-2)} \\ &= Q(\alpha_{n-1}\theta^2Qb - \alpha_{n-1}\theta^2Q^2x^{(n-1)} - \beta_{n-1}x^{(n-1)} - x^{(n-2)} + \beta_{n-1}x^{(n-2)}) \\ &= Q(\alpha_{n-1}\theta^2Q(b - Qx^{(n-1)}) - \beta_{n-1}x^{(n-1)} - (1 - \beta_{n-1})x^{(n-2)}), \end{aligned}$$

und somit ist

$$x^{(n)} = \beta_{n-1}x^{(n-1)} - \alpha_{n-1}\theta^2Q(b - Qx^{(n-1)}) + (1 - \beta_{n-1})x^{(n-2)}$$

die Rekursionsvorschrift für die Iterierten. \square

Die Rekursionsvorschrift (4.26) für die hermitesche Schwinger Matrix Q kann zurücktransformiert werden auf entsprechende Größen der originären Schwinger Matrix A :

Lemma 4.44. Es sei $Q = \Gamma_5 A$ und $b = \Gamma_5 \varphi$. Dann gilt

$$x^{(n)} = \beta_{n-1} x^{(n-1)} - \alpha_{n-1} \theta^2 A^H (\varphi - Ax^{(n-1)}) + (1 - \beta_{n-1}) x^{(n-2)}.$$

Beweis. Klar. □

Bemerkung 4.45. Die hier vorgestellte semiiterative Glättung ist eng verwandt mit dem CG-Verfahren als Glättungsiteration, vgl. [37], Lemma 3.3.5. Tatsächlich sind für die hermitesche Schwinger Matrix die Glättungsfaktoren bei n Schritten einer semiiterativen Glättung und bei einer Glättung durch n Schritte des nicht-präkonditionierten CG-Verfahrens für $Q^2 = A^H A$ identisch.

4.6 Zusammenfassung

Analog zur Diskussion der Schurkomplement-Präkonditionierer für die Laplace Matrix können auch für die Schwinger Matrix entsprechende Präkonditionierer erfolgreich konstruiert werden. Exemplarisch wurde die Vorgehensweise für die odd-even reduzierte Schwinger Matrix vorgestellt.

Da im Falle der Schwinger Matrix pro Gitterpunkt zwei Freiheitsgrade existieren, wurde die Charakterisierung der Kopplungen in starke und schwache übertragen auf Blöcke der Größe 2×2 . Es wurde bewiesen, daß bei der verwendeten Charakterisierung der Kopplungen in starke und schwache die Kopplungen zwischen feinen (groben) Gitterpunkten schwach, und daß die Kopplungen zwischen groben (feinen) und feinen (groben) Gitterpunkten stark sind. Somit entfällt auch für die Schwinger Matrix eine aufwendige dynamische Konstruktion der groben und feinen Gitter.

Verglichen wurden wiederum der Jacobi- und der ILU-Präkonditionierer der Matrix $A_{ff}^{(e)}$. Beide Präkonditionierer reflektieren die γ_5 -Symmetrie. Der ILU-Präkonditionierer ist dem Jacobi-Präkonditionierer vorzuziehen. Untersucht wurden zwei ebenfalls γ_5 -symmetrische Schurkomplement-Approximationen, so daß die γ_5 -Symmetrie in Lanczos-basierten Krylov-Unterraum-Verfahren ausgenutzt werden kann. Beide Ansätze sind 9-Punkte-Sterne, und es wurde gezeigt, daß eine rekursive Vorgehensweise zur Konstruktion echter Multilevel-Verfahren ohne Probleme realisierbar ist.

Die numerischen Ergebnisse veranschaulichen, daß die vorgestellten Schurkomplement-Präkonditionierer die Anzahl der BiCGStab-Iterationen für die hier untersuchten Konfigurationen im Mittel um einen Faktor 4-5 (5-6) zu reduzieren vermögen. Für größere Gitter fällt die Verbesserung der Iterationszahlen deutlich höher aus.

Ist ein Schurkomplement-Präkonditionierer der Schwinger Matrix gegeben, so kann in einfacher Weise ein Schurkomplement-Präkonditionierer der hermiteschen Schwinger Matrix bestimmt werden. Für die hermitesche Schwinger Matrix ist der Nachweis der Glättungseigenschaft für eine Reihe von einfachen Iterationsverfahren möglich. Somit ist eine Kombination von Schurkomplement-Präkonditionierern und Glättungsiterationen [49] möglich. Es wurde exemplarisch der Beweis der Glättungseigenschaft für die modifizierte Jacobi-Glättung geführt, anschließend ein semiiterativer Glätter hergeleitet und die Optimalität dieses Glätters bewiesen. Obgleich die semiiterative Glättung relativ teuer ist, steht mit ihr eine optimale Glättungsiteration zur Verfügung. Für die hermitesche Schwinger Matrix ist auch die Glättung durch das CG-Verfahren optimal.

Skizziert wurde außerdem eine einfache Art der (Eich-)Fixierung einer gegebenen Konfiguration, welche zu geometrisch glatteren Eigenvektoren betragskleiner Eigenwerte führt. Der Ansatz wurde nicht weiter verfolgt, da es ein Ziel dieser Arbeit ist, zur Präkonditionierung ungeordneter Systeme rein algebraische Eigenschaften zu verwenden.

Kapitel 5

Wilson-Fermionen

Die Quantenchromodynamik (QCD) ist eine physikalische Theorie zur Beschreibung der Wechselwirkungen zwischen Elementarteilchen als Bausteine der Materie. Mit Hilfe der QCD sollen Eigenschaften von Quarks vorhergesagt und mit Messungen verglichen werden. Eine wesentliche Teilaufgabe ist dabei die Lösung eines sehr großen, jedoch dünn besetzten linearen Gleichungssystems. Die Matrix dieses Systems heißt Wilson Fermion Matrix.

5.1 Einführung

Das Gleichungssystem des Modellproblems beschreibt eine Nächste-Nachbar-Kopplung auf einem 4-dimensionalen Raum-Zeit-Gitter

$$\Omega = \{x = (x_1, x_2, x_3, x_4) \mid x_i = 1, \dots, n_i, \quad n_i = 2^{N_i}\} \quad (5.1)$$

mit periodischer Fortsetzung an den Rändern. Bezeichnen wir die Gitterdimensionen mit n_1, n_2, n_3 (Raum) und n_4 (Zeit), so gibt es insgesamt $n_1 n_2 n_3 n_4$ Gitterpunkte. Aufgrund der Periodizität sind Komponenten von x außerhalb des jeweiligen Bereiches modulo Indexgrenze zu verstehen.

Auf jedem Gitterpunkt x gibt es zwölf Variablen, die zu einem Vektor $\psi(x) \in \mathbb{C}^{12}$ zusammengefaßt werden. Diese 12 Komponenten entsprechen in der Physik den möglichen Kombinationen von 3 Farben und 4 Dirac-Komponenten. Jede Nächste-Nachbar-Kopplung wird beschrieben durch eine Matrix $U \in SU(3)$, die auf die Farben wirkt, und eine Matrix der Form $I \pm \gamma_\mu \in \mathbb{C}^{4 \times 4}$, die auf die Dirac-Komponenten wirkt.

Wir setzen im folgenden stets voraus, daß N_i hinreichend groß ist. Insbesondere gibt es also zu jedem Gitterpunkt x genau 8 verschiedene nächste Nachbarn.

Definition 5.1. Die Matrizen $\gamma_i \in \mathbb{C}^{4 \times 4}$ für $i = 1, \dots, 4$ sind wie folgt definiert:

$$\gamma_1 = \begin{pmatrix} 0 & 0 & 0 & i \\ 0 & 0 & i & 0 \\ 0 & -i & 0 & 0 \\ -i & 0 & 0 & 0 \end{pmatrix}, \quad \gamma_2 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix},$$

$$\gamma_3 = \begin{pmatrix} 0 & 0 & i & 0 \\ 0 & 0 & 0 & -i \\ -i & 0 & 0 & 0 \\ 0 & i & 0 & 0 \end{pmatrix}, \quad \gamma_4 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}.$$

Definition 5.2. Eine Konfiguration ist eine Sammlung von $4n_1n_2n_3n_4$ Matrizen $U_\mu(x) \in SU(3)$, wobei x Gitterpunkt und $\mu \in \{1, 2, 3, 4\}$ ist. Die Konfiguration heißt kalt, falls alle $U_\mu(x) = I$ sind. Sie heißt heiß, falls alle $U_\mu(x)$ gleichverteilt sind. Eine Konfiguration, die weder kalt noch heiß ist, heißt realistisch.

Definition 5.3. Das lineare Gleichungssystem des Modellproblems für eine gegebene Konfiguration und den reellen Parameter $\kappa \geq 0$ ist gegeben durch

$$\psi(x) - \kappa \left(\sum_{\mu=1}^4 [(I - \gamma_\mu) \otimes U_\mu(x)] \psi(x + e_\mu) + \sum_{\mu=1}^4 [(I + \gamma_\mu) \otimes U_\mu^H(x - e_\mu)] \psi(x - e_\mu) \right) = \varphi(x) \quad (5.2)$$

für $x \in \{1, \dots, n_1\} \times \dots \times \{1, \dots, n_4\}$.

Durch (5.2) sind $n = 12n_1n_2n_3n_4$ Gleichungen für $n_1n_2n_3n_4$ Unbekannte $\psi(x) \in \mathbb{C}^{12}$ gegeben. Die rechte Seite φ wird auch als Quelle bezeichnet.

Die numerischen Berechnungen der QCD gehören zu den rechenintensivsten Anwendungen überhaupt. Bei heutigen physikalisch bedeutsamen Simulationen, s. z.B. [52], werden sukzessive $\mathcal{O}(10^6)$ Gleichungssysteme der Gestalt (5.2) auf einem $64^3 \times 112$ -Gitter gelöst.

5.2 Eigenschaften

Wir diskutieren nun völlig analog zu Kapitel 4.2 dieser Arbeit einige wichtige Eigenschaften der Wilson Fermion Matrix.

Anordnung der Gitterpunkte

Wir übertragen die Definition der lexikographischen und der odd-even Anordnung der Gitterpunkte auf das gegebene 4-dimensionale Gitter Ω aus (5.1):

Definition 5.4. Für zwei Gitterpunkte $x = (x_1, x_2, x_3, x_4)$ und $y = (y_1, y_2, y_3, y_4)$ gelte $x < y$ genau dann, wenn entweder $x_1 < y_1$ ist oder es ein i_0 gibt, so daß $x_i = y_i$ für alle $i = 1, \dots, i_0 - 1$ und $x_{i_0} < y_{i_0}$ gilt.

Die Anordnung einer Menge von Gitterpunkten bezüglich der Relation „ $<$ “ heißt lexikographische Anordnung.

Definition 5.5. Ein Gitterpunkt $x = (x_1, x_2, x_3, x_4)$ heißt gerade (even), falls die Summe seiner Komponenten gerade ist,

$$x_1 + x_2 + x_3 + x_4 \equiv 0 \pmod{2},$$

ansonsten heißt er ungerade (odd).

Werden zuerst alle ungeraden, dann alle geraden Gitterpunkte gezählt, so ist heißt die resultierende Anordnung der Gitterpunkte auch odd-even Anordnung.

Odd-even Symmetrie

Bei einer odd-even Anordnung der Gitterpunkte bekommt die Matrix des Modellproblems mit $I \in \mathbb{C}^{n \times n}$ und $D_{eo}, D_{oe} \in \mathbb{C}^{\frac{n}{2} \times \frac{n}{2}}$ die Gestalt

$$A = I - \kappa D, \quad D = \begin{pmatrix} 0 & D_{oe} \\ D_{eo} & 0 \end{pmatrix}, \quad (5.3)$$

wodurch sich mit $\Psi = (\psi(x))$ und $\Phi = (\varphi(x))$ das lineare Gleichungssystem

$$A\Psi = \Phi \quad (5.4)$$

ergibt. Die Matrix A wird als Wilson Fermion Matrix bezeichnet. Der Nicht-diagonalanteil D heißt Hopping-Term, das Gewicht κ Hopping-Parameter.

Die γ_5 -Symmetrie

Definition 5.6. Die Matrix $\gamma_5 \in \mathbb{C}^{4 \times 4}$ ist definiert durch

$$\gamma_5 = -\gamma_4\gamma_1\gamma_2\gamma_3 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}. \quad (5.5)$$

Lemma 5.7. Es gilt

- a) $\gamma_5 = \gamma_5^H = \gamma_5^T = \gamma_5^{-1}$
- b) $\gamma_5\gamma_\mu\gamma_5 = -\gamma_\mu^H$ für alle $\mu = 1, \dots, 4$

Beweis. Nachrechnen. □

Definition 5.8. Die Matrix Γ_5 ist definiert durch $\Gamma_5 = I \otimes (\gamma_5 \otimes I_3)$.

Γ_5 ist die Permutation, welche auf jedem Gitterpunkt für jede Farbe die Dirac-Komponenten 1 mit 3 und 2 mit 4 vertauscht.

Satz 5.9. Die Wilson Fermion Matrix ist γ_5 -symmetrisch, $A\Gamma_5 = \Gamma_5A^H$.

Beweis. Analog zum Beweis von Satz 4.16. □

Satz 5.10. Die Wilson Fermion Matrix ist i.a. nicht normal, d.h. $A^HA \neq AA^H$.

Beweis. Analog zum Beweis von Satz 4.17. □

Das Spektrum

Definition 5.11. Der kritische Wert κ_c ist definiert als $\kappa_c = 1 / \max_{\lambda \in \sigma(D)} \operatorname{Re} \lambda$.

Die Wilson Fermion Matrix $A(\kappa) = I - \kappa D$ ist positiv reell für $\kappa \in [0, \kappa_c)$ mit $\kappa_c \in [\frac{1}{8}, \frac{1}{4}]$.

Physikalisch interessant ist die Lösung des Modellproblems (5.4) für $\kappa \approx \kappa_c$. Diese Wahl entspricht einer kleinen Quarkmasse auf dem Gitter, vgl. (4.3).

Aufgrund der odd-even Symmetrie und der γ_5 -Symmetrie gilt:

Satz 5.12. Ist λ ein Eigenwert von A , so sind auch $\bar{\lambda}$, $2 - \lambda$ und $2 - \bar{\lambda}$ Eigenwerte von A .

Beweis. Vgl. Abschnitt 4.2. □

Abbildung 5.1 zeigt das Spektrum der Wilson Fermion Matrix und der odd-even reduzierten Matrix für eine realistische Konfiguration.

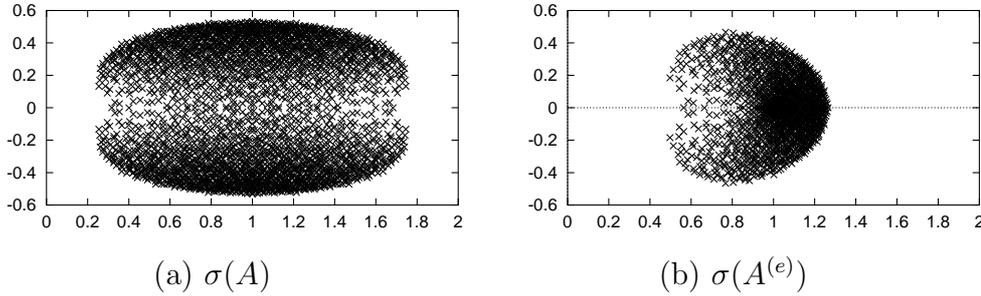


Abbildung 5.1: Das Spektrum der Wilson Fermion Matrix A und der odd-even reduzierten Wilson Fermion Matrix $A^{(e)}$ für eine realistische Konfiguration mit einem moderaten Wert $\kappa \ll \kappa_c$.

5.3 Schurkomplement-Verfahren

In diesem Abschnitt werden wir erklären, warum sich die in Kapitel 3 und Kapitel 4 hergeleiteten Schurkomplement-Präkonditionierer der odd-even reduzierten Matrix nicht sinnvoll auf die odd-even reduzierte Wilson Fermion Matrix übertragen lassen. Statt dessen werden Schurkomplement-Präkonditionierer für die originäre Wilson Fermion Matrix bestimmt.

Anordnung der Gitterpunkte

Wir verallgemeinern die in Def. 2.13 eingeführte Einfärbung der Gitterpunkte wie folgt:

Definition 5.13. Für $i = 0, \dots, 4$ sei die Menge $L_i(\Omega)$ von Gitterpunkten definiert durch

$$L_i(\Omega) = \{x = (x_1, x_2, x_3, x_4) \in \Omega \mid x \text{ besitzt genau } i \text{ gerade Komponenten}\}.$$

Im folgenden gehen wir stets davon aus, daß zunächst alle ungeraden und anschließend alle geraden Gitterpunkte gezählt werden. Dabei werden die Gitterpunkte aus $L_1(\Omega)$ vor den Gitterpunkten aus $L_3(\Omega)$, die Gitterpunkte aus $L_2(\Omega)$ vor den Gitterpunkten aus $L_0(\Omega)$ und die Gitterpunkte aus $L_0(\Omega)$ vor den Gitterpunkten aus $L_4(\Omega)$ angeordnet.

Bei dieser Anordnung der Gitterpunkte besitzt die Wilson Fermion Matrix A die Gestalt

$$A = I - \kappa \begin{pmatrix} 0 & D_{oe} \\ D_{eo} & 0 \end{pmatrix}$$

mit

$$D_{eo} = \begin{pmatrix} D_{L_2, L_1} & D_{L_2, L_3} \\ D_{L_0, L_1} & 0 \\ 0 & D_{L_4, L_3} \end{pmatrix}, \quad D_{oe} = \begin{pmatrix} D_{L_1, L_2} & D_{L_1, L_0} & 0 \\ D_{L_3, L_2} & 0 & D_{L_3, L_4} \end{pmatrix}.$$

Lemma 5.14. Es sei N gerade, $n_i = N$ für $i = 1, \dots, 4$ und Ω mit $|\Omega| = n$ Gitterpunkten gegeben. Dann gilt

$$|L_0(\Omega)| = |L_4(\Omega)| = \frac{1}{16}n, \quad |L_1(\Omega)| = |L_3(\Omega)| = \frac{1}{4}n, \quad |L_2(\Omega)| = \frac{3}{8}n.$$

Beweis. Abzählen. □

Die odd-even reduzierte Matrix

Lemma 5.15. Es gilt

$$D_{eo}D_{oe} = \begin{pmatrix} D_{L_2, L_1}D_{L_1, L_2} + D_{L_2, L_3}D_{L_3, L_2} & D_{L_2, L_1}D_{L_1, L_0} & D_{L_2, L_3}D_{L_3, L_4} \\ D_{L_0, L_1}D_{L_1, L_2} & D_{L_0, L_1}D_{L_1, L_0} & 0 \\ D_{L_4, L_3}D_{L_3, L_2} & 0 & D_{L_4, L_3}D_{L_3, L_4} \end{pmatrix}.$$

Beweis. Nachrechnen. □

Lemma 5.16. Die odd-even reduzierte Matrix $A^{(e)} = I - \kappa^2 D_{eo}D_{oe}$ beschreibt einen 33-Punkte-Stern.

Beweis. Der Gitterpunkt x koppelt mit sich selbst, seinen 8 Nachbarn $x \pm 2e_\mu$ in Gitterrichtung μ sowie für $\mu < \nu$ mit je 4 Gitterpunkten $x \pm e_\mu \pm e_\nu$ in insgesamt 6 verschiedenen $\mu\nu$ -Ebenen. □

Wir bestimmen die einzelnen Kopplungen der beteiligten Teilmatrizen genauer:

Lemma 5.17. Die Blockdiagonale (mit Blöcken der Größe 12×12) der Matrix $D_{eo}D_{oe}$ verschwindet.

Beweis. Analog zum Beweis von Lemma 4.21. □

Lemma 5.18. Die Matrizen $D_{L_0, L_1}D_{L_1, L_0}$ und $D_{L_4, L_3}D_{L_3, L_4}$ besitzen genau 8 Nichtnullblöcke in der Blockzeile x ($\mu = 1, \dots, 4$),

$$\begin{aligned} (x, x + 2e_\mu) &: 2(I - \gamma_\mu) \otimes U_\mu(x)U_\mu(x + e_\mu) \\ (x, x - 2e_\mu) &: 2(I + \gamma_\mu) \otimes U_\mu^H(x - e_\mu)U_\mu^H(x - 2e_\mu). \end{aligned}$$

Beweis. Analog zum Beweis von Lemma 4.22. □

Die Matrizen $D_{L_0, L_1} D_{L_1, L_0}$ und $D_{L_4, L_3} D_{L_3, L_4}$ beschreiben nach Lemma 5.18 die Kopplung eines Gitterpunktes x mit seinen nächsten Nachbarn in Gitterrichtung $\mu = 1, \dots, 4$. Somit können $D_{L_0, L_1} D_{L_1, L_0}$ und $D_{L_4, L_3} D_{L_3, L_4}$ jeweils durch einen 9-Punkte-Stern repräsentiert werden.

Gemäß Lemma 5.16 und Lemma 5.17 koppelt jeder Gitterpunkt x via $D_{eo} D_{oe}$ mit insgesamt 32 anderen Gitterpunkten. Zusammen mit Lemma 5.18 ist daher klar, daß die Matrizen $D_{L_0, L_1} D_{L_1, L_2}$ und $D_{L_4, L_3} D_{L_3, L_2}$ jeweils Kopplungen eines Gitterpunktes x mit 24 Nachbarn aus $L_2(\Omega)$ beschreiben. Wir präzisieren:

Lemma 5.19. Die Matrizen $D_{L_0, L_1} D_{L_1, L_2}$ und $D_{L_4, L_3} D_{L_3, L_2}$ besitzen jeweils genau 24 Nichtnullblöcke in der Blockzeile x ,

$$\begin{aligned}
(x, x + e_\mu + e_\nu) &: (I - \gamma_\mu)(I - \gamma_\nu) \otimes U_\mu(x)U_\nu(x + e_\mu) \\
&\quad + (I - \gamma_\nu)(I - \gamma_\mu) \otimes U_\nu(x)U_\mu(x + e_\nu) \\
(x, x + e_\mu - e_\nu) &: (I - \gamma_\mu)(I + \gamma_\nu) \otimes U_\mu(x)U_\nu^H(x + e_\mu - e_\nu) \\
&\quad + (I + \gamma_\nu)(I - \gamma_\mu) \otimes U_\nu^H(x - e_\nu)U_\mu(x - e_\nu) \\
(x, x - e_\mu + e_\nu) &: (I + \gamma_\mu)(I - \gamma_\nu) \otimes U_\mu^H(x - e_\mu)U_\nu(x - e_\mu) \\
&\quad + (I - \gamma_\nu)(I + \gamma_\mu) \otimes U_\nu(x)U_\mu^H(x - e_\mu + e_\nu) \\
(x, x - e_\mu - e_\nu) &: (I + \gamma_\mu)(I + \gamma_\nu) \otimes U_\mu^H(x - e_\mu)U_\nu^H(x - e_\mu - e_\nu) \\
&\quad + (I + \gamma_\nu)(I + \gamma_\mu) \otimes U_\nu^H(x - e_\nu)U_\mu^H(x - e_\mu - e_\nu),
\end{aligned}$$

wobei $\mu < \nu$ ist.

Beweis. Analog zum Beweis von Lemma 4.23. □

Um die Diskussion der verbleibenden Kopplungen zu vereinfachen, führen wir eine Unterteilung der Gitterpunkte $L_2(\Omega)$ in sechs disjunkte Mengen ein:

Definition 5.20. Für $\mu < \nu$ sei die Menge $L_2^{(\mu, \nu)}(\Omega)$ von Gitterpunkten definiert als

$$L_2^{(\mu, \nu)}(\Omega) = \{x = (x_1, x_2, x_3, x_4) \in L_2(\Omega) \mid x_\mu, x_\nu \text{ gerade}\}.$$

Sind die Komponenten μ und ν des Gitterpunktes $x \in L_2(\Omega)$ gerade, so koppelt x mit insgesamt 4 Gitterpunkten aus $L_0(\Omega)$:

Lemma 5.21. Die Matrix $D_{L_2, L_1} D_{L_1, L_0}$ besitzt genau 4 Nichtnullblöcke in der Blockzeile $x \in L_2^{(\mu, \nu)}(\Omega)$, $\mu < \nu$,

$$\begin{aligned}
(x, x + e_\mu + e_\nu) &: (I - \gamma_\mu)(I - \gamma_\nu) \otimes U_\mu(x)U_\nu(x + e_\mu) \\
&\quad + (I - \gamma_\nu)(I - \gamma_\mu) \otimes U_\nu(x)U_\mu(x + e_\nu) \\
(x, x + e_\mu - e_\nu) &: (I - \gamma_\mu)(I + \gamma_\nu) \otimes U_\mu(x)U_\nu^H(x + e_\mu - e_\nu) \\
&\quad + (I + \gamma_\nu)(I - \gamma_\mu) \otimes U_\nu^H(x - e_\nu)U_\mu(x - e_\nu) \\
(x, x - e_\mu + e_\nu) &: (I + \gamma_\mu)(I - \gamma_\nu) \otimes U_\mu^H(x - e_\mu)U_\nu(x - e_\mu) \\
&\quad + (I - \gamma_\nu)(I + \gamma_\mu) \otimes U_\nu(x)U_\mu^H(x - e_\mu + e_\nu) \\
(x, x - e_\mu - e_\nu) &: (I + \gamma_\mu)(I + \gamma_\nu) \otimes U_\mu^H(x - e_\mu)U_\nu^H(x - e_\mu - e_\nu) \\
&\quad + (I + \gamma_\nu)(I + \gamma_\mu) \otimes U_\nu^H(x - e_\nu)U_\mu^H(x - e_\mu - e_\nu).
\end{aligned}$$

Beweis. Analog zum Beweis von Lemma 4.23. \square

Eine entsprechende Aussage gilt natürlich auch für die Matrix $D_{L_2, L_3} D_{L_3, L_4}$:

Lemma 5.22. Die Matrix $D_{L_2, L_3} D_{L_3, L_4}$ besitzt genau 4 Nichtnullblöcke in der Blockzeile $x \in L_2^{(\mu', \nu')}(\Omega)$. Ist $\mu, \nu \in \{1, 2, 3, 4\} \setminus \{\mu', \nu'\}$ mit $\mu < \nu$, so sind die Nichtnullblöcke an den Positionen

$$\begin{aligned}
(x, x + e_\mu + e_\nu) &: (I - \gamma_\mu)(I - \gamma_\nu) \otimes U_\mu(x)U_\nu(x + e_\mu) \\
&\quad + (I - \gamma_\nu)(I - \gamma_\mu) \otimes U_\nu(x)U_\mu(x + e_\nu) \\
(x, x + e_\mu - e_\nu) &: (I - \gamma_\mu)(I + \gamma_\nu) \otimes U_\mu(x)U_\nu^H(x + e_\mu - e_\nu) \\
&\quad + (I + \gamma_\nu)(I - \gamma_\mu) \otimes U_\nu^H(x - e_\nu)U_\mu(x - e_\nu) \\
(x, x - e_\mu + e_\nu) &: (I + \gamma_\mu)(I - \gamma_\nu) \otimes U_\mu^H(x - e_\mu)U_\nu(x - e_\mu) \\
&\quad + (I - \gamma_\nu)(I + \gamma_\mu) \otimes U_\nu(x)U_\mu^H(x - e_\mu + e_\nu) \\
(x, x - e_\mu - e_\nu) &: (I + \gamma_\mu)(I + \gamma_\nu) \otimes U_\mu^H(x - e_\mu)U_\nu^H(x - e_\mu - e_\nu) \\
&\quad + (I + \gamma_\nu)(I + \gamma_\mu) \otimes U_\nu^H(x - e_\nu)U_\mu^H(x - e_\mu - e_\nu).
\end{aligned}$$

Beweis. Analog zum Beweis von Lemma 4.23. \square

Um die Kopplungen der Matrix $D_{L_2, L_1} D_{L_1, L_2} + D_{L_2, L_3} D_{L_3, L_2}$ zu beschreiben, ist das folgende Lemma nützlich:

Lemma 5.23. Es sei $\mu_1 < \mu_2$ und $\mu_3, \mu_4 \in \{1, 2, 3, 4\} \setminus \{\mu_1, \mu_2\}$ mit $\mu_3 < \mu_4$. Dann koppeln Gitterpunkte aus $L_2^{(\mu_1, \mu_2)}(\Omega)$ nicht mit Gitterpunkten aus $L_2^{(\mu_3, \mu_4)}(\Omega)$.

Beweis. Offensichtlich. \square

Nun gilt:

Lemma 5.24. Die Matrix $D_{L_2, L_1} D_{L_1, L_2} + D_{L_2, L_3} D_{L_3, L_2}$ beschreibt einen 24-Punkte-Stern. In der Blockzeile $x \in L_2^{(\mu_1, \mu_2)}(\Omega)$ gibt es 8 Nichtnullblöcke an den Positionen ($\mu = 1, \dots, 4$)

$$\begin{aligned} (x, x + 2e_\mu) &: 2(I - \gamma_\mu) \otimes U_\mu(x) U_\mu(x + e_\mu) \\ (x, x - 2e_\mu) &: 2(I + \gamma_\mu) \otimes U_\mu^H(x - e_\mu) U_\mu^H(x - 2e_\mu) \end{aligned}$$

und 16 Nichtnullblöcke an den Positionen

$$\begin{aligned} (x, x + e_\mu + e_\nu) &: (I - \gamma_\mu)(I - \gamma_\nu) \otimes U_\mu(x) U_\nu(x + e_\mu) \\ &\quad + (I - \gamma_\nu)(I - \gamma_\mu) \otimes U_\nu(x) U_\mu(x + e_\nu) \\ (x, x + e_\mu - e_\nu) &: (I - \gamma_\mu)(I + \gamma_\nu) \otimes U_\mu(x) U_\nu^H(x + e_\mu - e_\nu) \\ &\quad + (I + \gamma_\nu)(I - \gamma_\mu) \otimes U_\nu^H(x - e_\nu) U_\mu(x - e_\nu) \\ (x, x - e_\mu + e_\nu) &: (I + \gamma_\mu)(I - \gamma_\nu) \otimes U_\mu^H(x - e_\mu) U_\nu(x - e_\mu) \\ &\quad + (I - \gamma_\nu)(I + \gamma_\mu) \otimes U_\nu(x) U_\mu^H(x - e_\mu + e_\nu) \\ (x, x - e_\mu - e_\nu) &: (I + \gamma_\mu)(I + \gamma_\nu) \otimes U_\mu^H(x - e_\mu) U_\nu^H(x - e_\mu - e_\nu) \\ &\quad + (I + \gamma_\nu)(I + \gamma_\mu) \otimes U_\nu^H(x - e_\nu) U_\mu^H(x - e_\mu - e_\nu), \end{aligned}$$

wobei $\mu \in \{\mu_1, \mu_2\}$ und $\nu \in \{1, 2, 3, 4\} \setminus \{\mu_1, \mu_2\}$ ist.

Beweis. Analog zum Beweis von Lemma 4.22 und Lemma 4.23. \square

Abbildung 5.2(a) zeigt das Besetztheitsmuster der odd-even reduzierten Wilson Fermion Matrix bei einer Anordnung

$$L_2(\Omega) - L_0(\Omega) - L_4(\Omega),$$

wobei die Gitterpunkte innerhalb der Mengen jeweils lexikographisch angeordnet sind. (Die Notation $L_i(\Omega) - L_j(\Omega)$ bedeutet, daß alle Gitterpunkte aus $L_i(\Omega)$ vor den Gitterpunkten aus $L_j(\Omega)$ gezählt werden.) In Abbildung 5.2(b) ist ebenfalls das Besetztheitsmuster der odd-even reduzierten Wilson Fermion Matrix zu sehen. Die Gitterpunkte sind nun aber in der Anordnung

$$\begin{aligned} &L_2^{(1,2)}(\Omega) - L_2^{(3,4)}(\Omega) - L_2^{(1,3)}(\Omega) - L_2^{(2,4)}(\Omega) \\ &- L_2^{(1,4)}(\Omega) - L_2^{(2,3)}(\Omega) - L_0(\Omega) - L_4(\Omega) \end{aligned}$$

gegeben. Zusätzlich wurde innerhalb der Mengen eine odd-even Anordnung der (stets geraden) Gitterpunkte verwendet, wobei Gitterpunkte x mit $x_1 + x_2 + x_3 + x_4 \equiv 2 \pmod{4}$ odd und Gitterpunkte x mit $x_1 + x_2 + x_3 + x_4 \equiv 0 \pmod{4}$ even genannt werden.

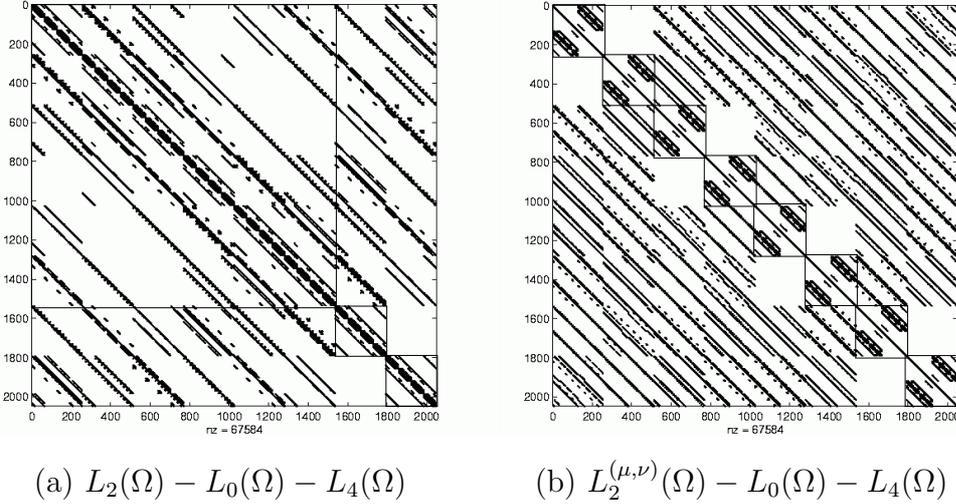


Abbildung 5.2: Das Besetzungsmuster der odd-even reduzierten Wilson Fermion Matrix $A^{(e)}$.

Wie in Kapitel 4 können die Kopplungen des Gitterpunktes x mit seinen 8 Nachbarn $x \pm 2e_\mu$ als schwach und die verbleibenden 24 Kopplungen als stark charakterisiert werden. In Abbildung 5.2(b) bilden die schwachen Kopplungen die Blockdiagonale, die starken Kopplungen befinden sich außerhalb der Blockdiagonalen. Da wir aber im folgenden aufzeigen werden, daß der in Kapitel 4 erfolgreich eingesetzte Schurkomplement-Präkonditionierer sich nicht sinnvoll auf die odd-even reduzierte Wilson Fermion Matrix übertragen läßt, werden wir nicht näher auf die Einteilung in starke und schwache Kopplungen eingehen.

Lemma 5.25. Mit den Bezeichnungen aus Lemma 5.23 sind die Mengen $L_0(\Omega) \cup L_4(\Omega)$ und $L_2^{(\mu_1, \mu_2)}(\Omega) \cup L_2^{(\mu_3, \mu_4)}(\Omega)$ jeweils maximal unabhängige Knotenmengen des reduzierten Graphen $G_{A^{(e)}}(V, E_S^*)$.

Beweis. Klar. □

Der in Kapitel 4 erfolgreich eingesetzte Schurkomplement-Präkonditionierer würde im Falle der odd-even reduzierten Wilson Fermion Matrix für alle unabhängigen Knotenmengen aus Lemma 5.25 lediglich auf einer Menge von $n/8$ feinen Gitterpunkten basieren. Das Grobgitter bestünde aus insgesamt $3n/8$ Gitterpunkten, und das exakte Schurkomplement wäre (oBdA mit $L_2(\Omega)$ als grobes Gitter) von der Gestalt

$$\begin{aligned}
& I - \kappa^2 D_{L_2, L_1} D_{L_1, L_2} \\
& - \kappa^4 \left(D_{L_2, L_1} D_{L_1, L_0} (I - \kappa^2 D_{L_0, L_1} D_{L_1, L_0})^{-1} D_{L_0, L_1} D_{L_1, L_2} \right. \\
& \quad \left. + D_{L_2, L_3} D_{L_3, L_4} (I - \kappa^2 D_{L_4, L_3} D_{L_3, L_4})^{-1} D_{L_4, L_3} D_{L_3, L_2} \right).
\end{aligned}$$

Bei einer Approximation der Inversen durch eine Diagonalmatrix beschreibt die resultierende Schurkomplement-Approximation einen 145-Punkte-Stern! Da der Stern selbst bei dieser einfachen Approximation der Inversen viel zu groß ist, ist der Schurkomplement-Präkonditionierer für die odd-even reduzierte Wilson Fermion Matrix nicht praktikabel.

Also betrachten wir statt dessen ab jetzt die originäre Wilson Fermion Matrix.

F/C-Anordnung der Gitterpunkte

Im folgenden wollen wir Schurkomplement-Präkonditionierer der Wilson Fermion Matrix mit

$$\Omega_c = L_0(\Omega) \cup L_4(\Omega), \quad \Omega_f = \Omega \setminus \Omega_c \quad (5.6)$$

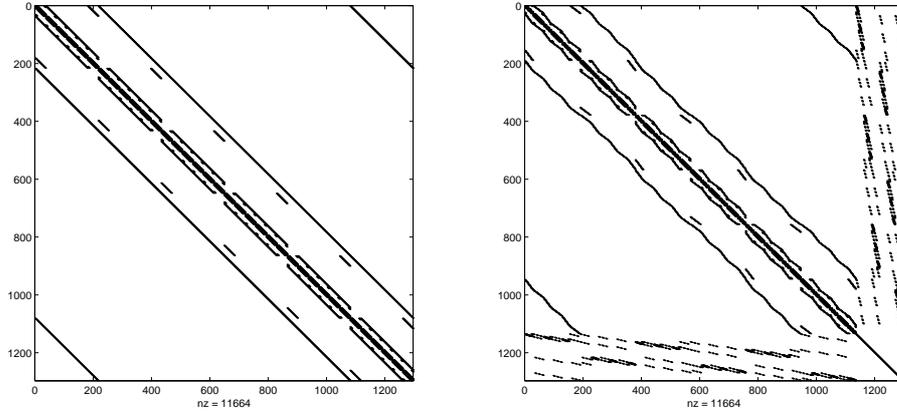
konstruieren [48]. Diese Wahl der groben und feinen Gitterpunkte entspricht einer Partition der Hopping-Matrix D mit

$$D_{ff} = \begin{pmatrix} 0 & 0 & D_{L_1, L_2} \\ 0 & 0 & D_{L_3, L_2} \\ D_{L_2, L_1} & D_{L_2, L_3} & 0 \end{pmatrix}$$

und

$$D_{fc} = \begin{pmatrix} D_{L_1, L_0} & 0 \\ 0 & D_{L_3, L_4} \\ 0 & 0 \end{pmatrix}, \quad D_{cf} = \begin{pmatrix} D_{L_0, L_1} & 0 & 0 \\ 0 & D_{L_4, L_3} & 0 \end{pmatrix}, \quad D_{cc} = 0.$$

Abbildung 5.3 zeigt das Besetztheitsmuster der Wilson Fermion Matrix, wobei hier die feinen Gitterpunkte in einer lexikographischen Anordnung vorliegen.



(a) lexikographisch

(b) F/C-Anordnung

Abbildung 5.3: Das Besetzmuster der Wilson Fermion Matrix A bei einer lexikographischen und einer F/C-Anordnung (5.6).

Restriktion und Prolongation

Zur Konstruktion einer matrixabhängigen Restriktion und Prolongation benötigen wir eine approximative Inverse der Teilmatrix $A_{ff} = I - \kappa D_{ff}$. Da die Matrix A_{ff} eine γ_5 -symmetrische Matrix ist, beschränken wir uns auf approximative Inverse, welche die γ_5 -Symmetrie reflektieren,

$$\tilde{A}_{ff}^{-1} = \Gamma_5 \tilde{A}_{ff}^{-H} \Gamma_5. \quad (5.7)$$

Auf den ersten Blick ist es nicht notwendig, daß für die Restriktion und die Prolongation identische approximative Inverse gewählt werden,

$$R = \begin{pmatrix} -A_{cf} \tilde{A}_{ff}^{-1} & I \end{pmatrix}, \quad P = \begin{pmatrix} -\tilde{A}_{ff}^{-1} A_{fc} \\ I \end{pmatrix}, \quad (5.8)$$

jedoch erscheint dieser Ansatz natürlich, da zusammen mit (5.7) die Restriktion und Prolongation zueinander γ_5 -adjungiert sind,

$$\Gamma_5 R^H = P \Gamma_5. \quad (5.9)$$

Satz 5.26. Für $0.072 \leq \kappa < \kappa_c$ ist die Matrix $A_{ff} = I - \kappa D_{ff}$ nicht diagonaldominant.

Beweis. Die Matrix D_{ff} besitzt in der Blockzeile $x \in L_2$ genau 8 Nichtnullblöcke an den Positionen ($\mu = 1, \dots, 4$)

$$\begin{aligned} (x, x + e_\mu) &: (I - \gamma_\mu) \otimes U_\mu(x) \\ (x, x - e_\mu) &: (I + \gamma_\mu) \otimes U_\mu^H(x - e_\mu). \end{aligned}$$

Wir erhalten

$$\begin{aligned} \sum_{y \neq x} |(D_{ff})_{x,y}| &= \sum_{\mu=1}^4 |(I - \gamma_\mu) \otimes U_\mu(x)| + |(I + \gamma_\mu) \otimes U_\mu^H(x - e_\mu)| \\ &= \begin{pmatrix} A(x) & 0 & C(x) & D(x) \\ 0 & A(x) & D(x) & C(x) \\ C(x) & D(x) & B(x) & 0 \\ D(x) & C(x) & 0 & B(x) \end{pmatrix} \end{aligned}$$

mit

$$\begin{aligned} A(x) &= \sum_{\mu=1}^3 |U_\mu(x)| + |U_\mu^H(x - e_\mu)| + 2|U_4^H(x - e_4)|, \\ B(x) &= \sum_{\mu=1}^3 |U_\mu(x)| + |U_\mu^H(x - e_\mu)| + 2|U_4(x)| \end{aligned}$$

und

$$C(x) = |U_3(x)| + |U_3^H(x - e_3)|, \quad D(x) = \sum_{\mu=1}^2 |U_\mu(x)| + |U_\mu^H(x - e_\mu)|.$$

Da die beteiligten Matrizen $U_\mu(x)$ unitär sind, gilt für die Zeilensummen der Betragsmatrizen

$$\sum_{j=1}^3 |U_\mu(x)|_{ij} \geq 1, \quad \sum_{j=1}^3 |U_\mu^H(x - e_\mu)|_{ij} \geq 1, \quad i = 1, \dots, 3.$$

Für $i = 1, \dots, 3$ folgt mit $A(x) = (A_{ij}(x))$, $B(x) = (B_{ij}(x))$, $C(x) = (C_{ij}(x))$ und $D(x) = (D_{ij}(x))$

$$\sum_{j=1}^3 A_{ij}(x) \geq 8, \quad \sum_{j=1}^3 B_{ij}(x) \geq 8, \quad \sum_{j=1}^3 C_{ij}(x) \geq 2, \quad \sum_{j=1}^3 D_{ij}(x) \geq 4.$$

Daher ist für $\kappa \geq 1/14$ wegen

$$\sum_{j \neq i} |(A_{ff})_{i,j}| = \kappa \sum_{j \neq i} |(D_{ff})_{i,j}| \geq 14\kappa \geq 1$$

die Matrix A_{ff} nicht diagonaldominant. \square

Aus Satz 5.26 folgern wir, daß auch die Wilson Fermion Matrix A für $0.072 \leq \kappa < \kappa_c$ nicht diagonaldominant ist.

Obwohl die Matrix A_{ff} nicht diagonaldominant ist, beobachten wir in numerischen Experimenten, daß die Beträge der Komponenten der Inversen der Matrix A_{ff} mit zunehmender Entfernung von der Diagonalen exponentiell fallen, vgl. Abb. 5.4. Wir schließen daher, daß die Matrix A_{ff} gut konditioniert.

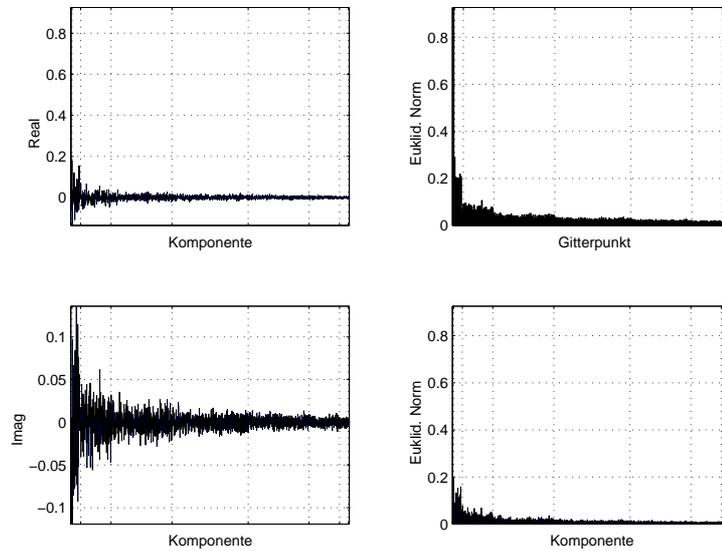


Abbildung 5.4: Die Lösung Ψ_f des Systems $A_{ff}\Psi_f = \Phi_f$ mit $\Phi_f(x^*) = e_1$ und $\Phi_f(x) = 0$ für alle $x \neq x^*$ bei einer levelweisen Anordnung der Gitterpunkte. Definieren wir die Mengen M_l induktiv durch (a) $M_0 = \{x^*\}$ und (b) $x \in M_l$, falls $x \notin M_k$, $k < l$, und es existiert ein $y \in M_{l-1} \cap N(x)$, so bedeutet eine levelweise Anordnung gerade, daß alle Gitterpunkte aus M_l vor Gitterpunkten aus M_{l+1} angeordnet werden.

Diese Beobachtung rechtfertigt eine Approximation der Inversen der Matrix A_{ff} durch eine trunkierte Neumann-Reihe. Damit zum einen Gitterpunkte aus $L_2(\Omega)$ sowohl zur Restriktion beitragen als auch Information durch die Prolongation erhalten und zum anderen eine Matrix-Vektor-Multiplikation billig ist, verwenden wir zunächst den Ansatz

$$\tilde{A}_{ff}^{-1} = I + \kappa D_{ff}, \quad (5.10)$$

bzw. mit einer zusätzlichen Gewichtung,

$$\tilde{A}_{ff}^{-1} = \alpha_0 I + \alpha_1 \kappa D_{ff}. \quad (5.11)$$

Für eine kalte Konfiguration werden mit $\alpha_0^{-1} = 2\kappa$ und $\alpha_1^{-1} = 16\kappa^2$ konstante Gitterfunktionen auf konstante Gitterfunktionen abgebildet. Da in diesem Fall die konstante Gitterfunktion $\mathbf{1}$ auch Eigenvektor zum kleinsten Eigenwert $1 - 4\kappa^2$ der Matrix A ist, ist diese Wahl optimal, sofern die Schurkomplement-Approximation die Wirkung des exakten Schurkomplements auf $\mathbf{1}$ gut reflektiert, vgl. auch Seite 59. Mit \tilde{S}_A aus (5.16) werden wir im nächsten Abschnitt eine solche Schurkomplement-Approximation vorstellen. Für realistische bzw. heiße Konfigurationen ist die optimale Wahl der Gewichte jedoch nicht bekannt und kann lediglich konfigurationsabhängig experimentell ermittelt werden.

Wie bereits in Kapitel 4 diskutiert, ist eine unvollständige LU-Zerlegung bzw. eine SGS-Zerlegung

$$\tilde{A}_{ff}^{-1} = \begin{pmatrix} I & 0 & \kappa D_{L_1, L_2} \\ 0 & I & \kappa D_{L_3, L_2} \\ 0 & 0 & I \end{pmatrix} \begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ \kappa D_{L_2, L_1} & \kappa D_{L_2, L_3} & I \end{pmatrix} \quad (5.12)$$

dem Ansatz (5.10) vorzuziehen, da u. a. bei gleichem Aufwand die Restriktion und die Prolongation bei dieser Vorgehensweise identisch ist zur Restriktion und Prolongation mit der approximativen Inversen

$$\tilde{A}_{ff}^{-1} = I + \kappa D_{ff} + (\kappa D_{ff})^2. \quad (5.13)$$

Für eine approximative Inverse der Gestalt (5.10) erhalten wir

$$-A_{cf} \tilde{A}_{ff}^{-1} = \kappa \begin{pmatrix} D_{L_0, L_1} & 0 & 0 \\ 0 & D_{L_4, L_3} & 0 \end{pmatrix} + \kappa^2 \begin{pmatrix} 0 & 0 & \kappa^2 D_{L_0, L_1} D_{L_1, L_2} \\ 0 & 0 & \kappa^2 D_{L_4, L_3} D_{L_3, L_2} \end{pmatrix},$$

bei einer durch die ILU-Zerlegung (5.12) induzierten approximativen Inversen hingegen

$$\begin{aligned} -A_{cf} \tilde{A}_{ff}^{-1} &= \kappa \begin{pmatrix} D_{L_0, L_1} & 0 & 0 \\ 0 & D_{L_4, L_3} & 0 \end{pmatrix} + \kappa^2 \begin{pmatrix} 0 & 0 & \kappa^2 D_{L_0, L_1} D_{L_1, L_2} \\ 0 & 0 & \kappa^2 D_{L_4, L_3} D_{L_3, L_2} \end{pmatrix} \\ &+ \kappa^3 \begin{pmatrix} D_{L_0, L_1} D_{L_1, L_2} D_{L_2, L_1} & D_{L_0, L_1} D_{L_1, L_2} D_{L_2, L_3} & 0 \\ D_{L_4, L_3} D_{L_3, L_2} D_{L_2, L_1} & D_{L_4, L_3} D_{L_3, L_2} D_{L_2, L_3} & 0 \end{pmatrix}. \end{aligned}$$

Dieser Ansatz berücksichtigt bei der Restriktion auf den Gitterpunkt x zusätzlich Information von Gitterpunkten y mit

$$y \notin H(x) = \{y \neq x \mid |x_\mu - y_\mu| \leq 1, \quad \mu = 1, \dots, 4\}.$$

Des weiteren sind kompliziertere matrixabhängige Restriktionen und Prolongationen denkbar, die nicht notwendig auf einer trunkierten Neumann-Reihe basieren. Als ein Beispiel einer solchen matrixabhängigen Restriktion möge ein Restriktion dienen, bei der Information von allen Gitterpunkten $y \in H(x)$ auf den Gitterpunkt x transportiert wird.

Die Schurkomplement-Approximation

Wegen der Wahl der groben Gitterpunkte Ω_c koppeln grobe Gitterpunkte untereinander nur, wenn diese bzgl. Ω über Wege gerader Länge verbunden sind, vgl. Lemma 3.22. Somit folgt für das exakte Schurkomplement

$$S_A = I - \kappa^2 D_{cf} \sum_{\nu=0}^{\infty} (\kappa D_{ff})^{2\nu} D_{fc}.$$

Um eine rekursive Vorgehensweise zu ermöglichen, ist es zwingend, das Besetztheitsmuster der Schurkomplement-Approximation so vorzuschreiben, daß zum einen Auffüllung vermieden wird und zum anderen die Schurkomplement-Approximation lokal wirkt. Wird als Schurkomplement-Approximation die Matrix

$$\tilde{S}_A = I - \kappa^2 D_{cf} D_{fc}$$

gewählt, so repräsentiert diese Grobgittermatrix eine Nächste-Nachbar-Kopplung auf den groben Gitterpunkten Ω_c ,

$$\begin{aligned} \psi_c(x) - 2\kappa^2 \left(\sum_{\mu=1}^4 ((I_4 - \gamma_\mu) \otimes U_\mu(x) U_\mu(x + e_\mu)) \psi_c(x + 2e_\mu) \right. \\ \left. + \sum_{\mu=1}^4 ((I_4 + \gamma_\mu) \otimes U_\mu^H(x - e_\mu) U_\mu^H(x - 2e_\mu)) \psi_c(x - 2e_\mu) \right) = \varphi_c(x). \end{aligned} \quad (5.14)$$

Tatsächlich existieren zwei voneinander unabhängige Grobgittersysteme auf den Gitterpunkten $L_0(\Omega)$ und $L_4(\Omega)$, vgl. Abb. 5.5.

Alle Eigenschaften der Wilson Fermion Matrix bleiben erhalten, so daß die Schurkomplement-Approximation als neue Wilson Fermion Matrix mit

$$\hat{\kappa} = 2\kappa^2, \quad \hat{U}_\mu(x) = U_\mu(x) U_\mu(x + e_\mu) \quad (5.15)$$

angesehen werden kann, vgl. (4.6), (4.7).

Ein Vergleich mit Lemma 5.18 offenbart, daß die Wilson Fermion Matrix auf den groben Gitterpunkten sich gerade aus schwachen Kopplungen der odd-even reduzierten Wilson Fermion Matrix zusammensetzt. Insgesamt ist die Matrix sehr gut konditioniert. Da eine schlecht konditionierte Matrix sich nicht sinnvoll durch eine gut konditionierte Matrix approximieren läßt, ist es notwendig, den Nichtdiagonalanteil (mindestens) durch einen zusätzlichen Parameter zu steuern,

$$\tilde{S}_A = I - \omega \kappa^2 D_{cf} D_{fc}. \quad (5.16)$$

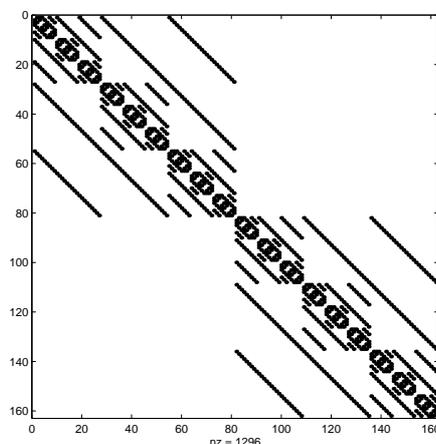


Abbildung 5.5: Das Besetzmuster der Schurkomplement-Approximation. Zu erkennen sind die zwei unabhängigen Grobgittersysteme auf Ω_c .

Für eine kalte Konfiguration ist $\omega = 1/(1 - 48\kappa^2)$ eine geeignete Wahl, da man in diesem Fall analog zum Beweis von Lemma 3.29 ohne Mühe beweisen kann, daß die konstante Gitterfunktion ein Eigenvektor zum betragskleinsten Eigenwert des exakten Schurkomplements und $\tilde{S}_A \mathbf{1} = S_A \mathbf{1}$ ist.

Eine weitere Heuristik ist es, als Grobgittergewicht $\omega^{-1} = 2\kappa$ zu wählen. In diesem Fall sind die effektiven Hopping-Parameter identisch,

$$2\omega\kappa^2 = \kappa.$$

Allerdings ist, obwohl eine kalte (heiße) Konfiguration beim Übergang zum Grobgittersystem eine kalte (heiße) Konfiguration bleibt, für realistische Konfigurationen die Konfiguration auf dem Grobgitter deutlich „heiß“, d.h. für die kritischen Werte κ_c und $\hat{\kappa}_c$ gilt $\kappa_c \ll \hat{\kappa}_c$, vgl. Abb. 5.6.

Die Verteilung der Konfiguration $U = \{U_\mu(x)\}$ wird in der Theoretischen Physik durch einen Parameter $\beta \geq 0$ gesteuert. Erste Messungen zeigen, daß realistische Konfigurationen bei $\beta = 6.0$ zu realistischen Grobgitterkonfigurationen bei $\beta = 3.0$ führen.

Eine systematische Untersuchung des physikalischen Zusammenhangs zwischen der originären und der resultierenden Konfiguration steht seitens der Physik aus, jedoch glauben wir, daß dies zum Verständnis des funktionalen Zusammenhangs zwischen optimalen Restriktions-, Prolongations- und Grobgittergewichten unerlässlich ist. Deshalb wurde der optimale Wert von ω mittels numerischer Experimente bestimmt.

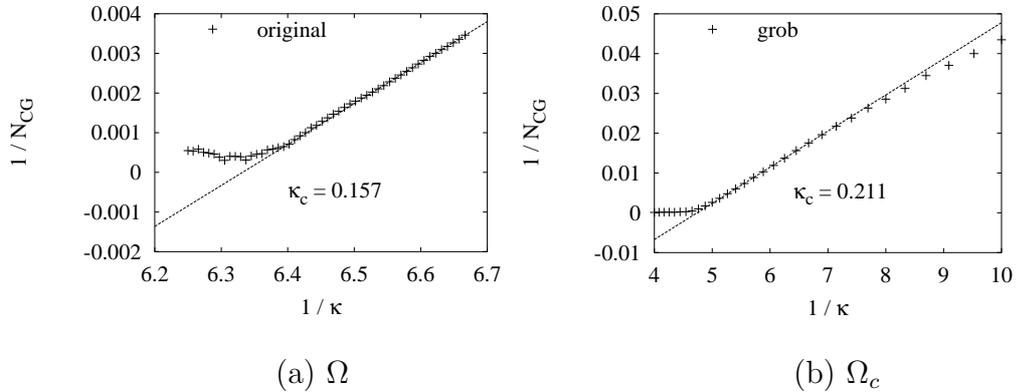


Abbildung 5.6: Die Extrapolation des kritischen Wertes κ_c für eine realistische Konfiguration auf Ω und für die korrespondierende Konfiguration auf Ω_c . Als Maß für den Wert von κ wird hier die Anzahl der CG-Iterationen zur Lösung der Normalengleichung verwendet.

Abbildung 5.7 zeigt die Anzahl benötigter Iterationen eines einfachen Zweigitterverfahrens für eine kalte Konfiguration bei $\kappa = 0.124$. Deutlich ist zu erkennen, daß zum einen eine Relaxation der Grobgitterkopplungen für eine schnelle Konvergenz des Verfahrens notwendig ist und zum anderen das Grobgittergewicht ω nicht unabhängig von den Parametern $\alpha_i = \alpha$ der Restriktion und Prolongation gewählt werden kann. Empfehlenswert ist $\alpha = 2$ und $\omega \approx 3.8$.

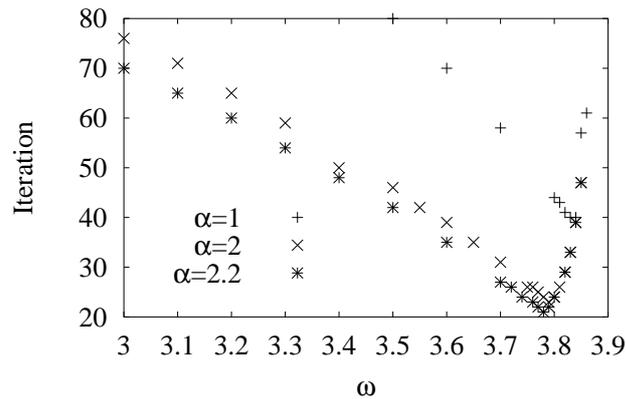


Abbildung 5.7: Die Anzahl benötigter Iterationsschritte einer einfachen Zweigitteriteration für eine kalte Konfiguration und $\kappa = 0.124$ bei drei verschiedenen Restriktions- bzw. Prolongationsparametern $\alpha_i = \alpha$ in Abhängigkeit vom Grobgittergewicht ω .

In Abbildung 5.8 ist der Iterationsverlauf einer einfachen Zweigitteriteration für eine realistische Konfiguration und einem moderaten Hopping-Parameter $\kappa = 0.15$ zu sehen. Zum Vergleich sind die entsprechenden Residuen bei einer GS-Iteration (und odd-even Anordnung der Gitterpunkte) aufgetragen.

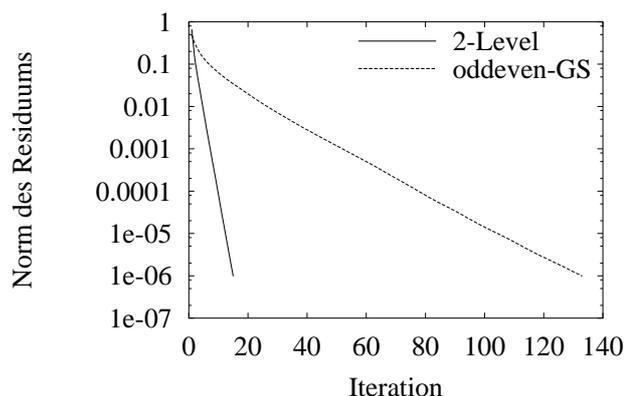


Abbildung 5.8: Das Konvergenzverhalten einer einfachen Zweigitteriteration und des GS-Verfahrens (bei einer odd-even Anordnung) für eine realistische Konfiguration mit $\kappa = 0.15$, $\alpha_i = 1.6$ und $\omega = 3.8$.

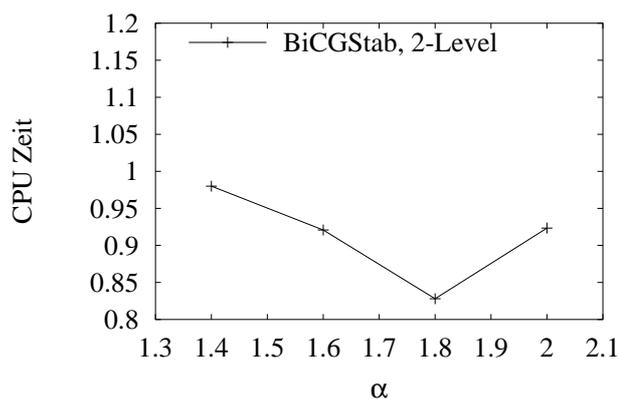


Abbildung 5.9: Die Ausführungszeiten des BiCGStab und des Schurkomplement-präkonditionierten BiCGStab-Verfahrens für eine realistische Konfiguration auf einem 16^4 -Gitter mit $\kappa = 0.155$ und $\omega = 3.7$. Dem Präkonditionierer sind zusätzlich eine Jacobi-Iteration vor- und zwei GS-Iterationen (odd-even Anordnung) zur Glättung nachgeschaltet.

Die Ausführungszeiten des BiCGStab und des Schurkomplement-präkonditionierten BiCGStab-Verfahrens für eine realistische Konfiguration auf einem 16^4 -Gitter mit $\kappa = 0.155$ und $\omega = 3.7$ in Abhängigkeit der Gewichte $\alpha_i = \alpha$ sind in Abbildung 5.9 wiedergegeben. Dem Präkonditionierer sind zusätzlich eine Jacobi-Iteration vor- und zwei GS-Iterationen (odd-even Anordnung) zur Glättung nachgeschaltet. Zu erkennen ist, daß das Schurkomplement-präkonditionierten BiCGStab-Verfahren für $\alpha = 1.8$ lediglich ca. 80% der CPU-Zeit des unpräkonditionierten BiCGStab benötigt.

5.4 Schwarz-Iteration

Ein weiterer Ansatz, der im Rahmen dieser Arbeit untersucht wurde, basiert auf einer Gebietszerlegung, bei der die Unterräume auf einer Zerlegung der Wilson Fermion Matrix nach Symmetrien beruhen. Die Diskussion dieses Ansatzes mit Vertretern der Theoretischen Physik führte zur Entwicklung der *Single-Flavour-Algorithmen* [46].

Definition 5.27. Die Unterräume V_1 und V_2 seien definiert als $V_1 = \text{kern}(I - \Gamma_5)$ und $V_2 = \text{kern}(I + \Gamma_5)$.

Lemma 5.28. Es sei $x \in V_1$ und $y \in V_2$, so gilt $x = \Gamma_5 x$ und $y = -\Gamma_5 y$.

Beweis. Trivial. □

Das Skalarprodukt $\langle \cdot, \cdot \rangle_A$ ist symmetrisch auf V_i .

Lemma 5.29. Für $x, y \in V_i$ ist $\langle x, Ay \rangle = \langle Ax, y \rangle$.

Beweis. Es seien $x, y \in V_1$. Wir erhalten

$$\langle x, Ay \rangle = \langle \Gamma_5 x, Ay \rangle = \langle x, \Gamma_5 Ay \rangle = \langle x, A^H \Gamma_5 y \rangle = \langle x, A^H y \rangle = \langle Ax, y \rangle.$$

Der zweite Teil wird analog bewiesen. □

Für $x \in V_i, y \in V_j, i \neq j$, hingegen ist $\langle x, y \rangle_A$ antisymmetrisch.

Lemma 5.30. Es sei $x \in V_i$ und $y \in V_j, i \neq j$. Dann gilt $\langle x, Ay \rangle = \langle -Ax, y \rangle$.

Beweis. Es seien $x \in V_1$ und $y \in V_2$. Dann gilt

$$\langle x, Ay \rangle = \langle \Gamma_5 x, Ay \rangle = \langle x, \Gamma_5 Ay \rangle = \langle x, A^H \Gamma_5 y \rangle = -\langle x, A^H y \rangle = \langle Ax, y \rangle$$

Der zweite Teil wird analog bewiesen. □

Das folgende Lemma zeigt, daß das Skalarprodukt $\langle \cdot, \cdot \rangle_A$ unter gewissen Voraussetzungen sogar symmetrisch positiv definit auf V_i ist.

Lemma 5.31. Es sei $x \in V_i$. Ist $0 \leq \kappa < \kappa_c$ hinreichend klein, so ist $\langle x, Ax \rangle > 0$.

Beweis. $\langle Ax, x \rangle = \langle x, Ax \rangle = \langle x, x \rangle - \kappa \langle x, Dx \rangle > 0$. □

Definition 5.32. Die Matrizen P_i seien definiert als $P_1 = \frac{1}{2}(I + \Gamma_5)$ und $P_2 = \frac{1}{2}(I - \Gamma_5)$.

Die Matrizen P_i sind orthogonale Projektionen auf V_i .

Lemma 5.33. Es gilt $P_i^2 = P_i$, $P_i^H = P_i^T = P_i$, $P_2 = I - P_1$, $P_1v = \Gamma_5 P_1v$ und $P_2v = -\Gamma_5 P_2v$ für alle $v \in \mathbb{C}^n$.

Beweis. Per Definition von Γ_5 klar. □

Satz 5.34. Die Matrizen P_i sind die eindeutig bestimmten orthogonalen Projektionen auf V_i .

Beweis. Es sei $v = v_1 + v_2$ mit $v_i \in V_i$. Wir erhalten $v = v_1 + v_2 = \Gamma_5 v_1 - \Gamma_5 v_2 = \Gamma_5(v_1 - v_2)$. Aus $v = v_1 + v_2$ und $v_1 - v_2 = \Gamma_5 v$ folgt sofort $v_i = P_i v$. □

Sowohl bei der additiven als auch bei der multiplikativen Schwarz-Iteration [38] ist es erforderlich, Teilsysteme mit den Matrizen $P_i A P_i^T$ zu lösen. Aus diesem Grund betrachten wir diese Matrizen genauer.

Lemma 5.35. Es gilt $P_i A P_i = \frac{1}{2} P_i (A + A^H) = \frac{1}{2} (A + A^H) P_i$.

Beweis. Wir zeigen $P_1 A P_1 = \frac{1}{2} P_1 (A + A^H) = \frac{1}{2} (A + A^H) P_1$.

$$4P_1 A P_1 = A + \Gamma_5 A + A \Gamma_5 + \Gamma_5 A \Gamma_5 = A + A^H + (\Gamma_5 A + A \Gamma_5)$$

Mit $\Gamma_5 A + A \Gamma_5 = \Gamma_5 (A + A^H) = (A + A^H) \Gamma_5$ und

$$\begin{aligned} 2P_1 (A + A^H) &= A + A^H + \Gamma_5 (A + A^H) \\ &= A + A^H + (A + A^H) \Gamma_5 = 2(A + A^H) P_1 \end{aligned}$$

ist die Behauptung für $i = 1$ bewiesen. Analog für $i = 2$. □

Lemma 5.36. Für $i \neq j$ gilt $P_i A P_j = \frac{1}{2} P_i (A - A^H) = \frac{1}{2} (A - A^H) P_j$.

Beweis. Wir zeigen $2P_1AP_2 = \frac{1}{2}P_1(A - A^H) = \frac{1}{2}(A - A^H)P_2$.

$$4P_1AP_2 = A + \Gamma_5A - A\Gamma_5 - \Gamma_5A\Gamma_5 = A - A^H + (\Gamma_5A - A\Gamma_5)$$

Mit $\Gamma_5A - A\Gamma_5 = \Gamma_5(A - A^H) = -(A - A^H)\Gamma_5$ und

$$\begin{aligned} 2P_1(A - A^H) &= A - A^H + \Gamma_5(A - A^H) \\ &= A - A^H - (A - A^H)\Gamma_5 = 2(A - A^H)P_2 \end{aligned}$$

ist die Behauptung bewiesen. \square

Zerlegen wir die Wilson Fermion Matrix in ihren hermiteschen und schiefhermiteschen Anteil,

$$A = A_+ - A_-, \quad A_+ = \frac{A + A^H}{2}, \quad A_- = \frac{A - A^H}{2}, \quad (5.17)$$

so gilt nach Lemma 5.35 und Lemma 5.36 offensichtlich

$$P_iAP_i = P_iA_+ = A_+P_i, \quad P_iAP_j = P_iA_- = A_-P_j, \quad i \neq j. \quad (5.18)$$

Lemma 5.37. Ist $A_+v = \lambda v$, so folgt $A_+\Gamma_5v = \lambda\Gamma_5v$.

Beweis. $(A + A^H)\Gamma_5 = A\Gamma_5 + \Gamma_5A = \Gamma_5(A + A^H)$. \square

Die Eigenwerte der Matrix A_+ treten also paarweise auf. Wir wechseln nun die Eigenvektor-Basis und zeigen, daß die Spektren der Matrizen P_iAP_i und A_+ identisch sind.

Lemma 5.38. Ist $A_+v = \lambda v$ und $A_+\Gamma_5v = \lambda\Gamma_5v$, so gilt auch $A_+P_iv = \lambda P_iv$.

Beweis. $A_+P_iv = A_+(v + \Gamma_5v)/2 = (A_+v + A_+\Gamma_5v)/2 = \lambda(v + \Gamma_5v)/2 = \lambda P_iv$. \square

Lemma 5.39. Es gilt $\sigma(P_1AP_1) = \sigma(P_2AP_2) = \{0\} \cup \sigma(A_+)$.

Beweis. Es sei $A_+P_iv = \lambda P_iv$. Dann gilt $(P_iAP_i)P_iv = P_iAP_i^2v = P_iAP_iv = A_+P_iv = \lambda P_iv$. Umgekehrt folgt aus $P_iAP_iv = \lambda v$ mit $v \in V_i$ sofort $\lambda v = P_iAP_iv = A_+P_iv = A_+v$. Für $v \notin V_i$ ist $P_iAP_iv = 0$ trivial. \square

Lemma 5.40. Mit $D_+ = (D + D^H)/2$ und $D_- = (D - D^H)/2$ gilt $A_+ = I - \kappa D_+$ sowie $A_- = -\kappa D_-$.

Beweis. Trivial. \square

Konfiguration	$\kappa_c(A)$	$\kappa_c(A_+)$
cold	0.125	0.125
hot	0.250	0.188

Tabelle 5.1: Der kritischer Wert κ_c für A sowie für den hermiteschen Anteil A_+ von A

Für die Matrix $A_+ = I - \kappa D_+$ definieren wir den kritischen Wert $\kappa_c(A_+)$ analog zum kritischen Wert $\kappa_c = \kappa_c(A)$ der Wilson Fermion Matrix A als

$$\kappa_c(A_+) = 1 / \max_{\lambda \in \sigma(D_+)} \operatorname{Re} \lambda.$$

Aus der Tabelle 5.1 ist ersichtlich, daß sowohl für A als auch A_+ im Falle einer kalten Konfiguration $\kappa_c(A) = 0.125$ und $\kappa_c(A_+) = 0.125$ ist.

Da jedoch für A_+ im Falle einer heißen Konfiguration $\kappa_c(A_+) \ll 0.25$ ist, ist es wesentlich schwieriger, für ein hinreichend großes κ ein System mit dem hermiteschen Anteil A_+ der Matrix A zu lösen als eines mit der Matrix A . Die Spektren der Matrizen $P_i A_i P_i$ und A_+ sind aber identisch, also ist eine Schwarz-Iteration zur Präkonditionierung der Wilson Fermion Matrix A nicht geeignet.

Die obige Diskussion läßt sich völlig analog auch für das Schwinger Modell (s. Kapitel 4) führen. In Abbildung 5.10 sind die Spektren von D , D_+ und D_- im Falle einer heißen Konfiguration aufgetragen. Deutlich zu erkennen ist $\kappa_c(A_+) \ll \kappa_c(A)$.

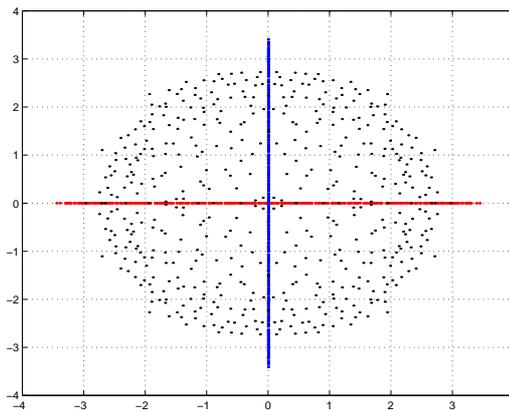


Abbildung 5.10: Die Spektren von D , des hermiteschen Anteils D_+ (reelle Achse) und des schief-hermiteschen Anteils D_- (imaginäre Achse) von D im Falle der Schwinger Matrix und einer heißen Konfiguration, $N = 16$.

5.5 Zusammenfassung

Es wurden insgesamt vier verschiedene maximal unabhängige Mengen des reduzierten Graphen $G_{A^{(e)}}(V, E_S^*)$ der odd-even reduzierten Wilson Fermion Matrix bestimmt. Jede dieser Mengen beinhaltet aber lediglich $n/8$ Gitterpunkte. Wird eine dieser maximal unabhängigen Mengen als Menge der feinen Gitterpunkte ausgewählt, so sind selbst naive Schurkomplement-Präkonditionierer der odd-even reduzierten Wilson Fermion Matrix nicht geeignet, da die resultierende Schurkomplement-Approximation einen zu großen Stern beschreibt.

Anschließend wurden Schurkomplement-Präkonditionierer der originären Wilson Fermion Matrix untersucht, wobei als Grobgitterpunkte zwei ineinander verschränkte Standard-Grobgitter gewählt wurden.

Als approximative Inverse der Matrix A_{ff} bieten sich sowohl der durch die Jacobi- als auch der durch die ILU- bzw. SGS-Zerlegung induzierte Präkonditionierer an. In beiden Fällen sind die Präkonditionierer γ_5 -symmetrisch, und die matrixabhängigen Restriktionen und Prolongationen sind zueinander γ_5 -adjungiert.

Die Schurkomplement-Approximation besteht aus zwei voneinander unabhängigen Nächste-Nachbar-Kopplungen, welche wiederum als Wilson Fermion Matrizen interpretiert werden können. Somit ist prinzipiell eine rekursive Vorgehensweise einfach umzusetzen, wobei von einer Gitterebene zur nächsten sich die Anzahl der Grobgittersysteme jeweils verdoppelt.

Numerische Experimente zeigen, daß eine zusätzliche Gewichtung der Grobgittermatrizen, aber auch der Restriktionen und Prolongationen zu einer erheblichen Reduktion der Iterationszahlen führen. Im Falle einer kalten Konfiguration kann die Wahl geeigneter Gewichte motiviert werden, für eine realistische bzw. heiße Konfiguration ist dies zunächst nicht möglich. Da die Grobgittermatrizen aber Wilson Fermion Matrizen sind, erscheint eine physikalisch motivierte Wahl der Parameter denkbar. Erste Ansätze wie ein Festhalten der effektiven Hopping-Werte κ oder der effektiven Massen m_q wurden untersucht und verworfen, da sich außer dem Gitterabstand a und dem kritischen Wert κ_c auch weitere physikalische Größen, beispielsweise der die Verteilung der Matrizen $U_\mu(x)$ steuernde Parameter β beim Übergang von Ω auf das grobe Gitter Ω_c ändern.

Es sei an dieser Stelle darauf hingewiesen, daß völlig analog zum Beweis der Glättungseigenschaft des (modifizierten) Jacobi-Glätters und der Herleitung

eines optimalen semiiterativen Glätters im Falle der hermiteschen Schwinger Matrix entsprechende Aussagen für die hermitesche Wilson Fermion Matrix $\Gamma_5 A$ formuliert und bewiesen werden können.

Schlußendlich wurden ausgehend von der γ_5 -Symmetrie der Wilson Fermion Matrix Schwarz-Iterationen untersucht. Dabei wurden die Unterräume so gewählt, daß die resultierenden Teilsysteme hermitesch positiv definit sind. Es wurde bewiesen, daß der Ansatz im wesentlichen identisch ist zu einer Zerlegung der Wilson Fermion Matrix in ihren hermiteschen und schieferhermiteschen Anteil. Obwohl gezeigt werden konnte, daß die Teilsysteme schwieriger zu lösen sind als das ursprüngliche System, ist dieser Präkonditionierungsansatz als Erfolg zu werten, da diese Untersuchungen den Anstoß zur Entwicklung der *Single-Flavour*-Algorithmen gaben [46].

Anhang A

Statistik

Lemma A.1. Seien x und y gleichverteilte und unabhängige Zufallsvariablen auf dem Einheitskreis S . Dann ist auch das Produkt xy gleichverteilt auf S .

Beweis. Aufgrund der Periodizität auf S ist dies klar. \square

Satz A.2. Es seien x und y gleichverteilte und unabhängige Zufallsvariablen auf dem Einheitskreis S . Dann ist $E = 4/\pi$ der Erwartungswert von $|x + y|$.

Beweis. Zunächst sei $x = \exp(2\pi it)$ und $y = \exp(2\pi is)$ mit $s, t \in [0, 1)$. Der gesuchte Erwartungswert ist definiert durch

$$\begin{aligned} & \int_0^1 \int_0^1 |\exp(2\pi it) + \exp(2\pi is)| \, ds \, dt \\ &= \int_0^1 \int_0^1 |\exp(2\pi it)| |1 + \exp(2\pi i(s - t))| \, ds \, dt = \int_0^1 |1 + \exp(2\pi it)| \, dt. \end{aligned}$$

Mit $z = 1 + \exp(2\pi it)$ folgt $|z|^2 = 2(1 + \cos(2\pi t))$, und wir erhalten

$$\int_0^1 |1 + \exp(2\pi it)| \, dt = \sqrt{2} \int_0^1 \sqrt{1 + \cos(2\pi t)} \, dt.$$

Aus $\cos(2\pi t) = \cos^2(\pi t) - \sin^2(\pi t) = 2\cos^2(\pi t) - 1$ folgt nun

$$\begin{aligned}\sqrt{2} \int_0^1 \sqrt{1 + \cos(2\pi t)} dt &= \sqrt{2} \int_0^1 \sqrt{2\cos^2(\pi t)} dt = 2 \int_0^1 |\cos(\pi t)| dt \\ &= 4 \int_0^{\frac{1}{2}} \cos(\pi t) dt = \frac{4}{\pi} \sin(\pi t) \Big|_0^{\frac{1}{2}} = \frac{4}{\pi}.\end{aligned}$$

Damit ist der Satz bewiesen. \square

Satz A.3. Es seien x und y gleichverteilte und unabhängige Zufallsvariablen auf dem Einheitskreis S . Dann gilt für den Erwartungswert:

$$\begin{aligned}E(|(1+i)x + (1-i)y|) &= E(|(1-i)x + (1+i)y|) = 4\sqrt{2}/\pi, \\ E(|(1+i)(x+y)|) &= E(|(1-i)(x+y)|) = 4\sqrt{2}/\pi.\end{aligned}$$

Beweis. Es sei $x = \exp(2\pi it)$ und $y = \exp(2\pi is)$ mit $s, t \in [0, 1)$. Dann gilt

$$\begin{aligned}&\int_0^1 \int_0^1 |(1+i)\exp(2\pi it) + (1-i)\exp(2\pi is)| ds dt \\ &= \int_0^1 \int_0^1 |(1+i)\exp(2\pi it)| \left| 1 + \frac{1-i}{1+i} \exp(2\pi i(s-t)) \right| ds dt \\ &= \sqrt{2} \int_0^1 |1 - i \exp(2\pi it)| dt.\end{aligned}$$

Mit $z = 1 - i \exp(2\pi it)$ folgt $|z|^2 = 2(1 + \sin(2\pi t))$, und wir erhalten

$$\begin{aligned}\sqrt{2} \int_0^1 |1 - i \exp(2\pi it)| dt &= 2 \int_0^1 \sqrt{1 + \sin(2\pi t)} dt \\ &= 2 \int_0^1 \sqrt{1 + \cos(2\pi(t-1/4))} dt = 2 \int_{-1/4}^{3/4} \sqrt{1 + \cos(2\pi t)} dt \\ &= 2\sqrt{2} \int_{-1/4}^{3/4} |\cos(\pi t)| dt = 2\sqrt{2} \left(\int_{-1/4}^{1/2} \cos(\pi t) dt - \int_{1/2}^{3/4} \cos(\pi t) dt \right) \\ &= 2\sqrt{2}/\pi \left(\sin(\pi t) \Big|_{-1/4}^{1/2} - \sin(\pi t) \Big|_{1/2}^{3/4} \right) = 4\sqrt{2}/\pi.\end{aligned}$$

Der zweite Teil der Behauptung folgt wegen

$$E(|(1+i)(x+y)|) = \sqrt{2}E(|(x+y)|)$$

bzw.

$$E(|(1-i)(x+y)|) = \sqrt{2}E(|(x+y)|)$$

aus Satz A.2. □

Lemma A.4. Es seien $x = \exp(2\pi is)$, $y = \exp(2\pi it)$ komplexe Zahlen mit $s, t \in [0, 1)$. Des weiteren seien $z_1 = x + y + i(x - y)$ und $z_2 = x + y - i(x - y)$ definiert. Dann gilt

$$|z_1|^2 = 4(1 - \sin(2\pi(s-t))), \quad |z_2|^2 = 4(1 + \sin(2\pi(s-t))),$$

also insbesondere $\max\{|z_1|, |z_2|\} \geq 2$.

Beweis. Wir rechnen nach:

$$\begin{aligned} |z_1|^2 &= ((1+i)\exp(2\pi is) + (1-i)\exp(2\pi it)) \\ &\quad ((1-i)\exp(-2\pi is) + (1+i)\exp(-2\pi it)) \\ &= 2(1+i)(1-i) + (1+i)^2 \exp(2\pi i(s-t)) + (1-i)^2 \exp(2\pi i(s-t)) \\ &= 4 + 2i \exp(2\pi i(s-t)) - 2i \exp(2\pi i(s-t)) = 4 - 4 \sin(2\pi(s-t)). \end{aligned}$$

Analog gilt

$$\begin{aligned} |z_2|^2 &= ((1-i)\exp(2\pi is) + (1+i)\exp(2\pi it)) \\ &\quad ((1+i)\exp(-2\pi is) + (1-i)\exp(-2\pi it)) \\ &= 2(1+i)(1-i) + (1-i)^2 \exp(2\pi i(s-t)) + (1+i)^2 \exp(2\pi i(s-t)) \\ &= 4 - 2i \exp(2\pi i(s-t)) + 2i \exp(2\pi i(s-t)) = 4 + 4 \sin(2\pi(s-t)). \end{aligned}$$

Nun folgt die Behauptung unmittelbar. □

Literaturverzeichnis

- [1] MATLAB 5.3. The MathWorks Inc., www.mathworks.com.
- [2] R. E. Alcouffe, A. Brandt, J. E. Deny, J. W. Painter. The multigrid method for the diffusion equation with strongly discontinuous coefficients. *SIAM J. Sci. Stat. Comp.*, 2:430–454, 1981.
- [3] W. E. Arnoldi. The principle of minimized iteration in the solution of the matrix eigenvalue problem. *Quart. Appl. Math.*, 9:17–29, 1951.
- [4] Owe Axelsson. *Iterative Solution Methods*. Cambridge University Press, Cambridge, UK, 1994.
- [5] M. Bäker, T. Kalkreuter, G. Mack, M. Speh. Neural multigrid for gauge theories and other disordered systems. *Int. J. Mod. Phys.*, C 4:239, 1993.
- [6] R. Barrett et al. *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods*. SIAM, Philadelphia, 1994.
- [7] R. Ben-Av, A. Brandt, M. Harmatz, E. Katznelson, P G. Lauwers, S. Solomon, K. Wolowesky. Fermionic simulations using parallel transported multigrid. *Phys. Lett.*, B 253:185, 1991.
- [8] R. Ben-Av, A. Brandt, S. Solomon. The fermionic matrix, instantons, zero modes and multigrid. *Nucl. Phys.*, B 329:139, 1990.
- [9] A. Brandt. Multi-level adaptive technique for fast numerical solution to boundary value problems. In *Third International Conference on Numerical Methods in Fluid Dynamics, Paris 1972, Lecture Notes in Physics 18*. Springer, Berlin, 1973.

-
- [10] W. Briggs. *A multigrid tutorial*. SIAM, Philadelphia, PA, 1987.
- [11] R. C. Brower, R. G. Edwards, C. Rebbi, E. Vicari. Projective multigrid for Wilson fermions. *Nucl. Phys.*, B 366:689, 1991.
- [12] R. C. Brower, K. Moriarty, E. Myers, C. Rebbi. The multigrid method for fermion calculations in quantum chromodynamics. In S. F. McCormick, Hrsg., *Multigrid Methods: Theory, Applications and Supercomputing, Lecture Notes in Pure and Applied Mathematics 110*. Springer Verlag, Berlin, 1988.
- [13] R. C. Brower, K. Moriarty, C. Rebbi, E. Vicari. Variational multigrid for non-abelian gauge theory. *Nucl. Phys.*, B 20:89, 1991.
- [14] R. C. Brower, C. Rebbi, E. Vicari. Multigrid propagators in the presence of disordered U(1) gauge fields. *Phys. Rev.*, D 43:1965, 1991.
- [15] A. M. Bruaset. *A survey of preconditioned Iterative Methods*. Addison Wesley Longman, Harlow, Essex, UK, 1992.
- [16] B. Buzbee, G. Golub, C. Nielsen. On direct methods for solving Poisson's equation. *SIAM J. Numer. Anal.*, 7:627–656, 1970.
- [17] J. D. F. Cosgrove, J. C. Diaz, A. Griewank. Approximate inverse preconditionings for sparse matrices. *Int. J. Comp. Math.*, 44:91–110, 1992.
- [18] P. M. de Zeeuw. Matrix-dependent prolongations and restrictions in a black box multigrid solver. *J. Comp. Appl. Math.*, 33:1–27, 1990.
- [19] J. W. Demmel. *Applied Numerical Linear Algebra*. SIAM, Philadelphia, 1997.
- [20] J. E. Deny. Black box multi-grid. *J. Comp. Phys.*, 48:366–386, 1982.
- [21] R. G. Edwards. *Numerical Simulations in Lattice Gauge Theories and Statistical Mechanics*. Dissertation, Department of Physics, New York University, 1989.
- [22] R. P. Fedorenko. A relaxation scheme for the solution of elliptic differential equations. *UdSSR Comput. Math. Phys.*, 1:1092–1096, 1961.
- [23] P. Fiebach. *Krylov-Verfahren zur Lösung von $P(A)x = b$* . Dissertation, Fachbereich Mathematik, BUGH Wuppertal, 1995.

-
- [24] R. Fletcher. Conjugate gradient methods for indefinite systems. In G. A. Watson, Hrsg., *Numerical Analysis Dundee 1975, Lecture Notes in Mathematics 506*, pages 73–89. Springer Verlag, Berlin, 1976.
- [25] R. W. Freund, N. M. Nachtigal. QMR: A quasi-minimal residual method for non-hermitian linear systems. *Numer. Math.*, 60:315–339, 1991.
- [26] A. Frommer. *Lösung linearer Gleichungssysteme auf Parallelrechnern*. Vieweg, Braunschweig, 1990.
- [27] A. Frommer, U. Glässner. Restarted GMRES for shifted linear systems. *SIAM J. Sci. Comp.*, 19:15–26, 1998.
- [28] A. Frommer, S. Güsken, T. Lippert, B. Nöckel, K. Schilling. Many masses on one stroke: Economic computation of quark propagator. *Int. J. Mod. Phys., C* 6:627–638, 1995.
- [29] A. Frommer, P. Maaß. Fast CG-based methods for Tikhonov-Phillips-regularization. *SIAM J. Sci. Comp.*, 20:1831–1850, 1999.
- [30] A. Frommer, B. Medeke. Structure in Krylov subspace methods for the Wilson fermion matrix. In A. Sydow, Hrsg., *15th IMACS World Congress on Scientific Computation, Modelling and Applied Mathematic*. Wissenschaft & Technik Verlag, Berlin, 1997.
- [31] G. Golub, D. Vanderstraeten. On the preconditioning of matrices with a dominant skew-symmetric component. TR SCCM-98-16, Stanford University, 1998.
- [32] G. H. Golub, C. van Loan. *Matrix Computations*. John Hopkins University Press, Baltimore, 3. Auflage, 1996.
- [33] N. I. M. Gould, J. A. Scott. Sparse approximate inverse preconditioners using norm-minimization techniques. *SIAM J. Sci. Comp.*, 19:605–625, 1998.
- [34] R. Gupta. Introduction to lattice QCD. TR LA-UR-98-3174, Los Alamos Preprint, 1998.
- [35] W. Hackbusch. Ein iteratives Verfahren zur schnellen Auflösung elliptischer Randwertprobleme. TR 76-12, Math. Fakultät, Universität Köln, 1976.
- [36] W. Hackbusch, Hrsg., *Multigrid Methods, Lecture Notes in Mathematics 960*. Springer Verlag, 1982.

-
- [37] W. Hackbusch. *Multi-Grid Methods and Applications*. Springer, Berlin, 1985.
- [38] W. Hackbusch. *Iterative Lösung großer schwachbesetzter Gleichungssysteme*. Teubner, Stuttgart, 1991.
- [39] M. R. Hestenes, E. Stiefel. Methods of conjugate gradients for solving linear systems. *J. Res. Nat. Bur. Stand.*, 49:409–436, 1952.
- [40] A. Hulsebos, J. Smit, J.C. Vink. Multigrid inversion of the staggered fermion matrix. *Nucl. Phys.*, B 20:94, 1991.
- [41] A. Hulsebos, J. Smit, J.C. Vink. Multigrid inversion of lattice fermion operators. *Nucl. Phys.*, B 368:379, 1992.
- [42] A. Hulsebos, J. Smit, J.C. Vink. Multigrid inversion of the staggered fermion matrix with U(1) and SU(2) gauge fields. *Int. J. Mod. Phys.*, C 3:161, 1992.
- [43] T. Kalkreuter. Multigrid methods for the computation of propagators in gauge fields. *Phys. Comput.*, 5:629, 1994.
- [44] C. Lanczos. An iteration method for the solution of the eigenvalue problem of linear differential and integral operators. *J. Res. Nat. Bur. Stand.*, 45:255–282, 1950.
- [45] C. Lanczos. Solution of systems of linear equations by minimized iterations. *J. Res. Nat. Bur. Stand.*, 49:33–53, 1952.
- [46] Th. Lippert. One-flavour hybrid Monte Carlo with Wilson fermions. In A. Frommer, Th. Lippert, B. Medeke, K.Schilling, Hrsg., *Numerical Challenges in Lattice Quantum Chromodynamics, Lecture Notes in Computational Science and Engineering*. Springer Verlag, Heidelberg, 2000.
- [47] S. F. McCormick, J. W. Ruge. Multigrid methods for variational problems. *SIAM J. Numer. Anal.*, 19:924–929, 1982.
- [48] B. Medeke. On algebraic multilevel preconditioners in lattice gauge theory. In A. Frommer, Th. Lippert, B. Medeke, K.Schilling, Hrsg., *Numerical Challenges in Lattice Quantum Chromodynamics, Lecture Notes in Computational Science and Engineering*. Springer Verlag, Heidelberg, 2000.

-
- [49] Y. Notay. Optimal V-cycle algebraic multilevel preconditioning. *Num. Lin. Alg. Appl.*, 5:441–459, 1998.
- [50] J. M. Ortega. *Matrix Theory*. Plenum Press, NY, 1987.
- [51] C. Paige, M. Saunders. Solution of sparse indefinite systems of linear equations. *SIAM J. Numer. Anal.*, 12:617–629, 1975.
- [52] CP-PACS Project. www.rccp.tsukuba.ac.jp/cppacs/project-e.html.
- [53] S. Quell. Das look-ahead Lanczosverfahren für J -symmetrische, J -hermitesche und geshiftete Matrizen. Diplomarbeit, Fachbereich Mathematik, BUGH Wuppertal, 1996.
- [54] A. Reusken. An approximate cyclic reduction scheme. In W. Hackbusch, G. Wittum, Hrsg., *Multigrid Methods V, Lecture Notes in Computational Science and Engineering*. Springer, Heidelberg, 1994.
- [55] A. Reusken. Multigrid with matrix-dependent transfer operators for convection-diffusion problems. In P. W. Hemker, P. Wesseling, Hrsg., *Seventh International Symposium on Domain Decomposition Methods for Partial Differential Equations*. Birkhäuser, Basel, 1994.
- [56] T. J. Rivlin. *The Chebyshev Polynomials: from Approximation Theory to Algebra and Number Theory*. J. Wiley and Sons, NY, 1990.
- [57] H. J. Rothe. *Lattice Gauge Theories - An Introduction, Lecture Notes in Physics Vol. 43*. World Scientific, 1985.
- [58] J. W. Ruge, K. Stüben. Algebraic multigrid. In S. F. McCormick, Hrsg., *Multigrid Methods*. SIAM, Philadelphia, 1987.
- [59] Y. Saad. *Iterative Methods for Sparse Linear Systems*. PWS Publishing, Boston, 1996.
- [60] Y. Saad, M. H. Schultz. GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM J. Sci. Stat. Comput.*, 7:856–869, 1986.
- [61] Y. Saad, H. van der Vorst. Developments and trends in the parallel solution of linear systems. *Parallel Computing*, 25(13-14):1931–1970, 1999.
- [62] Y. Saad, H. van der Vorst. Iterative solution of linear systems in the 20-th century. *J. Comp. Appl. Math.*, to appear.

- [63] Y. Shapira. Two-level analysis and multigrid methods for spd, non-normal and indefinite problems. www.mgnet.org, 1995.
- [64] G. L. G. Sleijpen, H. A. van der Vorst, D. R. Fokkema. BiCGStab(l) and other hybrid Bi-CG methods. *Numerical Algor.*, 7:75–109, 1994.
- [65] K. Stüben. Algebraic multigrid (AMG): An introduction with applications. TR 53, GMD Report, 1999.
- [66] L. N. Trefethen, D. Bau. *Numerical Linear Algebra*. SIAM, Philadelphia, 1997.
- [67] H. A. van der Vorst. BiCGStab: A fast and smoothly convergent variant of Bi-CG for the solution of nonsymmetric linear systems. *SIAM J. Sci. Stat. Comput.*, 13:631–644, 1992.
- [68] P. Vanek, J. Mandel, M. Brezina. Algebraic multigrid by smoothed aggregation for second and fourth order elliptic problems. *Computing*, 56:179–196, 1996.
- [69] R. S. Varga. *Matrix Iterative Analysis*. Prentice Hall, Englewood Cliffs, NJ, 1962.
- [70] J. C. Vink and U. J. Wiese. Gauge fixing on the lattice without ambiguity. *Phys. Lett.*, B 289:122–126, 1992.
- [71] J.C. Vink. Multigrid inversion of fermion operators with SU(2) gauge fields in two and four dimensions. *Nucl. Phys.*, B 26:607, 1992.
- [72] C. Wagner. Introduction to algebraic multigrid, 1999. Course Notes of an Algebraic Multigrid Course at the University of Heidelberg.
- [73] C. Wagner, W. Kinzelbach, G. Wittum. Schur-complement multigrid – a robust method for groundwater flow and transport problems. *Num. Math.*, 75:523–545, 1997.
- [74] D. Young. *Iterative Solution of Large Linear systems*. Academic Press, NY, 1971.