Algebraische Mehrgitterverfahren, Eigenlöser und Gitter-QCD

Dissertation zum Erlangen des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

am Fachbereich Physik, Mathematik und Informatik
der Johannes Gutenberg-Universität
in Mainz

vorgelegt von
Benjamin Müller
aus Wiesbaden
Mainz, Januar 2018

1. Berichterstatter:
2. Berichterstatter:

Datum der mündlichen Prüfung 16.04.2018

Zusammenfassung

Die vorliegende Arbeit beschäftigt sich mit numerischen Lösungsmethoden von sehr großen linearen Gleichungssystemen mit Anwendung im Bereich von Gitter-QCD Simulationen. Diese gehören zu den rechenintensivsten Problemen des aktuellen Hochleistungsrechnens. Die zentrale Herausforderung besteht dabei aus dem Lösen der diskretisierten Dirac-Gleichung, welche durch ein dünnbesetztes lineares Gleichungssystem mit einer halben Milliarde und mehr Unbekannten gegeben ist. Wir stellen ein hochperformantes adaptives Mehrgitterverfahren auf Basis von Gebietszerlegungsmethoden vor. Dabei werden Schwarz-Alternierende-Methode mit Aggregat-basierten Gitterhierarchien kombiniert. Das Krylov-Unterraumverfahren FGMRES bildet das Rückgrat unseres Mehrgitterverfahrens.

Weiter werden neue Verfahren zur Spektralapproximation des symmetrisierten Dirac-Operators vorgestellt, die auf Shift-Invertier-Ansätze wie der Rayleigh-Quotienten-Iteration und dem Jacobi-Davidson-Verfahren basieren. Dazu wird das Mehrgitterverfahren angepasst und mit den genannten Verfahren kombiniert. Wir zeigen, dass die resultierenden Verfahren mit in der Gitter-QCD etablierten Vorgehensweisen konkurrieren können und durch besseres Skalierungsverhalten auch und insbesondere bei zukünfig größeren Simulationen überlegen sind. Wir demonstrieren dies für physikalisch relevante Szenarien.

Abstract

This thesis deals with numerical methods solving very large linear systems of equations arising in the field of lattice QCD simulations. These are among the most computationally intensive problems of modern high-performance computing. The central challenge is to solve a discretised Dirac equation, which is given by a sparse linear system of equations with half a billion and more unknowns. We present an efficient adaptive multigrid method based on domain decomposition methods. In doing so, the Schwarz alternating procedure is combined with aggregate-based grid hierarchies. The Krylov subspace method FGMRES forms the backbone of our multigrid process.

In addition, new methods for spectral approximations of the symmetrized Dirac operator based on shift-invert approaches such as the Rayleigh quotient iteration and the Jacobi-Davidson method will be presented. For this purpose, the multigrid method is adapted and combined with the aforementioned methods. We show that the algorithms can compete with the ones currently in use in lattice QCD and may even be superior for forthcomming larger simulations due to better scaling behavior. We demonstrate this for physically relevant scenarios.

Vorwort

Das Gelingen dieser Arbeit hing von vielen Personen ab, denen ich meinen ausdrücklichen Dank ausspreche. Unter diesen Personen ist zu allererst mein Betreuer zu nennen, der mir dieses Projekt ermöglichte und mir stets mit seinem umfangreichen Wissen zur Seite stand. Ihm ist insbesondere das Zustandekommen der Kooperation mit der Arbeitsgruppe Angewandte Mathematik der Bergischen Universität Wuppertal zu verdanken. Der Zusammenarbeit mit besagter Arbeitsgruppe gilt mein besonderer Dank; die Aufenthalte in Wuppertal werden mir immer in positiver Erinnerung bleiben. Ebenfalls danke ich ausdrücklich meinem zweiten Betreuer für die vielen gemeinsamen interdisziplinären Diskussionen.
Ich bedanke mich ebenso beim Exzellenzcluster „PRISMA“ (Precision Physics, Fundamental Interactions and Structure of Matter) für die Finanzierung eines Großteils meiner Promotion, sowie beim Forschungsschwerpunkt „SRFN“ (Schwerpunkt für Rechnergestützte Forschungsmethoden in den Naturwissenschaften) für die zahlreichen interdisziplinären Workshops.
Darüber hinaus bedanke ich mich bei allen aktuellen und ehemaligen Mitgliedern der Arbeitsgruppe Numerik in Mainz, sowie der Arbeitsgruppe Funktionalanalysis Mainz, für das freundliche und produktive Umfeld (und die unzähligen Tischkicker-Spiele).

Inhaltsverzeichnis

1. Einleitung

Simulationen zur Gitter-Quantenchromodynamik (Gitter-QCD) gehören zu den rechenintensivsten Problemen im Bereich des Hochleistungsrechnens, und ein nicht unerheblicher Teil der aktuell verfügbaren Rechenleistung wird für Gitter-QCD-Simulationen aufgewendet [87]. Die zentrale Herausforderung besteht dabei aus dem Lösen der diskretisierten DIRAC[1]-Gleichung, welche im Wesentlichen durch ein *sehr* großes, dünn besetztes, lineares Gleichungssystem

$$Dz = b \tag{1.1}$$

gegeben ist. Das Ziel dieser Arbeit ist es, ein hochperformantes adaptives Mehrgitterverfahren zum Lösen der DIRAC-Gleichung vorzustellen und auf dieser Grundlage Spektralapproximationen für den symmetrisierten DIRAC-Operator zu berechnen.

Der Operator $D \equiv D(U, m)$ ist hierbei eine Diskretisierung des kontinuierlichen DIRAC-Operators aus der QCD, typischerweise die WILSON[2]-Diskretisierung, auf einem vierdimensionalen Raum-Zeit-Gitter. Der WILSON-DIRAC-Operator D hängt dabei von einem Eichfeld U und einem Massenparameter m ab. Aktuelle Simulationen arbeiten mit Gittern bestehend aus 144×64^3 Knoten und mehr, was in Gleichungssystemen mit mindestens einer halben Milliarde Unbekannten mündet [5].

Üblicherweise werden die Gleichungssysteme (1.1) mit iterativen numerischen Verfahren gelöst, überwiegend durch KRYLOV-Unterraumverfahren. Die Konvergenzrate dieser Verfahren verschlechtert sich jedoch enorm, wenn große Gitterkonfigurationen und/oder physikalisch relevante Massenparameter erreicht werden. Um dem entgegen zu wirken, müssen Präkonditionierer für besagte Unterraumverfahren entwickelt werden, die dazu im Stande sind, die Skalierungsprobleme zu reduzieren. In der Gitter-QCD sind bereits „odd-even"-Präkonditionierung, Deflation und Gebietszerlegungsmethoden verbreitet und liefern signifikante Laufzeitverbesserungen gegenüber nicht-präkonditionierten Methoden. Deren Skalierungsverhalten ist dennoch nahezu unverändert schlecht.

Hier kommen Mehrgitterverfahren ins Spiel, die in der Gitter-QCD bereits eine hohe Reputation genießen wegen ihrem Potential hohe Konvergenzraten praktisch unabhängig von Gitterweiten zu erreichen, z. B. im Bereich der elliptischen partiellen Differentialgleichungen. Aufgrund der in der Gitter-QCD involvierten Eichfelder und ihrer stochastischen Natur sind *geometrische* Mehrgitterverfahren, also Verfahren, die ausschließlich mit der vorliegenden partiellen Differentialgleichung arbeiten, trotz jahrzehntelanger Forschung nicht praktikabel [19, 52]. Daher wurden in den letzten Jahren vermehrt *adaptive algebraische* Mehrgitterverfahren konstruiert [4, 17], die direkt auf der Struktur der Operatormatrix ansetzen und z. B. in der Programmbibliothek QOPQDP [81] implementiert sind.

Ein ähnlicher in der Gitter-QCD weitverbreiteter Löser-Ansatz namens „inexact deflation“ wurde von M. Lüscher in [61] vorgeschlagen und eine (verbesserte) Implementierung ist in [62] verfügbar.

Der Fokus dieser Arbeit liegt zu Beginn auf der Herleitung des Aggregat-basierten adaptiven Mehrgitterverfahrens DD-αAMG [35], welches Aspekte der vorher genannten Verfahren aufgreift, jedoch an anderen Stellen signifikante Unterschiede aufweist. Letztendlich stellt dieses Verfahren aber durch seine Skalierbarkeit eine wesentliche Verbesserung zu den aktuell in der Gitter-QCD verbreiteten Gleichungssystemlöser dar. Die hier vorgestellten Ansätze spiegeln Resultate einer engen Kooperation mit der Arbeitsgruppe A. Frommer (Angewandte Informatik) der Universität Wuppertal wider.

Der wesentliche Beitrag der Arbeit besteht darin, das vorgestellte Mehrgitterverfahren anzupassen und mit numerischen Algorithmen zur Eigenwertbestimmung zu kombinieren. Es werden neue Verfahren vorstellt, die signifikante Verbesserungen im Bereich der Eigenmodenberechnung des symmetrischen DIRAC-Operators $Q := \Gamma_5 D$ aufweisen. Varianten der Verfahren wurden bereits in [7] veröffentlicht, eine weitere Publikation ist in Vorbereitung.

Die vorliegende Arbeit ist genauer wie folgt aufgebaut:

Kapitel 2 gibt einen Überblick zu physikalischen Hintergründen und der Herleitung der WILSON-Diskretisierung inklusive des Clover-Korrekturterms, sowie zu gewissen Eigenschaften des hergeleiteten Operators. Darüber hinaus wird statisches Präkonditionieren und das Konzept des *Smearing*, einhergehend mit einer Normalitätsanalyse von D, erläutert.

Kapitel 3 möchte eine knappe Einführung in das Gebiet der KRYLOV-Unterraumverfahren vermitteln, mit Schwerpunkt auf dem robusten GMRES-Verfahren und dessen *flexible* Variante FGMRES.

Kapitel 4 wendet sich Gebietszerlegungsmethoden zu und legt mit der Einführung der SAP-Methode als Glätter den Grundstein für das genannte Mehrgitterverfahren.

Kapitel 5 stellt die Aggregat-basierte Interpolation vor, mit der zwischen verschiedenen Gitterebenen kommuniziert wird. Ebenso ist die Adaptivität des resultierenden Verfahrens ein hier vorgestellter wichtiger Aspekt. Das Kapitel schließt mit numerischen Ergebnissen und Vergleichen zu verbreiteten anderen Lösern in der Gitter-QCD.

Schließlich präsentiert Kapitel 6 neue Resultate aus der Kombination des hergeleiteten Mehrgitterverfahrens mit numerischen Verfahren zur Eigenwertberechnung mit Fokus auf Shift-Invertier- bzw. Projektionsverfahrens-Ansätzen. Das Kapitel schließt mit Ergebnissen einer aktuellen physikalischen Anwendung, die im Rahmen einer Kooperation entstand, in die zusätzlich die Arbeitsgruppe G. Bali (Hochenergiephysik) der Universität Regensburg eingebunden war.

Anmerkungen*

[1] Paul Adrien Maurice Dirac (* 8. August 1902 in Bristol; † 20. Oktober 1984 in Tallahassee) war ein britischer Physiker, Nobelpreisträger und Mitbegründer der Quantenphysik. Eine von Diracs wichtigsten Entdeckungen ist in der Dirac-Gleichung von 1928 beschrieben, in der Einsteins spezielle Relativitätstheorie und die Quantenphysik erstmals zusammengebracht werden konnten.

[2] Kenneth Geddes Wilson (* 8. Juni 1936 in Waltham, Massachusetts; † 15. Juni 2013 in Saco, Maine) war ein US-amerikanischer Physiker und Nobelpreisträger. Er war Schüler von Murray Gell-Mann.

*Alle Angaben aus der deutschen Wikipedia, stand 2017

2. Quantenchromodynamik

Als Teil der Teilchenphysik, genauer der relativistischen Quantenmechanik, beschreibt die Quantenchromodynamik (QCD) die starke Wechselwirkung der kleinsten der Menschheit bekannten Elementarteilchen, die der Quarks. Im Unterschied zur Eichtheorie der Quantenelektrodynamik (QED) besitzen Quarks neben der elektrischen- noch eine zusätzliche Ladung, die Farbladung (daher der Name *Chromo*dynamik). Quarks bilden zusammen mit den Leptonen (z. B. dem Elektron) und den Eichbosonen (insbesondere dem Gluon) die fundamentalen Bestandteile der Materie (siehe Abbildung 2.1).

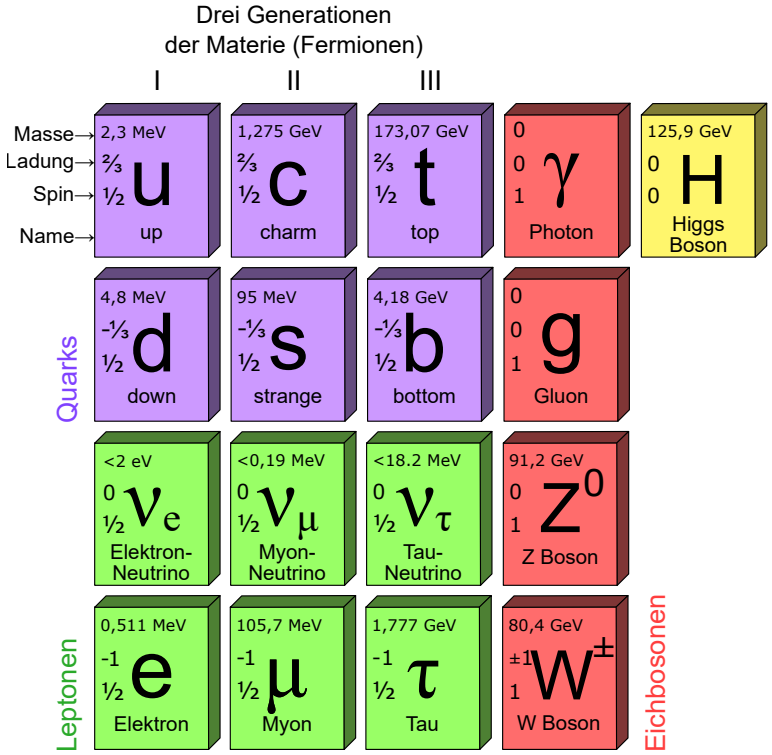


Abbildung 2.1: Das Standardmodell mit Quarks, Leptonen und Eichbosonen*.

Quarks treten in insgesamt sechs sog. Quark-Flavours auf: Up, Down, Charm, Strange, Top und Bottom und deren jeweiligen Antiteilchen (gegeben durch eine Anti-Farbladung). Quarks treten allerdings niemals einzeln, sondern in Gruppierungen, sog. Hadronen, auf; Dieses Phänomen wird als *Confinement* [110] bezeichnet. Das Proton, als Beispiel eines stabilen Hadrons, besteht aus zwei Up-Quarks und einem Down-Quark. Der Zustand eines Fermions (insbesondere eines Quarks, vgl. Abbildung 2.1) wird durch die DIRAC-Gleichung beschrieben, welche bereits im Jahre

*Ursprüngliche Quelle: Fermilab, Office of Science, United States Department of Energy, Particle Data Group

1928 entwickelt wurde. Der zugehörige DIRAC-Operator, kann in der kontinuierlichen Theorie geschrieben werden als

$$\mathcal{D} = \sum_{\mu=1}^{4} \gamma_\mu \otimes (\partial_\mu + A_\mu),$$

wobei die Summierung über $\mu = 1, 2, 3, 4$ zum vierdimensionalen Raumzeitkontinuum korrespondiert. Weiter ist $\partial_\mu = \partial/\partial x_\mu$, wobei x_μ die Komponenten des Raumzeitpunkts $x \in \mathbb{R}^4$ bezeichnet. Die Eichfelder A_μ repräsentieren die Eichmatrizen $A_\mu(x) \in \mathfrak{su}(3)$, welche Elemente der LIE[3]-Algebra[†] über der speziellen unitären Gruppe $SU(3)$[‡] sind.

Die DIRAC-Matrizen $\gamma_1, \gamma_2, \gamma_3, \gamma_4 \in \mathbb{C}^{4\times4}$ sind hermitesche, unitäre Generatoren der CLIFFORD[4]-Algebra $Cl_{0,4}(\mathbb{R})$. Details zu diesen Objekten folgen im nächsten Abschnitt.

Erwähnt sei, dass es (viele) alternative Notationen gibt, wobei geringere Abweichungen z. B. darin bestehen, dass die imaginäre Einheit i aus den Eichfeldern A_μ ausgeklammert und vorangestellt wird.

Lösungen der DIRAC-Gleichung, bzw. Vorhersagen über Hadronen (u. a. über ihre Observablen wie Masse, beteiligte Fermionen, gebundene Eichbosonen, vgl. [55]) können in der QCD weder analytisch noch durch klassische Störungstheorie, sondern nur durch numerische Simulation bestimmt werden (zumindest bei großer Kopplungskonstante, wie sie im Niederenergiebereich auftritt). Hierfür werden wir im folgenden Kapitel Konzepte der Gitter-QCD sowie vor allem die WILSON-Diskretisierung des DIRAC-Operators vorstellen und herleiten. Das folgende Kapitel beruht größtenteils auf den QCD-Theorie Abschnitten in [39, 16, 89] und [35]. Für einen tieferen Einblick in die Materie sei hier auf [39, 24] sowie [69] verwiesen.

2.1 Kontinuierliche Theorie

Die physikalischen Hintergründe weitestgehend beiseitelassend, konzentrieren wir uns in diesem Kapitel auf die mathematische Konstruktion des DIRAC-Operators. Wir beginnen mit dem Notieren von Quarks und Gluonen in mathematisch handhabbaren Ausdrücken:

2.1.1 Definition

Seien $\mathcal{C} := \{1, 2, 3\}$ die Menge der *Farb*-Indizes, $\mathcal{S} := \{0, 1, 2, 3\}$ die *Spin*- oder DIRAC-Indizes sowie

$$\begin{aligned}\psi \,:\, \mathbb{R}^4 &\to \mathbb{C}^{12} \cong \mathbb{C}^{\mathcal{C}\times\mathcal{S}},\\ x &\mapsto \big(\psi_{10}(x), \psi_{20}(x), \psi_{30}(x), \psi_{11}(x), \ldots, \psi_{33}(x)\big)^T\end{aligned}$$

[†]Die zur $SU(n)$ gehörende LIE-Algebra $\mathfrak{su}(n)$ entspricht dem Tangentialraum am Einselement der Gruppe. Sie besteht aus dem Raum aller schiefhermiteschen $n\times n$-Matrizen mit Spur Null.

[‡]Die spezielle unitäre Gruppe $SU(n)$ besteht aus den unitären $n\times n$-Matrizen mit komplexen Einträgen, deren Determinante Eins beträgt.

eine differenzierbare Funktion. Dann wird ψ als *Quarkfeld*, bzw. $\psi(x)$ als DIRAC-*Spinor* bezeichnet. Wir sammeln diese in $\mathcal{M} := \{\psi : \psi \text{ ist Quarkfeld}\}$. Für $\mu = 1, 2, 3, 4$ sind

$$\begin{aligned} A_\mu : \mathbb{R}^4 &\to \mathfrak{su}(3), \\ x &\mapsto A_\mu(x) \end{aligned}$$

sog. *Eichfelder*, welche die Gluonen, also die Kopplung der Quarks, repräsentieren. ◊

Die Komponenten des Spinors $\psi(x)$ werden typischerweise mit $\psi_{c\sigma}(x)$ notiert, wobei sich $c \in \mathcal{C}$ auf den Farb- und $s \in \mathcal{S}$ auf den Spin-Index bezieht. Die Notation für einen fixierten Spin-Index $\sigma \in \mathcal{S}$, $\psi_\sigma(x) = \big(\psi_{1\sigma}(x), \psi_{2\sigma}(x), \psi_{3\sigma}(x)\big)^T$, ist ebenso verbreitet, da die Eichfelder nicht-trivial auf die Farbkomponenten und trivial auf die Spinkomponenten des Spinors $\psi(x)$ im Sinne von

$$(A_\mu\psi)(x) := (I_4 \otimes A_\mu(x))\psi(x)$$

wirken. Essentiell für die Wirkung des DIRAC-Operators auf die Quarkfelder ψ sind, neben den oben bereits verwendeten Tensorprodukten (bzw. KRONECKER[5]-Produkten), bestimmte 4×4-Matrizen:

2.1.2 Definition

Die vier hermiteschen, unitären Matrizen $\gamma_\mu \in \mathbb{C}^{4\times4}$, $\mu = 1, 2, 3, 4$ erzeugen die CLIFFORD-*Algebra* $Cl_{0,4}(\mathbb{R})$ genau dann, wenn für alle $\mu, \nu = 1, 2, 3, 4$ gilt

$$\gamma_\mu\gamma_\nu + \gamma_\nu\gamma_\mu = \begin{cases} 2I_4, & \mu = \nu, \\ 0, & \text{sonst.} \end{cases} \tag{2.1}$$

Die (nicht-eindeutigen) Matrizen γ_μ werden als DIRAC- oder γ-*Matrizen* bezeichnet. ◊

Hintergründe zur Notation und Details zur Rolle der CLIFFORD-Algebren in diesem Teil der Physik kann z. B. im Buch [55] nachgegangen werden.

Wichtig ist anzumerken, dass im Gegensatz zu den Eichmatrizen $A_\mu(x)$ und dem Spinor $\psi(x)$, die γ-Matrizen nicht von der Raumzeit x abhängen. Die Multiplikation einer γ-Matrix mit einem Quarkfeld ist wie folgt definiert:

$$(\gamma_\mu\psi)(x) := (\gamma_\mu \otimes I_3)\psi(x).$$

Nun sind wir in der Lage, den DIRAC-Operator und dessen Wirkung auf Quarkfelder zu definieren.

2.1.3 Definition

Sei $\mathcal{M}$ der Raum der Quarkfelder, dann ist der kontinuierliche (massefreie) DIRAC-Operator eine lineare Abbildung

$$\mathcal{D} : \mathcal{M} \to \mathcal{M}$$

definiert durch

$$\mathcal{D} := \sum_{\mu=1}^{4} \gamma_\mu \otimes (\partial_\mu + A_\mu), \tag{2.2}$$

wobei $\partial_\mu = \partial/\partial x_\mu$ die partielle Ableitung nach x_μ, $\mu = 1, 2, 3, 4$, bezeichnet. Die Auswertung von $\mathcal{D}\psi$ an einem Punkt $x \in \mathbb{R}^4$ ist via

$$(\mathcal{D}\psi)(x) = \sum_{\mu=1}^{4} \big(\gamma_\mu(\partial_\mu + A_\mu)\psi\big)(x).$$

gegeben. ◊

Der Differentialoperator $\partial_\mu + A_\mu$ ist quantenmechanisch eine „minimale Kopplung“ und darüber hinaus so konstruiert, dass $\big((\partial_\mu + A_\mu)\psi\big)(x)$ auf dieselbe Weise unter lokalen Eichtransformationen transformiert, wie $\psi(x)$, er ist also *(Eich-)kovariant*. Diese Eichtransformationen entsprechen einem lokalen Wechsel des Farb-Koordinatensystems. A_μ, als Teil des kovarianten Differentialoperators, kann hierbei als Kopplung verschiedener, infinitesimal nah beieinander liegender, Raumzeitpunkte verstanden werden. Eigenschaften der γ-Matrizen γ_μ garantieren auch, dass $\mathcal{D}\psi(x)$ unter Transformationen der speziellen Relativitätstheorie genau so transformiert wie der Spinor $\psi(x)$. Dies nennt man lokale Eichinvarianz und ist ein zentrales Prinzip des Standardmodells der elementaren Teilchenphysik. Für Details siehe z. B. [86].

2.2 Gitter-QCD

Die Diskretisierung der Raumzeit des Euklid[6]ischen Kontinuums durch ein hyperkubisches Gitter mit regelmäßigem Gitterabstand, wohl aber unterschiedlicher Ausdehnung in Zeit N_t und Raum N_r, hat vor allem zwei Gründe: Sie ermöglicht numerische Simulationen zum Einen und das Erforschen von nicht-perturbativen Phänomenen der QCD zum Anderen. Die auf Wilson [110] zurückgehende Gitter-QCD assoziiert die fermionischen Spinor-Felder $\psi(x)$ dabei mit Gitterknoten $x = [x_1, x_2, x_3, x_4]^T$ und die Eichfelder werden durch *Links* zwischen diesen Knoten dargestellt. Kernaufgabe bei Gitter-QCD Simulationen wird es auf lange Sicht immer sein, die diskretisierte Dirac-Gleichung für eine gegebene (oder oft mehrere) rechte Seite zu Lösen. In diesem Abschnitt werden wir, neben der Einführung in die Prinzipien der Wilson-Diskretisierung, auch auf die numerischen Eigenschaften des resultierenden linearen Operators eingehen.

2.2.1 Definition

Wir betrachten ein periodisches regelmäßiges vierdimensionales Raumzeitgitter $\mathcal{L}$ mit Gitterweite a. Für je zwei $x, y \in \mathcal{L}$ soll ein $p \in \mathbb{Z}^4$ existieren, sodass

$$y = x + ap.$$

So können wir Shiftoperationen mittels eines Shiftvektors $\hat{\mu} \in \mathbb{R}^4$

$$\hat{\mu}_\nu := \begin{cases} a, & \nu = \mu, \\ 0, & \text{sonst}, \end{cases}$$

für $\mu \in \{1, 2, 3, 4\}$ und $\nu = 1, 2, 3, 4$, definieren. ◊

Um Quarkfelder ψ auf dem Gitter zu definieren, genügt es Auswertungen an den Gitterpunkten vorzunehmen:

$$\begin{aligned} \psi \,:\, \mathcal{L} &\to \mathbb{C}^{12}, \\ x &\mapsto \psi(x). \end{aligned}$$

Bis auf die Tatsache, dass diese Funktion nicht mehr differenzierbar ist, werden dieselben Notationen für den Spinor $\psi(x)$ übernommen, d. h. Spin- und Farbindizes $\psi_{c\sigma}(x)$ für $c \in \mathcal{C}$, $\sigma \in \mathcal{S}$ (vgl. Definition 2.1.1). Nun wenden wir uns den Eichfeldern A_μ zu, welche in der kontinuierlichen QCD infinitesimal nah beieinanderliegende Punkte der Raumzeit koppeln und durch diskrete Variablen $U_\mu(x)$ ersetzt werden müssen.

2.2.2 Definition

Zu jeder Eichmatrix $A_\mu(x)$ ist die korrespondierende Diskretisierung $U_\mu(x)$ durch ein (quantenmechanisches) Pfadintegral entlang der Kante $(x, x + \hat{\mu})$ gegeben, genauer:

$$U_\mu(x) := \exp\left(- \int_x^{x+\hat{\mu}} A_\mu(s) \mathrm{d}s \right) \approx e^{-aA_\mu(x+\frac{1}{2}\hat{\mu})}.$$

Das diskretisierte Eichfeld $U := \{U_\mu(x) \,:\, x \in \mathcal{L},\, \mu = 1, 2, 3, 4\}$ nennt man *(Eich-)Konfiguration*. ◊

Die analytische Transformation von A_μ zu U_μ ist allerdings nur von theoretischer Bedeutung, da Gitter-QCD-Berechnungen immer direkt von diskret berechneten Eichkonfigurationen U ausgehen. Anschaulich betrachtet, „lebt“ die Variable $U_\mu(x)$ nicht im oder auf dem Gitterknoten x, sondern auf der Kante $(x, x + \hat{\mu})$; Man spricht von *Linkvariablen*, denn $U_\mu(x)$ stellt die Kopplung zwischen x und $x + \hat{\mu}$ dar. Insbesondere ist die Kopplung von $x + \hat{\mu}$ und x durch $U_\mu(x)^{-1}$ gegeben, vgl. Definition 2.2.2. Darüber hinaus gilt

$$U_\mu(x) \in SU(3), \text{ insbesondere } U_\mu(x)^{-1} = U_\mu(x)^H,^{\S}$$

wobei $U_\mu(x)^H \equiv U_\mu^H(x)$ hier und im Folgenden die Adjungierte von $U_\mu(x)$ bezeichnet.

Abbildung 2.2 illustriert die hier verwendete Notationskonvention.

[§] Für A spurfrei gilt $\det\big(\exp(A)\big) = \exp\big(\mathrm{Spur}(A)\big) = 1$ sowie für A schiefhermitesch $\exp(A)^{-1} = \exp(A)^H$.

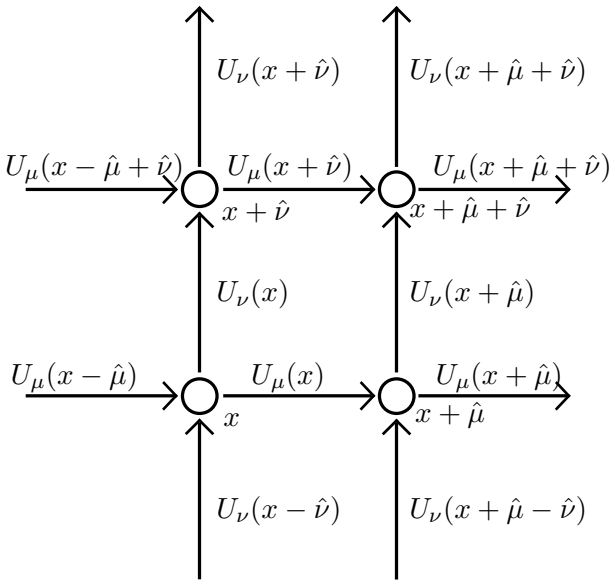


Abbildung 2.2: Die verwendete Notationskonvention auf dem Gitter.

Wir wenden uns nun der Diskretisierung der kovarianten Ableitung zu, welche auf verschiedene Weise vollzogen werden kann; die hier beschriebene ist die am häufigsten verwendete WILSON-Diskretisierung (vgl. [111]).

2.2.3 Definition

Sei A_μ ein Eichfeld und U_μ die zugehörige Eichkonfiguration. Die Definition der kovarianten rechtsseitigen Differenzenquotienten

$$(\Delta^\mu \psi_\sigma)(x) := \frac{U_\mu^H(x)\psi_\sigma(x+\hat{\mu}) - \psi_\sigma(x)}{a} \approx (\partial_\mu + A_\mu)\psi_\sigma(x)$$

und der kovarianten linksseitigen Differenzenquotienten

$$(\Delta_\mu \psi_\sigma)(x) := \frac{\psi_\sigma(x) - U_\mu(x-\hat{\mu})\psi_\sigma(x-\hat{\mu})}{a},$$

führt mittels kovarianter zentralem Differenzenquotienten zu folgender naiven Diskretisierung des (massefreien) DIRAC-Operators $\mathcal{D}$ (2.2):

$$D_N := \sum_{\mu=1}^{4} \gamma_\mu \otimes (\Delta_\mu + \Delta^\mu)/2. \tag{2.3}$$

$\Diamond$

Bevor wir zu Problemen und Verbesserungen dieser naiven Diskretisierung kommen, führen wir ein Lemma an, welches wir später benötigen.

2.2.4 Lemma

Für die vorwärts und rückwärts kovarianten finiten Differenzen gilt

$$(\Delta^\mu)^H \;=\; -\Delta_\mu.$$

Beweis. Seien ψ und η beliebige Quarkfelder. Dann gilt

$$\begin{aligned}\langle\psi_\sigma, \Delta^\mu\eta_\sigma\rangle &\equiv \sum_{x\in\mathcal{L}}\langle\psi_\sigma(x), \Delta^\mu\eta_\sigma(x)\rangle \\ &= \frac{1}{a}\sum_{x\in\mathcal{L}}\langle\psi_\sigma(x), U_\mu^H(x)\eta_\sigma(x+\hat{\mu})\rangle - \frac{1}{a}\sum_{x\in\mathcal{L}}\langle\psi_\sigma(x), \eta_\sigma(x)\rangle \\ &= (*).\end{aligned}$$

Die Periodizität des Gitters $\mathcal{L}$ erlaubt es nun, in der ersten Summe eine Indextransformation $x \mapsto x+\hat{\mu}$ anzuwenden

$$\begin{aligned}(*) &= \frac{1}{a}\sum_{x\in\mathcal{L}}\langle\psi_\sigma(x-\hat{\mu}), U_\mu^H(x-\hat{\mu})\eta_\sigma(x)\rangle - \frac{1}{a}\sum_{x\in\mathcal{L}}\langle\psi_\sigma(x), \eta_\sigma(x)\rangle \\ &= -\frac{1}{a}\sum_{x\in\mathcal{L}}\langle\psi_\sigma(x) - U_\mu(x-\hat{\mu})\psi_\sigma(x-\hat{\mu}), \eta_\sigma(x)\rangle \\ &= -\langle\Delta_\mu\psi_\sigma, \eta_\sigma\rangle.\end{aligned}$$

□

Insbesondere ist die zentrale (kovariante) finite Differenz $(\Delta^\mu + \Delta_\mu)/2$ schiefhermitesch. Da die γ-Matrizen γ_μ hermitesch sind (vgl. Definition 2.1), ergibt sich folgendes Korollar.

2.2.5 Korollar

Die naive Diskretisierung D_N aus (2.3) ist schiefhermitesch, d. h.,

$$D_N^H \;=\; -D_N.$$

Die naive Diskretisierung D_N erzeugt in dieser Form unphysikalische Eigenvektoren[¶], auch bekannt unter dem Problem der *Fermionenverdopplung*, welches untrennbar mit der *chiralen Symmetrie* auf dem Gitter verknüpft ist. Dies sind zwei der vier (hier nicht näher genannten) Eigenschaften für Diskretisierungen des DIRAC-Operators, die nach dem Nielson-Nimomiya-Theorem nicht gleichzeitig erfüllt werden können. Für Details verweisen wir auf [77] und [78]. Im Folgenden stellen wir eine wiederum von WILSON [111] vorgeschlagene Möglichkeit vor, die Fermionenverdopplung zu vermeiden (welche jedoch die chirale Symmetrie explizit bricht): Der Stabilisierungsterm $a\Delta_\mu\Delta^\mu$, welcher ein (kovarianter) zentraler Differenzenquotient zweiter Ordnung ist. Er wird auch WILSON-Fermion genannt.

[¶]Der Eigenraum eines Eigenwertes von D_N ist 16-dimensional, aber nur ein 1-dimensionaler Eigenraum korrespondiert zu jeweils einer Eigenfunktion des kontinuierlichen Operators $\mathcal{D}$.

2.2.6 Definition

Gegeben sei eine Konfiguration U auf einem Gitter $\mathcal{L}$ mit $n_{\mathcal{L}}$ Knoten, Gitterweite a, sowie einem Massenparameter m_0. Dann ist die WILSON-Diskretisierung des DIRAC-Operators (auch bekannt als WILSON-DIRAC-Operator) definiert durch

$$D_W := \frac{m_0}{a} I_{12n_{\mathcal{L}}} + \frac{1}{2} I_{n_{\mathcal{L}}} \sum_{\mu=1}^{4} \Big(\gamma_\mu \otimes (\Delta_\mu + \Delta^\mu) - a I_4 \otimes \Delta_\mu \Delta^\mu \Big). \tag{2.4}$$

◊

Zum Massenparameter m_0, welcher sich auf die Quarkmasse bezieht, sei an dieser Stelle nur gesagt, dass dieser die größten Schwierigkeiten insbesondere bei physikalisch relevanten kleinen Quarkmassen hervorruft, bis dahingehend, dass frühere Gitter-QCD-Simulationen überhaupt nur mit unrealistisch großen Werten möglich waren. Für physikalische Details zum Massenparameter sei exemplarisch auf [69] verwiesen.

Die Vertauschungseigenschaften (2.1) der γ-Matrizen implizieren eine nicht-triviale Symmetrie des WILSON-DIRAC-Operators D_W.

2.2.7 Lemma

Für $\gamma_5 := \gamma_1\gamma_2\gamma_3\gamma_4$ gilt $\gamma_5\gamma_\mu = -\gamma_\mu\gamma_5$, $\mu = 1,2,3,4$, vgl. (2.1), und mit $\Gamma_5 := I_{n_{\mathcal{L}}} \otimes \gamma_5 \otimes I_3$ weist der WILSON-DIRAC-Operator eine sog. Γ_5-Symmetrie auf, d. h.,

$$(\Gamma_5 D_W)^H = \Gamma_5 D_W.$$

Beweis. Aufgrund der Hermitizität von γ_μ und γ_5 ist $\gamma_5\gamma_\mu$ schiefhermitesch. Mit Hilfe von Lemma 2.2.4 sehen wir, dass $(\gamma_5\gamma_\mu) \otimes (\Delta_\mu + \Delta^\mu)$ als Tensorprodukt zweier schiefhermitschen Operatoren hermitesch ist. Dasselbe Lemma impliziert, dass

$$(\Delta_\mu\Delta^\mu)^H = (\Delta^\mu)^H(\Delta_\mu)^H = (-\Delta_\mu)(-\Delta^\mu) = \Delta_\mu\Delta^\mu$$

und somit auch $I_4 \otimes \Delta_\mu\Delta^\mu$ hermitesch ist. Außerdem kommutiert Γ_5 (bzw. genauer $\gamma_5 \otimes I_3$) mit diesem Summanden, sodass die Behauptung folgt. □

2.2.8 Bemerkung

Mittels der beiden Projektoren

$$\pi_\mu^+ := \frac{I_4 + \gamma_\mu}{2} \quad \text{und} \quad \pi_\mu^- := \frac{I_4 - \gamma_\mu}{2}$$

lässt sich D_W, angewendet auf einen diskretisierten Spinor $\psi(x)$, schreiben als

$$D_W\psi(x) = \frac{4 + m_0}{a}\psi(x) - \frac{1}{a}\sum_{\mu=1}^{4} \Big(\pi_\mu^- \otimes U_\mu^H(x)\psi(x+\hat{\mu}) + \pi_\mu^+ \otimes U_\mu(x-\hat{\mu})\psi(x-\hat{\mu}) \Big).$$

Aus dieser Formulierung geht nochmals hervor, dass D_W die Γ_5-Symmetrie aufweist, allerdings nicht hermitesch ist, da $(\pi_\mu^+)^H \neq \pi_\mu^-$ (Darüber hinaus ist D_W im Gegensatz zu $\mathcal{D}$ nicht normal, vgl. Abschnitt 2.4). Neben kleineren Nachteilen, wie einer notwendigen Reskalierung des Massenparameters, führt die Verwendung von WILSON-Fermionen zum Auftreten reeller Eigenwerte von D_W (vgl. Abbildung 2.4), auch *exzeptionelle Konfigurationen* genannt, was bei der Verwendung des hermiteschen Operators $\Gamma_5 D_W$ zu sehr kleinen Eigenwerten führt und dieser damit sehr schlecht konditioniert ist. Darüber hinaus führt der WILSON-Term im Gegensatz zum naiven Ansatz D_N dazu, dass Gitterartefakte nur noch in der Größenordnung $\mathcal{O}(a)$ verschwinden. Um dieses letztgenannte Problem zu beheben, ist die im Folgenden beschriebene *Clover-Wirkung*, welche auf Sheikholeslami und Wohlert [96] zurückgeht, eine (mit einem passenden Parameter) sinnvolle Modifikation von D_W. Zunächst benötigen wir hierfür die Definition einer *Plakette*.

2.2.9 Definition

Gegeben sei eine Konfiguration von Linkvariablen $\{U_\mu(x)\}$, dann ist die *Plakette* $Q_x^{\mu,\nu}$ am Gitterknoten x definiert durch

$$Q_x^{\mu,\nu} := U_\mu^H(x) U_\nu^H(x+\hat{\mu}) U_\mu(x+\hat{\nu}) U_\nu(x). \tag{2.5}$$

$\Diamond$

Eine Plakette ist also ein Produkt von Linkvariablen, entlang eines Zykels der Länge vier, was in der (μ,ν)-Ebene einem Quadrat entspricht:

$Q_x^{\mu,\nu} \quad \widehat{=}$  .

Entsprechend sind die Plaketten der anderen Quadranten wie folgt definiert

$Q_x^{\mu,-\nu} \quad \widehat{=}$  , $Q_x^{-\mu,-\nu} \quad \widehat{=}$  , $Q_x^{-\mu,\nu} \quad \widehat{=}$  .

Offenbar sind verschiedene benachbarte Plaketten miteinander konjugiert, in dem Sinne, dass nur der Startpunkt ein anderer ist. Beispielsweise gilt $Q_{x+\hat{\mu}}^{-\mu,\nu} = U_\mu(x) Q_x^{\mu\nu} U_\mu^H(x)$. Nun kommen wir zum Sheikholeslami-Wohlert- oder auch *Clover*-(Korrektur-)Term:

2.2.10 Definition

Mit

$$Q_{\mu\nu}(x) := Q_x^{\mu,\nu} + Q_x^{\mu,-\nu} + Q_x^{-\mu,\nu} + Q_x^{-\mu,-\nu}$$

wird der *Clover*-Term C definiert durch

$$C(x) := \frac{c_{\mathsf{sw}}}{32a} \sum_{\substack{\mu,\nu=1 \\ \mu\neq\nu}}^{4} (\gamma_\mu\gamma_\nu) \otimes \big(Q_{\mu\nu}(x) - Q_{\nu\mu}(x)\big), \tag{2.6}$$

wobei $c_{\mathsf{sw}} > 0$. Für eine Illustrierung der Wirkung des Clover-Terms auf dem Gitter siehe Abbildung 2.3. ◊

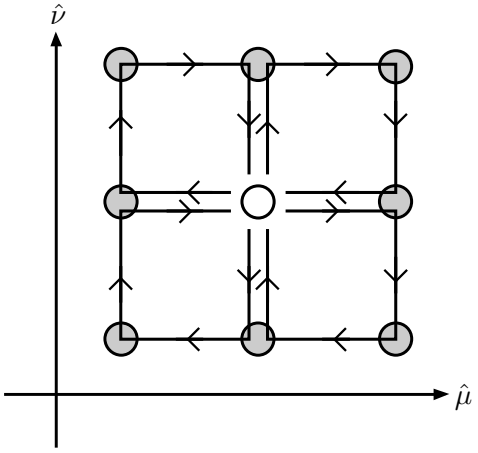


Abbildung 2.3: Graphische Darstellung des Clover- (zu Deutsch „Kleeblatt-")Terms.

Der Clover-Term reduziert die lokalen Gitterartefakte und gleicht damit die negativen Auswirkungen der WILSON-Fermionen aus. Diese Reduzierung hängt allerdings stark vom Sheikholeslami-Wohlert-Koeffizient c_{sw} ab, der geeignet gewählt werden muss. Während Sheikholeslami und Wohlert den Koeffizienten zunächst auf dem Wert Eins beließen [96], entwickelte Wohlert später Methoden, um diesen genauer zu bestimmen, bzw. mit der unrenormierten Kopplungskonstante g der QCD in Beziehung zu setzen ($c_{\mathsf{sw}} = 1 + 0.2659g^2$) [112]. Neuere Arbeiten zu diesem Parameter, vor allem der ALPHA-Kollaboration[‖], sind in [67, 29] und, speziell für den (wichtigen) Fall der zwei-Flavour-Theorie, in [51] zu finden.

2.2.11 Lemma

Der Clover-Term (2.6) ist hermitesch.

Beweis. Nach Definition 2.1.2 ist γ_μ hermitesch und es gilt

$$\gamma_\nu\gamma_\mu = -\gamma_\mu\gamma_\nu, \quad \text{solange nur } \mu \neq \nu.$$

Mit anderen Worten, $\gamma_\mu\gamma_\nu$ ist schiefhermitesch. Darüber hinaus ist

$$(Q_x^{\mu,\nu})^H = U_\nu^H(x)U_\mu^H(x+\hat{\nu})U_\nu(x+\hat{\mu})U_\mu(x) = Q_x^{\nu,\mu},$$

und daher $Q_{\mu\nu}^H(x) = Q_{\nu\mu}(x)$. Demnach ist auch $Q_{\mu\nu}(x) - Q_{\nu\mu}(x)$ schiefhermitesch und das Produkt $(\gamma_\mu\gamma_\nu) \otimes \big(Q_{\mu\nu}(x) - Q_{\nu\mu}(x)\big)$ hermitesch sowie damit der gesamte Clover-Term C. □

[‖] DESY Zeuthen, `https://www-zeuthen.desy.de/alpha/`.

2.2.12 Definition

Für ein Quarkfeld ψ und Gitterknoten x ist der WILSON-DIRAC-Operator mit Clover-Term gegeben durch

$$D\psi(x) := D_W\psi(x) - C(x)\psi(x). \tag{2.7}$$

Ausgeschrieben und wichtig für die Auffassung von D als Matrix ist folgende Darstellung:

$$\begin{aligned}(D\psi)(x) =& \frac{1}{a}\Big((m_0+4)I_{12} - \frac{c_{\mathsf{sw}}}{32}\sum_{\mu,\nu=1}^{4}(\gamma_\mu\gamma_\nu)\otimes\big(Q_{\mu\nu}(x) - Q_{\nu\mu}(x)\big)\Big)\psi(x)\\ &- \frac{1}{2a}\sum_{\mu=1}^{4}\big((I_4-\gamma_\mu)\otimes U_\mu^H(x)\big)\psi(x+\hat{\mu})\\ &- \frac{1}{2a}\sum_{\mu=1}^{4}\big((I_4+\gamma_\mu)\otimes U_\mu(x-\hat{\mu})\big)\psi(x-\hat{\mu}).\end{aligned} \tag{2.8}$$

$\Diamond$

Offenbar erhält man für $c_{\mathsf{sw}} = 0$ wieder den ursprünglichen WILSON-DIRAC-Operator. Der Clover-WILSON-DIRAC-Operator D erhält die Γ_5-Symmetrie, gewisse Spektralsymmetrien gehen aber verloren:

2.2.13 Lemma

(i) Der Clover-WILSON-DIRAC-Operator D ist Γ_5-symmetrisch, d. h.,

$$(\Gamma_5 D)^H = \Gamma_5 D.$$

(ii) Jeder Rechtseigenvektor ψ_λ zum Eigenwert λ von D korrespondiert zu einem Linkseigenvektor

$$\hat{\psi}_{\bar{\lambda}} = \Gamma_5\psi_\lambda$$

des Eigenwerts $\bar{\lambda}$ von D und umgekehrt. Mit anderen Worten, das Spektrum von D ist symmetrisch bezüglich der reellen Achse.

(iii) Das Spektrum von D_W ist symmetrisch bezüglich der vertikalen Gerade $\mathrm{Re}(z) = \frac{4+m_0}{a}$, d. h.,

$$\lambda \in \sigma(D_W) \quad \Rightarrow \quad \frac{2(4+m_0)}{a} - \lambda \in \sigma(D_W).$$

Beweis. (i) Wegen Lemma 2.2.11 ist C hermitesch. Es bleibt also nur die Vertauschbarkeit von C mit Γ_5 zu untersuchen; dies ist gewährleistet, da

$$\gamma_5(\gamma_\mu\gamma_\nu) = (\gamma_\mu\gamma_\nu)\gamma_5.$$

Also ist der Clover-Term Γ_5-symmetrisch. Nach Lemma 2.2.7 ist der WILSON-DIRAC-Operator D_W Γ_5-symmetrisch, insbesondere ist es auch D als Differenz aus beidem.

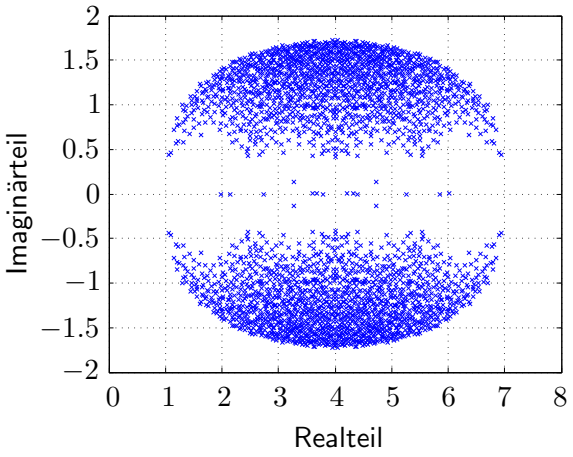


Abbildung 2.4: Spektrum eines 4^4 WILSON-DIRAC-Operator mit $m_0 = 0$ und $c_{\text{sw}} = 0$.

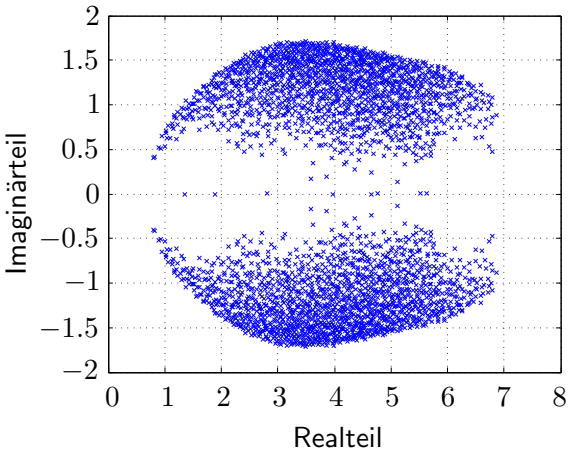


Abbildung 2.5: Spektrum eines 4^4 Clover-WILSON-DIRAC-Operator mit $m_0 = 0$ und $c_{\text{sw}} = 1$.

(ii) Wie eben gesehen, ist $D^H = \Gamma_5 D \Gamma_5$ und daher

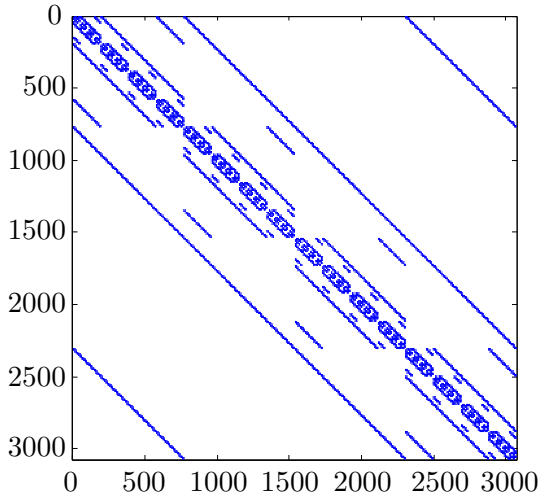
$$D\psi_\lambda = \lambda\psi_\lambda \quad \Leftrightarrow \quad \psi_\lambda^H D^H = \bar{\lambda}\psi_\lambda^H \quad \Leftrightarrow \quad (\Gamma_5\psi_\lambda)^H D = \bar{\lambda}(\Gamma_5\psi_\lambda)^H.$$

(iii) Da die Diskretisierung D_W ausschließlich über direkte Nachbarschaftsrelationen realisiert ist (im Gegensatz zum Operator C, welcher diagonale Relationen aufweist), existiert eine Rot-Schwarz-Ordnung [26] der Raumzeitpunkte so, dass der Operator durch Umsortieren dieser auf eine spezielle Blockstruktur gebracht werden kann (vgl. auch für ein 4^4-Gitter-Beispiel Abbildung 2.6):

$$D_W - \frac{4+m_0}{a} I_{12n_{\mathcal{L}}} = \begin{bmatrix} 0 & D_{\text{rs}} \\ D_{\text{sr}} & 0 \end{bmatrix}.$$

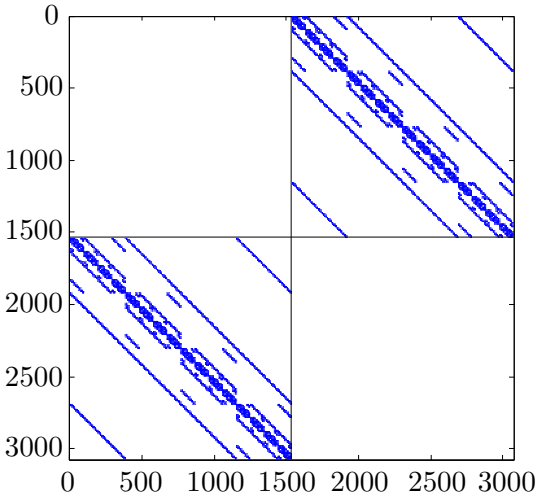
Falls nun $x = (x_{\text{r}}, x_{\text{s}})$ ein Eigenvektor von $D_W - \frac{4+m_0}{a} I_{12n_{\mathcal{L}}}$ zum Eigenwert λ ist, dann ist $x' = (x_{\text{r}}, -x_{\text{s}})$ offenbar ein Eigenvektor von $D_W - \frac{4+m_0}{a} I_{12n_{\mathcal{L}}}$ zum Eigenwert $-\lambda$. Ein Shift in Richtung $\frac{4+m_0}{a}$ liefert die Behauptung. □

Um D abschließend in Matrixform angeben zu können, müssen wir uns auf eine Repräsentation der γ-Matrizen wie folgt festlegen:



Anzahl der nicht-Null Einträge: 147456

Abbildung 2.6: Matrixstruktur von $D_W - \frac{4+m_0}{a} I_{12n_\mathcal{L}}$ *ohne* Rot-Schwarz-Umordnung (4^4-Gitter).



Anzahl der nicht-Null Einträge: 147456

Abbildung 2.7: Matrixstruktur von $D_W - \frac{4+m_0}{a} I_{12n_\mathcal{L}}$ *nach* Rot-Schwarz-Umordnung (4^4-Gitter).

$$\gamma_1 = \begin{bmatrix} & & & \mathrm{i} \\ & & \mathrm{i} & \\ & -\mathrm{i} & & \\ -\mathrm{i} & & & \end{bmatrix}, \gamma_2 = \begin{bmatrix} & & & -1 \\ & & 1 & \\ & 1 & & \\ -1 & & & \end{bmatrix}, \gamma_3 = \begin{bmatrix} & & \mathrm{i} & \\ & & & -\mathrm{i} \\ -\mathrm{i} & & & \\ & \mathrm{i} & & \end{bmatrix}, \gamma_4 = \begin{bmatrix} & & & 1 \\ & & 1 & \\ & 1 & & \\ 1 & & & \end{bmatrix}, \tag{2.9}$$

was auf

$$\gamma_5 = \gamma_1\gamma_2\gamma_3\gamma_4 = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & -1 & \\ & & & -1 \end{bmatrix} \tag{2.10}$$

führt und bedeutet, dass γ_5 trivial auf die Spins mit Index Null und Eins und als Vorzeichenwechsel bei den Spins mit Index Zwei und Drei wirkt.

Zusammengefasst, handelt es sich nach (2.8) bei $D \in \mathbb{C}^{n\times n}$ um eine dünnbesetzte Matrix, die auf einem vierdimensionalen Gitter mit $n_\mathcal{L} = N_t N_r^3$ Knoten operiert. Jeder Knoten beherbergt dabei 12 Variablen, wodurch sich $n = 12n_\mathcal{L}$ ergibt. Zusätzlich hängt die Matrix D von einer Linkvariablen-Konfiguration $\{U_\mu(x) : x \in \mathcal{L}, \mu = 1,2,3,4\}$, sowie vom Massenparameter m_0 und dem Sheikholeslami-Wohlert-Koeffizienten c_{sw} ab. In der Praxis ist der Massenparameter m_0 negativ und das Spektrum von D befindet sich in der rechten Halbebene, vgl. Abbildungen 2.4 und 2.5. Das Spektrum des „symmetrisierten“ Operators $Q := \Gamma_5 D$ ist reell und Q praktisch maximal indefinit (z. B. [41]), d. h., das reelle Spektrum von Q ist nahezu punktsymmetrisch zum Ursprung. Ab gewissen Gittergrößen ist eine explizite Formulierung bzw. Speicherung des Operators D als Matrix nicht mehr sinnvoll oder möglich, daher ist eine effiziente, parallel implementierte

Anwendung von D bei Gitter-QCD-Simulationen essentiell und Gegenstand der Forschung [53].

2.3 Präkonditionierung mit Schurkomplement

Eine weitverbreitete (statische) Prä- oder auch Vorkonditionierungstechnik der Gitter-QCD (z. B. [65], sog. *even-odd-preconditioning*) ist, statt mit dem gesamten System zu arbeiten, eine Umsortierung der Variablen vorzunehmen und nur auf der Hälfte der Knoten Berechnungen durchzuführen. Ähnlich zum Beweis von Lemma 2.2.13 (iii), führt eine Aufteilung der Gitterknoten in *gerade* Knoten zuerst und *ungerade* Knoten zuletzt zur Blockstruktur

$$D = \begin{bmatrix} D_{gg} & D_{gu} \\ D_{ug} & D_{uu} \end{bmatrix},$$

wobei ein Knoten $x = (x_1, x_2, x_3, x_4)$ als gerade bezeichnet wird, falls $x_1 + x_2 + x_3 + x_4$ gerade ist, andernfalls als ungerade. Die Inverse von D ist dann gegeben durch

$$D^{-1} = \begin{bmatrix} I & 0 \\ -D_{uu}^{-1} D_{ug} & I \end{bmatrix} \begin{bmatrix} D_S^{-1} & 0 \\ 0 & D_{uu}^{-1} \end{bmatrix} \begin{bmatrix} I & -D_{gu} D_{uu}^{-1} \\ 0 & I \end{bmatrix} \tag{2.11}$$

mit dem SCHUR[7]-Komplement

$$D_S := D_{gg} - D_{gu} D_{uu}^{-1} D_{ug}.$$

Die Blöcke D_{gg} und insbesondere D_{uu} sind nur auf der Diagonalen besetzt mit Blöcken der Größe 6×6, welche ausschließlich vom Clover-Term stammen. Das Lösen des Systems $D\psi = \eta$ für nur gerade Gitterknoten via SCHUR-Komplement liefert anschließend auch die Lösung auf den ungeraden Gitterknoten und damit die Lösung des gesamten Systems. Hierzu wird der Vektor auf der rechten Seite ebenfalls auf

$$\eta = \begin{bmatrix} \eta_g \\ \eta_u \end{bmatrix}$$

sortiert und wir lösen

$$\psi_g = D_S^{-1}(\eta_g - D_{gu} D_{uu}^{-1} \eta_u)$$

mit einem iterativem Löser für D_S, um danach über

$$\psi_u = D_{uu}^{-1} \eta_u - D_{uu}^{-1} D_{ug} \psi_g$$

die Lösung für die ungeraden Gitterknoten zu erhalten. Eine einmalige Vorberechnung von D_{uu}^{-1} wird algebraisch vollzogen und ist in der Berechnung nicht sehr teuer, daher benötigt eine Anwendung von D_S auf einen Vektor etwa denselben Aufwand wie mit D, während sich die Kondition von D_S typischerweise gegenüber der von D verbessert.

2.4 Normalität und Smearing

2.4.1 Definition

Ein linearer Operator A auf einem endlichdimensionalen Vektorraum heißt *normal*, wenn

$$A^H A = A A^H.$$

Äquivalent dazu ist A genau dann normal, wenn A ähnlich zu einer Diagonalmatrix bezüglich unitärer Transformation ist (z. B. [90]). Obige Definition 2.4.1 ist übertragbar auf stetige lineare Operatoren in unendlichdimensionalen Hilberträumen, wobei diese dann normal sind, wenn sie mit ihrer Adjungierten kommutieren. Der kontinuierliche DIRAC-Operator $\mathcal{D}$ ist (in geeigneten, hier nicht näher betrachteten Hilberträumen) schief-selbstadjungiert, also, im Gegensatz zu D (vgl. Bemerkung 2.2.8), normal. ◊

2.4.2 Definition

Für einen linearen Operator A auf einem endlichdimensionalen komplexen Vektorraum V ist der (numerische) *Wertebereich* von A definiert durch

$$\mathcal{F}(A) := \{v^H A v \,:\, v^H v = 1,\, v \in V\}.$$ ◊

Für normale Operatoren ist der Wertebereich die konvexe Hülle des Spektrums (nach z. B. [44]). Für numerische Gleichungssystemlöser, wie das *Generalized Minimal Residual*-Verfahren mit Restart $\big(\text{GMRES}(m)\big)$, ist bekannt, dass sie konvergieren, falls der Wertebereich den Ursprung nicht enthält [90]. Es hat also numerische Vorteile, wenn eine Diskretisierung D des DIRAC-Operators $\mathcal{D}$ möglichst normal ist und somit $\mathcal{F}(D)$ in der rechten Halbebene liegt (vgl. Abbildungen 2.4 und 2.5). Normalität von D tritt asymptotisch tatsächlich auf, wenn Diskretisierungseffekte bei zunehmender Gitterfeinheit und -größe ab- bzw. zunehmen. Um Normalität in der Diskretisierung D zu erreichen, gibt es eine Reihe sog. *Smearing*-Techniken wie „stout"- [72], APE- [2], HYP- [45] und HEX- [20] Smearing, *Wuppertal-Smearing* [42], sowie die *Destillation* [85]. Typischerweise ist Smearing ein iterativer Prozess, in dem Linkvariablen über ihre Nachbarn geglättet werden.

Die Abweichung zur Normalität von D_W, dem WILSON-DIRAC-Operator ohne Clover-Term, kann mit Hilfe der FROBENIUS[8]-Norm als Summe von Plaketten (vgl. Definition 2.2.9) dargestellt werden. Der folgende Abschnitt basiert auf [16] und [89] und soll zunächst die Berechnung der FROBENIUS-Norm zeigen, welche eng mit der WILSON-Wirkung verknüpft ist, um dann damit die Funktionsweise des „stout"-Smearings zu veranschaulichen.

Der Einfachheit halber vernachlässigen wir im Folgenden die Gitterweite, d. h., wir setzen $a = 1$ und betrachten das Gitter

$$\mathcal{L} = \{x = (x_1, x_2, x_3, x_4) \,:\, 1 \le x_1 \le N_t,\, 1 \le x_2, x_3, x_4 \le N_r\}.$$

Nun stehen die Abweichungen der Plaketten zur Identität und die Normalität von D_W in folgendem Zusammenhang.

2.4.3 Satz

Die FROBENIUS-Norm von $D_W^H D_W - D_W D_W^H$ genügt der Gleichung

$$\|D_W^H D_W - D_W D_W^H\|_F^2 \;=\; 16 \sum_x \sum_{\mu<\nu} \mathrm{Re}\left(\mathrm{Spur}(I_3 - Q_x^{\mu,\nu})\right), \tag{2.12}$$

wobei die erste Summe über alle Gitterknoten $x \in \mathcal{L}$ läuft und $\sum_{\mu<\nu}$ eine Abkürzung für $\sum_{\mu=1}^4 \sum_{\nu=\mu+1}^4$ ist.

Beweis. Es gilt die Einträge von $D_W^H D_W - D_W D_W^H$ zu untersuchen. Dazu benutzen wir die in Bemerkung 2.2.8 eingeführte Notation $\pi_\mu^\pm$ für die Matrizen

$$\pi_\mu^\pm \;=\; \frac{1}{2}(I_4 \pm \gamma_\mu), \quad \mu = 1,2,3,4.$$

Die Beziehungen (2.1) zwischen den γ-Matrizen implizieren, dass $\pi_\mu^\pm$ (kommutierende) Projektionen sind mit der zusätzlichen Eigenschaft

$$\pi_\mu^+ \pi_\mu^- \;=\; \pi_\mu^- \pi_\mu^+ \;=\; 0, \quad \mu = 1,2,3,4. \tag{2.13}$$

Tabelle 2.1 listet die Kopplungsterme an den Gitterknoten x und $x \pm \hat{\mu}$ von D_W bzw. D_W^H, welche in Matrixdarstellung Teilblöcke der Größe 12×12 von D_W bzw. D_W^H entsprechen (vgl. wieder mit Bemerkung 2.2.8). m steht hierbei für $m_0 + 4$ mit m_0 aus (2.4). Für das Produkt

	D_W	D_W^H
(x,x)	mI_{12}	mI_{12}
$(x, x+\hat{\mu})$	$-\pi_\mu^- \otimes U_\mu^H(x)$	$-\pi_\mu^+ \otimes U_\mu^H(x)$
$(x, x-\hat{\mu})$	$-\pi_\mu^+ \otimes U_\mu(x-\hat{\mu})$	$-\pi_\mu^- \otimes U_\mu(x-\hat{\mu})$

Tabelle 2.1: Kopplungsterme in D_W und D_W^H

$D_W^H D_W$ entstehen die in Tabelle 2.2 gelisteten Kopplungsterme aus Summen aller Pfade der Länge zwei auf dem vierdimensionalen Gitter sowie dem Produkt der entsprechenden Kopplungsterme in D_W^H und D_W. Die relevanten Pfade sind die folgenden (vgl. auch Abbildung 2.2):

- Für die Diagonalposition (x,x) existieren neun Pfade der Länge Zwei, $(x,x) \to (x,x) \to (x,x)$ und $(x,x) \to (x, x \pm \hat{\mu}) \to (x,x)$, mit $\mu = 1,2,3,4$.
- Für die nächsten Nachbarn $(x, x \pm \hat{\mu})$ haben wir jeweils zwei Pfade $(x,x) \to (x,x) \to (x, x \pm \hat{\mu})$ und $(x,x) \to (x, x \pm \hat{\mu}) \to (x, x \pm \hat{\mu})$.
- Für die Knoten $(x, x \pm 2\hat{\mu})$ gibt es lediglich den Pfad $(x,x) \to (x, x \pm \hat{\mu}) \to (x, x \pm 2\hat{\mu})$, für den aber das Produkt der Kopplungen wegen (2.13) Null ist.

- Schließlich gibt es für die restlichen vier übernächsten Nachbarn jeweils zwei Pfade, exemplarisch $(x,x) \to (x, x+\hat{\mu}) \to (x, x+\hat{\mu}-\hat{\nu})$ und $(x,x) \to (x, x-\hat{\nu}) \to (x, x+\hat{\mu}-\hat{\nu})$.

Die Kopplungsterme für $D_W D_W^H$ erhalten wir durch Vertauschen von π_μ^+ und π_μ^- sowie π_ν^+ und π_ν^-.

	$D_W^H D_W$
(x,x)	$m^2 I_{12} + \frac{1}{2}\sum_{\mu=1}^{4}(\pi_\mu^+ + \pi_\mu^-) \otimes I_3$
$(x, x+\hat{\mu})$	$-m(\pi_\mu^+ + \pi_\mu^-) \otimes U_\mu^H(x)$
$(x, x-\hat{\mu})$	$-m(\pi_\mu^+ + \pi_\mu^-) \otimes U_\mu^H(x-\hat{\mu})$
$(x, x \pm 2\hat{\mu})$	0
$\mu \neq \nu$:	
$(x, x+\hat{\mu}+\hat{\nu})$	$\pi_\mu^-\pi_\nu^+ \otimes U_\mu^H(x)U_\nu^H(x+\hat{\mu}) + \pi_\nu^-\pi_\mu^+ \otimes U_\nu^H(x)U_\mu^H(x+\hat{\nu})$
$(x, x+\hat{\mu}-\hat{\nu})$	$\pi_\mu^-\pi_\nu^- \otimes U_\mu^H(x)U_\nu(x+\hat{\mu}-\hat{\nu}) + \pi_\nu^+\pi_\mu^+ \otimes U_\nu(x-\hat{\nu})U_\mu^H(x-\hat{\nu})$
$(x, x-\hat{\mu}-\hat{\nu})$	$\pi_\mu^+\pi_\nu^- \otimes U_\mu(x-\hat{\mu})U_\nu(x-\hat{\mu}-\hat{\nu}) + \pi_\nu^+\pi_\mu^- \otimes U_\nu(x-\hat{\nu})U_\mu(x-\hat{\mu}-\hat{\nu})$

Tabelle 2.2: Kopplungsterme in $D_W^H D_W$. Kopplungen für $D_W D_W^H$ erhalten wir durch Vertauschen von π_μ^+ und π_μ^- sowie π_ν^+ und π_ν^-.

Damit ergibt sich für $D_W^H D_W - D_W D_W^H$, dass nur die Kopplungsterme an den Positionen $(x, x+\hat{\mu}+\hat{\nu})$, $(x, x+\hat{\mu}-\hat{\nu})$ und $(x, x-\hat{\mu}-\hat{\nu})$ für $\mu \neq \nu$ nicht wegfallen, vgl. Tabelle 2.3. Es

$\mu \neq \nu$	$D_W^H D_W - D_W D_W^H$
$(x, x+\hat{\mu}+\hat{\nu})$	$\frac{1}{2}(\gamma_\mu - \gamma_\nu) \otimes (I_3 - Q_x^{\mu,\nu})U_\nu^H(x)U_\mu^H(x+\hat{\nu})$
$(x, x+\hat{\mu}-\hat{\nu})$	$\frac{1}{2}(\gamma_\mu + \gamma_\nu) \otimes (I_3 - Q_x^{\mu,-\nu})U_\mu^H(x)U_\nu(x+\hat{\mu}-\hat{\nu})$
$(x, x-\hat{\mu}-\hat{\nu})$	$\frac{1}{2}(-\gamma_\mu + \gamma_\nu) \otimes (I_3 - Q_x^{-\mu,-\nu})U_\nu(x-\hat{\nu})U_\mu(x-\hat{\mu}-\hat{\nu})$

Tabelle 2.3: Nicht-verschwindende Kopplungsterme in $D_W^H D_W - D_W D_W^H$.

wurden dabei die Identitäten

$$\pm\pi_\mu^-\pi_\nu^- \mp \pi_\mu^+\pi_\nu^+ = \tfrac{1}{2}(\mp\gamma_\mu \mp \gamma_\nu) \quad \text{und}$$
$$\pm\pi_\mu^+\pi_\nu^- \mp \pi_\mu^-\pi_\nu^+ = \tfrac{1}{2}(\pm\gamma_\mu \mp \gamma_\nu)$$

verwendet, sowie die Plaketten aus Definition 2.2.9:

$$Q_x^{\mu,\nu} = U_\mu^H(x)U_\nu^H(x+\hat{\mu})U_\mu(x+\hat{\nu})U_\nu(x),$$
$$Q_x^{\mu,-\nu} = U_\nu(x-\hat{\nu})U_\mu^H(x-\hat{\nu})U_\nu^H(x+\hat{\mu}-\hat{\nu})U_\mu(x),$$
$$Q_x^{-\mu,-\nu} = U_\mu(x-\hat{\mu})U_\nu(x-\hat{\mu}-\hat{\nu})U_\mu^H(x-\hat{\mu}-\hat{\nu})U_\nu^H(x-\hat{\nu}).$$

Exemplarisch rechnen wir für den Eintrag an Position $(x, x+\hat{\mu}+\hat{\nu})$ in Tabelle 2.3 nach:

$$\begin{aligned}
&\pi_\mu^- \pi_\nu^+ \otimes U_\mu^H(x) U_\nu^H(x+\hat{\mu}) + \pi_\nu^- \pi_\mu^+ \otimes U_\nu^H(x) U_\mu^H(x+\hat{\nu}) \\
&- \big(\pi_\mu^+ \pi_\nu^- \otimes U_\mu^H(x) U_\nu^H(x+\hat{\mu}) + \pi_\nu^+ \pi_\mu^- \otimes U_\nu^H(x) U_\mu^H(x+\hat{\nu})\big) \\
&\quad = \tfrac{1}{2}(-\gamma_\mu + \gamma_\nu) \otimes U_\mu^H(x) U_\nu^H(x+\hat{\mu}) + \tfrac{1}{2}(\gamma_\mu - \gamma_\nu) \otimes U_\nu^H(x) U_\mu^H(x+\hat{\nu}) \\
&\quad = \tfrac{1}{2}(\gamma_\mu - \gamma_\nu) \otimes (I_3 - Q_x^{\mu,\nu}) U_\nu^H(x) U_\mu^H(x+\hat{\nu}).
\end{aligned}$$

Nun benutzen wir folgende generellen Eigenschaften der FROBENIUS-Norm

$$\begin{aligned}
\|AQ\|_F &= \|A\|_F \text{ für jede unitäre Matrix } Q \text{ (und solange } AQ \text{ definiert ist)}, \\
\|A \otimes B\|_F &= \|A\|_F \|B\|_F \text{ für alle Matrizen } A,\, B,
\end{aligned}$$

um das Quadrat der FROBENIUS-Norm der Kopplungen in Tabelle 2.3 darzustellen als**

$$\begin{aligned}
&2\|I_3 - Q_x^{\mu,\nu}\|_F^2 && \text{für Position } (x, x+\hat{\mu}+\hat{\nu}), \\
&2\|I_3 - Q_x^{\mu,-\nu}\|_F^2 && \text{für Position } (x, x+\hat{\mu}-\hat{\nu}), \\
&2\|I_3 - Q_x^{-\mu,-\nu}\|_F^2 && \text{für Position } (x, x-\hat{\mu}-\hat{\nu}).
\end{aligned}$$

Schließlich gilt für die FROBENIUS-Norm und unitäre Matrizen Q

$$\|I - Q\|_F^2 = \operatorname{Spur}\big((I-Q^H)(I-Q)\big) = 2\operatorname{Re}\big(\operatorname{Spur}(I-Q)\big),$$

womit wir das Quadrat der FROBENIUS-Norm $\|D_W^H D_W - D_W D_W^H\|_F^2$ durch Summieren über die Quadrate der FROBENIUS-Normen der einzelnen Kopplungen erhalten. Diese Summe erstreckt sich über insgesamt $24 n_{\mathcal{L}}$ Kopplungsmatrizen. Innerhalb dieser beziehen sich die Kopplungen dabei in Vierergruppen, bis auf Konjugation in $SU(3)$, auf dieselbe Plakette, d. h., $\operatorname{Spur}(I-Q)$ hat denselben Wert für alle vier Plaketten Q. Deshalb genügt es beispielsweise nur die Plakette $Q_x^{\mu,\nu}$ vierfach zu werten. Insgesamt ergibt sich

$$\|D_W^H D_W - D_W D_W^H\|_F^2 = 4 \sum_x \sum_{\mu<\nu} 2 \cdot 2 \cdot \operatorname{Re}\big(\operatorname{Spur}(I_3 - Q_x^{\mu,\nu})\big). \qquad \square$$

Der obige Satz zeigt also, dass der WILSON-DIRAC-Operator D_W normal ist, falls alle $Q_x^{\mu,\nu}$ gleich der Identität sind, d. h. alle Linkvariablen $U_\mu(x)$ sind gleich der Identität oder es gilt $U_\mu(x) \equiv U^H(x+\hat{\mu}) U(x)$ für gewisse $U(\,\cdot\,) \in SU(3)$ (vgl. (2.5)). Dies ist der Fall in der sog. *freien Theorie*. In physikalisch relevanten Konfigurationen ist D_W allerdings immer nicht-normal.

2.4.4 Definition

Für eine gegebene Linkvariablenkonfiguration $U = \{U_\mu(x)\}$ wird die Größe

$$S_W(U) := \sum_x \sum_{\mu<\nu} \operatorname{Re}\big(\operatorname{Spur}(I_3 - Q_x^{\mu,\nu})\big)$$

**Es gilt $\|(-1)^n \gamma_\mu + (-1)^m \gamma_\nu\|_F = \sqrt{8}, \quad \forall\, m, n,\ \mu \neq \nu$, vgl. (2.9).

als WILSON-*Eichfeldwirkung* bezeichnet[††]. ◊

Neben der numerischen Vorteile der Normalität ist Smearing in der Physik vor allem verbreitet, um sog. „cut-off"-Effekte zu reduzieren. Diese stehen eng in Verbindung mit lokalen Eigenvektoren, die zu sehr nahe an der Null liegenden Eigenwerten gehören.

Beim „stout"-Smearing [72] werden nun im Wesentlichen die Linkvariablen auf folgende Art modifiziert:

$$U_\mu(x) \rightsquigarrow \tilde{U}_\mu(x) := e^{\varepsilon Z_\mu^{\mathcal{U}}(x)} U_\mu(x), \tag{2.14}$$

wobei der Parameter ε positiv und hinreichend klein ist und

$$Z_\mu^{\mathcal{U}}(x) := -\frac{1}{2}\big(M_\mu(x) - M_\mu^H(x)\big) + \frac{1}{6}\,\mathrm{Spur}\,\big(M_\mu(x) - M_\mu^H(x)\big) I_3 \tag{2.15}$$

mit wiederum

$$M_\mu(x) := \sum_{\nu=1,\,\nu\neq\mu}^{4} \left(Q_x^{\mu,\nu} + Q_x^{\mu,-\nu}\right).$$

Insbesondere hängt $Z_\mu^{\mathcal{U}}(x)$ von den lokalen Plaketten um x ab.

Das folgende Resultat aus [63, 72] verbindet das „stout"-Smearing mit dem WILSON-Fluss $\mathcal{V}(\tau) := \{V_\mu(x,\tau) : x \in \mathcal{L}, \mu = 1,2,3,4\}$, definiert durch die Lösung des Anfangswertproblems

$$\frac{\partial}{\partial\tau} V_\mu(x,\tau) = -\{\partial S_W\big(\mathcal{V}(\tau)\big)\} V_\mu(x,\tau), \quad V_\mu(x,0) = U_\mu(x). \tag{2.16}$$

Insbesondere ist $V_\mu(x,\tau) \in SU(3)$ und ∂ der kanonische Differentialoperator bezüglich der Linkvariablen $V_\mu(x,\tau)$ mit Werten in $\mathfrak{su}(3)$.

2.4.5 Satz

Sei $\mathcal{V}(\tau)$ die Lösung des Anfangswertproblems (2.16). Dann gilt

(i) $\mathcal{V}(\tau)$ ist eindeutig für alle $\mathcal{V}(0)$ und alle $\tau \in \mathbb{R}$, sowie differenzierbar nach τ und $\mathcal{V}(0)$.

(ii) $S_W\big(\mathcal{V}(\tau)\big)$ ist als Funktion in τ monoton fallend.

(iii) Ein Schritt der LIE-EULER[9]-Integration mit Schrittweite ε von (2.16), gestartet bei $\tau = 0$, liefert eine Approximation $\widetilde{\mathcal{V}}(\varepsilon) = \{\widetilde{V}_\mu(x,\varepsilon)\}$ an $\mathcal{V}(\varepsilon)$ mit

$$\widetilde{V}_\mu(x,\varepsilon) = e^{\varepsilon Z_\mu^{\mathcal{U}}(x)} U_\mu(x),$$

wobei $Z_\mu^{\mathcal{U}}(x)$ aus (2.15) ist.

Wir verweisen zur Vertiefung dieses Themas auf [63, 72] sowie für Details zum Beweis von (i) und (ii) auf [12], wobei aus dieser Quelle erwähnt sei, dass die Lösung von (2.16) Eichkonfigurationen entlang der Richtung des steilsten Abstiegs im Raum der Eichkonfigurationen

[††] In unserem Kontext nicht wichtige Skalierungen sind nötig, um physikalische Relevanz zu erreichen, für Details siehe z. B. [110].

transportiert werden und somit S_W tatsächlich lokal minimiert wird. Teil (iii) folgt direkt aus einmaliger Anwendung des LIE-EULER-Schemas, vgl. [43].

Zusammengefasst impliziert der Satz, dass ein Schritt der LIE-EULER-Integration äquivalent zu einem Schritt „stout"-Smearing ist (vgl. (2.14)), wobei gleichzeitig die WILSON-Eichwirkung entlang der exakten Lösung von (2.16) minimiert wird. Wir können also, zumindest für hinreichend kleine ε, erwarten, dass auch die LIE-EULER-Approximation die WILSON-Eichwirkung ebenso minimiert und dadurch D_W letztendlich normalisiert wird.

Für den WILSON-DIRAC-Operator mit Clover-Term, d. h., für $D = D_W + C$ ergibt sich bezüglich der Normalität Folgendes:

$$\begin{aligned}\|D^H D - DD^H\|_F &= \|D_W^H D_W - D_W D_W^H + (D_W^H - D_W)C - C(D_W^H - D_W)\|_F \\ &\leq \|D_W^H D_W - D_W D_W^H\|_F + 2\|C\|_F \|D_W^H - D_W\|_F.\end{aligned}$$

Da alle Summanden von $\mathrm{Re}\left(\mathrm{Spur}(I_3 - Q_x^{\mu,\nu})\right)$ in (2.12) positiv sind (vgl. Beweis von 2.4.3), folgt aus $\|D_W^H D_W - D_W D_W^H\|_F^2 \to 0$ insbesondere, dass $Q_x^{\mu,\nu} \to I_3$ für alle x. Dies bedeutet, dass nach Definition 2.2.10 des Clover-Terms $Q_{\mu\nu}(x) - Q_{\nu\mu}(x) \to 0$ für alle x und μ, ν gilt und somit $\|C\|_F$ verschwindet. Demnach gilt

$$D_W \text{ wird normalisiert } \implies D \text{ wird normalisiert.}$$

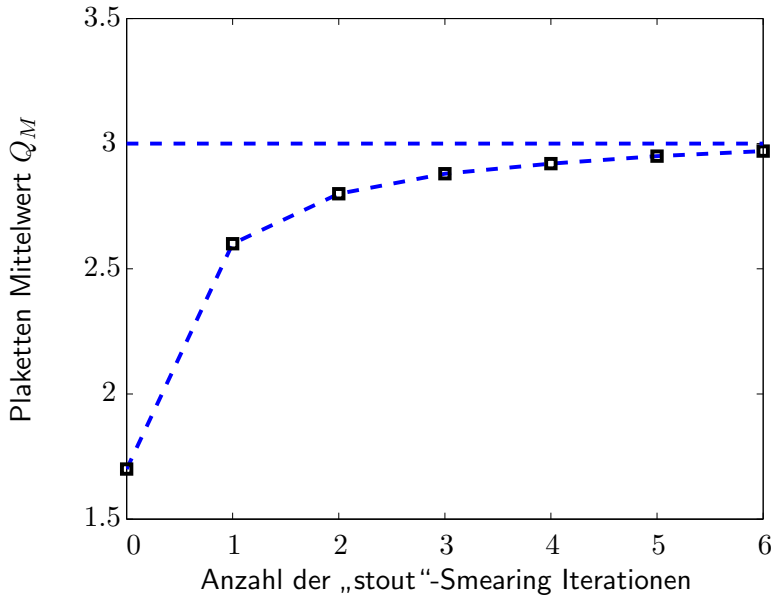


Abbildung 2.8: Effekt des „stout"-Smearings auf den Mittelwert der Plaketten[‡‡].

Um die Wirkung des „stout"-Smearings auf die Plaketten zu veranschaulichen, betrachten wir den Mittelwert über alle Plaketten

$$Q_M := \frac{1}{N_Q} \sum_x \sum_{\mu<\nu} \mathrm{Re}\left(\mathrm{Spur}(Q_x^{\mu,\nu})\right), \tag{2.17}$$

[‡‡] Daten aus [16].

wobei N_Q die Anzahl aller Plaketten bezeichnet. Damit vereinfacht sich (2.12) zu

$$\|D_W^H D_W - D_W D_W^H\|_F^2 = 16 N_Q (3 - Q_M).$$

Abbildung 2.8 zeigt, wie die WILSON-Wirkung in den ersten Iterationen des „stout“-Smearings rapide abnimmt.

Abschließend zu diesem Kapitel sei auf diverse Arbeiten verwiesen, die Beziehungen zwischen der Spektralstruktur und Werteverteilung in den Plaketten analysieren. In [75] wird beispielsweise gezeigt, dass der Abstand des Spektrums zum Ursprung damit zusammenhängt, ob $\mathrm{Re}\left(\mathrm{Spur}(I - Q_x^{\mu,\nu})\right)$ größer als eine gewisse Schranke ist, für alle Plaketten $Q_x^{\mu,\nu}$. Andere Arbeiten studieren den Zusammenhang zwischen Schwankungen der Plakettenwerte und räumlich-lokalen Eigenmoden nahe der Null [11, 74, 76], sowie den Einfluss von Smearing auf diese Eigenmoden [46].

Anmerkungen

[3] Marius Sophus Lie [li:] (* 17. Dezember 1842 in Nordfjordeid; † 18. Februar 1899 in Kristiania, heute Oslo) war ein norwegischer Mathematiker.

[4] William Kingdon Clifford (* 4. Mai 1845 in Exeter, Devon, England; † 3. März 1879 auf Madeira, Portugal) war ein britischer Philosoph und Mathematiker.

[5] Leopold Kronecker (* 7. Dezember 1823 in Liegnitz; † 29. Dezember 1891 in Berlin) war ein deutscher Mathematiker.

[6] Euklid von Alexandria war ein griechischer Mathematiker, der wahrscheinlich im 3. Jahrhundert v. Chr. in Alexandria gelebt hat.

[7] Issai Schur (* 10. Januar 1875 in Mogiljow (Weißrussland); † 10. Januar 1941 in Tel Aviv) war ein deutscher Mathematiker und Schüler von Frobenius.

[8] Ferdinand Georg Frobenius, genannt Georg, (* 26. Oktober 1849 in Berlin; † 3. August 1917 in Charlottenburg, heute ein Ortsteil von Berlin) war ein deutscher Mathematiker und Schüler von Karl Weierstraß und Ernst Eduard Kummer.

[9] Leonhard Euler (lateinisch Leonhardus Eulerus; * 15. April 1707 in Basel; † 7. September (jul.)/ 18. September 1783 (greg.) in Sankt Petersburg) war ein Schweizer Mathematiker und Physiker. Wegen seiner Beiträge zur Analysis, zur Zahlentheorie und zu vielen weiteren Teilgebieten der Mathematik gilt er als einer der bedeutendsten Mathematiker.

3. Krylov-Unterraumverfahren

In diesem Kapitel wollen wir uns einen Überblick im Gebiet der numerischen linearen Algebra verschaffen, insbesondere bezüglich der KRYLOV[10]-Unterraumverfahren zur Lösung von Gleichungssystemen und final zur Berechnung von Eigenpaaren (also Eigenwerten mit zugehörigen Eigenvektoren). Breiteren Überblick auf das Themengebiet liefern viele Lehrbücher, beispielsweise [44, 68].

Wir gehen von einem eindeutig lösbaren linearen Gleichungssystem

$$Ax = b, \qquad A \in \mathbb{C}^{n\times n},\, b \in \mathbb{C}^n,$$

aus, dessen Lösung $x = A^{-1}b$ unbekannt und gesucht ist. Aufbauen wollen wir auf einem fundamentalen Satz aus der linearen Algebra, dem Satz von CAYLEY[11]-HAMILTON[12]: Eine quadratische Matrix ist immer Nullstelle ihres charakteristischen Polynoms. Daraus leitete KRYLOV mit folgender simpler Überlegung

$$\begin{aligned} &\chi_A(A) = A^n + \alpha_{n-1}A^{n-1} + \cdots + A\alpha_1 + \alpha_0 = 0 && (\alpha_i \in \mathbb{C}, i = 0, \ldots, n-1) \\ \Longrightarrow\ &A^{-1} = \beta_{n-1}A^{n-1} + \beta_{n-2}A^{n-2} + \cdots + \beta_1 A + \beta_0 && (\beta_i \in \mathbb{C}, i = 0, \ldots, n-1) \end{aligned}$$

ab, dass die Inverse von A ein Polynom in A mit maximalem Grad $n-1$ ist. Insbesondere befindet sich die Lösung x in dem Raum

$$\mathcal{K}_m(A, b) = \mathrm{Spann}\{b, Ab, A^2b, \ldots, A^{m-1}b\},$$

für ein $m \leq n$, wobei Spann die lineare Hülle bezeichnet. Dieser Raum wurde erstmals 1931 im Artikel [54] von KRYLOV benannt. Verfahren, welche sich diesen Sachverhalt auf verschiedenste Arten zunutze machen, werden unter dem Begriff der KRYLOV-*Unterraumverfahren* zusammengefasst. Andere Autoren nennen in diesem Zusammenhang den Begriff *Projektionsverfahren*, da viele Verfahren Orthogonalbasen der (oder Derivate von) KRYLOV-Unterräume berechnen, um dann, mittels Projektion auf diese Unterräume, Gleichungssysteme zu lösen oder Eigenwerte zu berechnen.

KRYLOV-Unterraumverfahren basieren oft auf Umformulierung des linearen Gleichungssystems in ein Minimierungsproblem. Die zwei wohl bekanntesten Vertreter in diesem Bereich sind das Verfahren der *konjugierten Gradienten* (CG) von HESTENES[13] und STIEFEL[14] [47] aus dem Jahre 1952, sowie das 1986 von Saad und Schultz [91] entwickelte *Generalized minimal residual*-Verfahren (GMRES). Beide Verfahren bestimmen die optimale Approximation $x_m \in x_0 + \mathcal{K}_m(A, r_0)$ mit $r_0 = b - Ax_0$ an die Lösung $x = A^{-1}b$ mittels einer Orthogonalitäts- oder GALERKIN[15]-Bedingung, wobei in jeder Iteration die Dimension des KRYLOV-Unterraums

um Eins erhöht wird. Insbesondere würden beide Methoden bei exakter Arithmetik nach maximal n Schritten die richtige Lösung liefern.

Das CG-Verfahren, welches ausschließlich für symmetrische (bzw. hermitesche, im komplexen Fall) und positiv definite Matrizen funktioniert, minimiert das Funktional

$$\frac{1}{2}x^H Ax - x^H b$$

über einer Orthonormalbasis von $\mathcal{K}_m$. Der Orthogonalisierungsprozess ist dank der Symmetrie von A ohne Kenntnis aller vorherig berechneten Basisvektoren möglich, es handelt sich um ein Verfahren mit *kurzen Rekursionen*. Insbesondere müssen keine Zwischenergebnisse langfristig gespeichert werden, d. h., wachsende KRYLOV-Räume stellen kein Problem dar.

Nun ist klar, dass $A^H A$ für reguläres A immer symmetrisch und positiv definit ist und statt $Ax = b$ auch $A^H Ax = A^H b$ gelöst werden könnte, was aber ab bestimmten Systemgrößen, weder aus praktischer noch aus analytischer Sicht, zu empfehlen ist. Insbesondere quadriert sich dabei die Konditionszahl des Systems, die allgemein erheblichen Einfluss auf die Konvergenzgeschwindigkeit von KRYLOV-Unterraumverfahren hat.

Eine Möglichkeit die Konditionszahl des Systems zu verringern, besteht darin, möglichst „leicht"-invertierbare Matrizen P_L und $P_R \in \mathbb{C}^{n\times n}$ zu finden und statt der Ausgangsgleichung das folgende *präkonditionierte* System

$$\begin{aligned} P_L A P_R x^P &= P_L b, \\ x &= P_R x^P \end{aligned}$$

zu betrachten, mit dem Ziel, dass das Matrixprodukt $P_L A P_R$ eine möglichst gute Approximation an die Einheitsmatrix ist. Für das CG-Verfahren ist ein solch präkonditioniertes System aber problematisch, da $P_L A P_R$ in nur wenigen Fällen symmetrisch und positiv definit ist.

Ab bestimmten Konditionszahlen und/oder Systemgrößen werden KRYLOV-Unterraumverfahren selten ohne Präkonditionierung verwendet. Deshalb ist die Möglichkeit, auf bestimmte Probleme abgestimmte Präkonditionierungen flexibel verwenden zu können, sehr wichtig. Dies steht, bezogen auf unser Problem des Lösens der diskretisierten DIRAC-Gleichung, besonders im Vordergrund, weshalb wir uns im Folgenden auf das GMRES-Verfahren konzentrieren. Im Übrigen hat das zu GMRES symmetrisierte Verfahren *MINRES* von Paige und Saunders [84] aus dem Jahre 1975*, aus ähnlichen Gründen, wie oben beschrieben, in der QCD-Praxis wenig Relevanz, selbst wenn wir mit dem hermiteschen (aber indefiniten) Operator $Q := \Gamma_5 D$ arbeiten.

3.1 GMRES

Das „Generalized minimal residual"-Verfahren GMRES minimiert zunächst ähnlich zum CG-Verfahren das Funktional

$$F(x) := \|b - Ax\|_2^2$$

*Historisch gesehen ist das GMRES-Verfahren eine Verallgemeinerung des MINRES-Verfahren.

über einer mittels des ARNOLDI[16]-Verfahrens berechneten Orthogonalbasis des KRYLOV-Unterraums. Das ARNOLDI-Verfahren (siehe Algorithmus 1) arbeitet dabei ähnlich zum (numerisch weniger stabilen) GRAM-SCHMIDT-Orthogonalisierungs-Verfahren. Da in der Berechnung alle Vektoren v_j gebraucht werden, um das nächste v_m zu berechnen, handelt es sich hier um ein Verfahren mit *langen Rekursionen*.

Algorithmus 1: Arnoldi-Verfahren

Eingabe: Normierter Startvektor v_1, Krylov-Dimension m

Ausgabe: Orthonormalbasis V_m, Hessenbergmatrix H_m

for $i = 1, \ldots, m-1$ **do**
  $z \leftarrow Av_i$
  for $j = 1, \ldots, i$ **do**
    $h_{i,j} \leftarrow v_j^H z$
  end for
  $v_{i+1} \leftarrow z - \sum_{j=1}^{i} h_{j,i} v_j$
  $h_{i+1,i} \leftarrow \|v_{i+1}\|$ `/* Abbruch falls` $v_{i+1} = 0$ `*/`
  $v_{i+1} \leftarrow v_{i+1}/h_{i+1,i}$
end for
return $V_m = [v_1, \ldots, v_m],\ H_m = (h_{i,j})$

Vorausgesetzt, das ARNOLDI-Verfahren bricht nicht vor der Berechnung von $v_m \neq 0$ ab, stellen die Spalten von V_j eine Orthonormalbasis des j-ten KRYLOV-Unterraums $\mathcal{K}_j(A, v_1)$ für $j = 1, \ldots, m$ dar. Gilt ansonsten $v_{i+1} = 0$ mit $i+1 \leq m$, so ist $\mathcal{K}_{i+1}(A, v_1) = \mathcal{K}_i(A, v_1)$. D. h. $\mathcal{K}_i(A, v_1)$ (und somit auch V_i) enthält bereits alle Informationen.

Formal sind die Iterierten im GMRES-Verfahren gegeben durch das Minimierungsproblem

$$x_m = \underset{x \in x_0 + \mathcal{K}_m}{\operatorname{argmin}} F(x),$$

was, um wieder den Bogen hin zum KRYLOV-Unterraumverfahren zu spannen, äquivalent zur GALERKIN-Bedingung ist: Finde $x_m \in x_0 + \mathcal{K}_m(A, r_0)$ mit

$$b - Ax_m \perp \mathcal{L}_m := A\mathcal{K}_m(A, r_0).$$

Nachdem die Matrix $V_m \in \mathbb{C}^{n \times m}$ (mit orthogonalen Spalten) und die (obere) HESSENBERG[17]-Matrix $H_m \in \mathbb{C}^{m \times m}$ berechnet wurde, ist das weitere Vorgehen im GMRES-Verfahren wie folgt:

- Schreibe $x_m = x_0 + \sum_{j=1}^{m} \alpha_j v_j = x_0 + V_m \alpha$ mit $\alpha = [\alpha_1, \ldots, \alpha_m]^T \in \mathbb{C}^m$.
- Finde α, welches die Bedingung $J(\alpha) \leq J(a) := \|b - A(x_0 + V_m a)\|_2$ für alle $a \in \mathbb{C}^m$ erfüllt.

Unter Zuhilfenahme folgender Eigenschaften kann α explizit berechnet werden:

(i) $H_m = V_m A V_m$ mit $H_m = \begin{bmatrix} h_{11} & \dots & \dots & \dots & h_{1m} \\ h_{21} & \ddots & & & \vdots \\ 0 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & h_{m,m-1} & h_{mm} \end{bmatrix}$,

(ii) $AV_m = V_{m+1}\overline{H}_m$ mit $\overline{H}_m = \begin{bmatrix} H_m \\ 0 \dots 0\ h_{m+1,m} \end{bmatrix} \in \mathbb{C}^{(m+1)\times m}$.

Mit e_1, dem ersten Einheitsvektor in $\mathbb{C}^{m+1}$ ist folgende Gleichungskette wesentlich:

$$\begin{aligned} J(a) &= \|b - A(x_0 + V_m a)\|_2 \\ &\overset{r_0 = b - Ax_0}{=} \|r_0 - AV_m a\|_2 \\ &\overset{v_1 := r_0/\|r_0\|_2}{=} \left\| \|r_0\|_2 v_1 - AV_m a \right\|_2 \\ &\overset{\text{(ii)}}{=} \left\| V_{m+1}(\|r_0\|_2 e_1 - \overline{H}_m a) \right\|_2 \\ &= \left\| \|r_0\|_2 e_1 - \overline{H}_m a \right\|_2, \end{aligned}$$

wobei im letzten Schritt die Spaltenorthonormalität von V_{m+1} ausgenutzt wurde. Um das Minimum der letzten Zeile explizit zu berechnen, wird eine QR-Zerlegung[†] der erweiterten HESSENBERG-Matrix $\overline{H}_m$ berechnet, was in modernen GMRES-Implementierungen mittels GIVENS[18]-Rotationen realisiert wird. Diese Rotationen sind unitäre Matrizen G_i, welche im Wesentlichen Diagonalmatrizen mit einem 2×2-Block auf der Diagonalen sind. Je eine dieser Rotationen wird explizit in jeder GMRES-Iteration berechnet. $Q_m := G_m \cdots G_1 \in \mathbb{C}^{(m+1)\times(m+1)}$ ist dann ebenfalls eine unitäre Matrix, für die

$$Q_m \overline{H}_m = \overline{R}_m \tag{3.1}$$

mit

$$\overline{R}_m = \begin{bmatrix} \bar{r}_{11} & \dots & \dots & \bar{r}_{1m} \\ 0 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \bar{r}_{mm} \\ 0 & \dots & \dots & 0 \end{bmatrix} =: \begin{bmatrix} R_m \\ 0 \dots 0 \end{bmatrix} \in \mathbb{C}^{(m+1)\times m}$$

gilt, wobei R_m regulär ist[‡]. Unter nochmaliger Verwendung von $e_1 = [1, 0, \dots, 0]^T \in \mathbb{C}^{m+1}$ definieren wir den (Fehlergrößen-)Vektor

$$\overline{g}_m := \|r_0\|_2 Q_m e_1 =: [\gamma_1^{(m)}, \dots, \gamma_m^{(m)}, \gamma_{m+1}]^T =: [g_m^T, \gamma_{m+1}]^T \in \mathbb{C}^{m+1}$$

[†] $H = Q^H R$ mit unitärer Matrix Q (d. h. $Q^H = Q^{-1}$) und R einer rechten oberen Dreiecksmatrix.

[‡] Beweis per Induktion über m, vorausgesetzt das ARNOLDI-Verfahren terminiert nicht vorzeitig.

und es folgt

$$\begin{aligned}
\min_{a\in\mathbb{C}^m} J(a) &= \min_{a\in\mathbb{C}^m} \left\| \|r_0\|_2 e_1 - \overline{H}_m a \right\|_2 \\
&= \min_{a\in\mathbb{C}^m} \left\| Q_m(\|r_0\|_2 e_1 - \overline{H}_m a) \right\|_2 \\
&\overset{(3.1)}{=} \min_{a\in\mathbb{C}^m} \|\overline{g}_m - \overline{R}_m a\|_2 \\
&= \min_{a\in\mathbb{C}^m} \sqrt{|\gamma_{m+1}|^2 + \|g_m - R_m a\|_2^2} \\
&\geq |\gamma_{m+1}| = J(R_m^{-1} g_m).
\end{aligned}$$

Insbesondere erhalten wir das Minimum α durch einfaches Rückwärtseinsetzen der Gleichung $R_m\alpha = g_m$, und die Norm des Residuum $r_m := b - Ax_m$ der aktuellen Iterierten ist gegeben durch

$$\|r_m\|_2 = |\gamma_{m+1}|.$$

Algorithmus 2: GMRES-Verfahren (vereinfachte Darstellung)

Eingabe: Startvektor x_0, Iterationsanzahl $m_{\max}$

Ausgabe: Näherunglösung x_m

1 **for** $m = 1, \ldots, m_{\max}$ **do**
2 `Berechne` V_m `und` H_m `mit dem Arnoldi-Verfahren`
3 `Berechne QR-Zerlegung von` $\overline{H}_m$
4 $\alpha \leftarrow R_m^{-1} g_m$ `/* Rückwärtseinsetzen */`
5 **end for**
6 **return** $x_m = x_0 + V_m\alpha$

Falls im Arnoldi-Verfahren $v_{i+1} = 0$ auftritt und der Algorithmus vorzeitig abbricht, so terminiert auch das GMRES-Verfahren. Es bricht aber in diesem Fall nicht etwa zusammen (wie diverse andere Bi-CG-Verfahren), sondern liefert die exakte Lösung.

Darüber hinaus haben Saad und Schultz erkannt, dass der Rechenaufwand je Iteration (nur) linear mit m wächst, wenn die Rechenschritte in Zeile 2 und 3 geeignet in die Schleife des Arnoldi-Verfahrens integriert werden.

Da alle Berechneten Vektoren v_i gespeichert werden müssen und das Arnoldi-Verfahren mit wachsendem m immer aufwändiger wird, wird das GMRES-Verfahren in der Praxis nur in einer Variante mit *Neustarts* genutzt, d. h, die Größe des Krylov-Raum wird limitiert. Genauer startet das Verfahren nach $m_{\max}$-Iterationen mit der aktuellen Näherung $x_{m_{\max}}$ als neuen Startwert x_0 erneut. Es werden dann solange Neustarts vollzogen bis eine gewünschte Toleranz erreicht ist. Falls das Spektrum von A in der rechten Halbebene liegt und exakte Arithmetik vorliegt, ist die Konvergenz selbst für $m_{\max} = 2$ zwar noch garantiert (allerdings nicht mehr zwingend nach m Iterationen), in der Praxis kommt es aber bei zu kleinen Krylov-Räumen zur Stagnation im Residuenverlauf. Wir verwenden bei den im Verlauf der Arbeit vorgestellten Ergebnissen Krylov-Unterräume mit höchstens Dimension 25. Ebenso verwenden wir sog. „deflated“-Neustarts, vgl.

[70]. Hierbei wird bei einem Neustart nicht der gesamte KRYLOV-Unterraum verworfen, sondern es werden nach dem RAYLEIGH[19]-RITZ[20]-Prinzip [93] Informationen über das System beibehalten. Hierfür werden einige wenige Eigenwerte der (kleinen) Matrix

$$B := V_{m_{\max}}^H A V_{m_{\max}}$$

berechnet und die zugehörigen Eigenvekoren in den neuen KRYLOV-Unterraum übernommen.

3.2 Flexibles GMRES

Bekanntlich hat die Kondition des zu lösenden Systems erheblichen Einfluss auf die Konvergenz von KRYLOV-Unterraumverfahren. Deshalb ist es überaus sinnvoll, die Kondition des Systems durch Anwenden von (links-)Präkonditionierung (wie zu Beginn des Kapitels gesehen)

$$Ax = b \quad \Leftrightarrow \quad MAx = Mb$$

zu reduzieren. Für dieses Vorgehen ist das GMRES-Verfahren besonders gut geeignet, da es möglich ist, den KRYLOV-Unterraum derart zu modifizieren, dass jede Iteration mit einem eigenen Präkonditionierer M_j versehen werden kann. Das Verfahren arbeitet dann mit dem modifizierten KRYLOV-Unterraum

$$\widetilde{\mathcal{K}}_m(A, r_0) = \mathrm{Spann}\{r_0, M_1 A r_0, M_2 A M_1 A r_0, \ldots, M_{m-1} A M_{m-2} A \cdots M_2 A M_1 A r_0\}.$$

Hierbei kann M_j insbesondere selbst wieder aus einem iterativen Verfahren resultieren. Dieses Verfahren ist als *flexibles* GMRES-Verfahren (FMGRES) bekannt [92, 71]. Der (algorithmisch) einzige Nachteil gegenüber statischem Präkonditionieren ist der doppelte Speicheraufwand, da neben den v_j auch die präkonditionierten Vektoren $M_j v_j$ gespeichert werden müssen. Dies stellt aber bei relativ kleinen KRYLOV-Unterraum-Dimensionen kein größeres Problem dar. Die außerordentliche Robustheit (d. h. die numerische Stabilität) von FMGRES qualifiziert es, als äußerer Löser in Kombination mit den, im Verlauf der Arbeit vorgestellten, algebraischen Mehrgitterverfahren, zu wirken.

Anmerkungen[§]

[10] Arthur Cayley (* 16. August 1821 in Richmond upon Thames, Surrey; † 26. Januar 1895 in Cambridge) war ein englischer Mathematiker. Er befasste sich mit sehr vielen Gebieten der Mathematik von der Analysis, Algebra, Geometrie bis zur Astronomie und Mechanik, ist aber vor allem für seine Rolle bei der Einführung des abstrakten Gruppenkonzepts bekannt.

[11] Sir William Rowan Hamilton (* 4. August 1805 in Dublin; † 2. September 1865 in Dunsink bei Dublin) war ein irischer Mathematiker und Physiker, der vor allem für seine Beiträge zur Mechanik und für seine Einführung und Untersuchung der Quaternionen bekannt ist.

[12] Alexei Nikolajewitsch Krylow (* 3. (jul.)/ 15. August 1863 (greg.) in Wisjaga, Gouvernement Simbirsk (heute Oblast Uljanowsk); † 26. Oktober 1945 in Leningrad (heute St. Petersburg), Sowjetunion) war ein russischer Schiffbau-Ingenieur und Mathematiker.

[13] Magnus Hestenes (* 1906 in Bricelyn, Minnesota; † 31. Mai 1991) war ein US-amerikanischer Mathematiker.

[14] Eduard Ludwig Stiefel (* 21. April 1909 in Zürich; † 25. November 1978 ebenda) war ein Schweizer Mathematiker.

[15] Boris Grigorjewitsch Galjorkin (wiss. Transliteration Boris Grigor'evič Galërkin, häufig als Galerkin transkribiert; * 20. Februar (jul.)/ 4. März 1871 (greg.) in Polozk, heute Weißrussland; † 12. Juli 1945 in Leningrad) war ein sowjetischer Ingenieur und Mathematiker.

[16] Walter Edwin Arnoldi (* 14. Dezember 1917 in New York City; † 5. Oktober 1995) war ein US-amerikanischer Maschinenbau-Ingenieur, bekannt für eine Arbeit zur numerischen linearen Algebra.

[17] Karl Adolf Hessenberg (* 8. September 1904 in Frankfurt am Main; † 22. Februar 1959 ebenda) war ein deutscher Elektrotechnik-Ingenieur und Mathematiker.

[18] James Wallace Givens, Jr. (* 14. Dezember 1910 in Alberene bei Charlottesville; † 5. März 1993) war Mathematiker und Pionier der Informatik.

[19] John William Strutt, 3. Baron Rayleigh (* 12. November 1842 in Langford Grove, Maldon, England; † 30. Juni 1919 in Terlins Place bei Witham, England), war ein englischer Physiker. Er erhielt 1904 den Nobelpreis für Physik.

[20] Walter Ritz (oder Walther Ritz, * 22. Februar 1878 in Sion (Sitten); † 7. Juli 1909 in Göttingen) war ein Schweizer Mathematiker und Physiker. Er war ein bedeutender Schweizer Wissenschaftler und Forscher, obwohl er nach einer kurzen Karriere bereits mit 31 Jahren starb.

[§] Alle Angaben aus der deutschen Wikipedia, stand 2017

4. Gebietszerlegungsmethoden

Wie in Kapitel 3 betrachten wir das Raumzeitgitter $\mathcal{L}$ mit normierter Gitterweite $a = 1$

$$\mathcal{L} = \{x = (x_1, x_2, x_3, x_4) \,:\, 1 \leq x_1 \leq N_t,\, 1 \leq x_2, x_3, x_4 \leq N_r\}.$$

Gebietszerlegungsmethoden sind in der Gitter-QCD weit verbreitet und als zuverlässige Präkonditionierung bekannt, siehe z. B. [60, 65, 64]. Wir wollen uns im Folgenden die Grundidee der Blockzerlegung des Gitters anschauen und die SCHWARZ[21]schen Algorithmen einführen. Dieses Kapitel orientiert sich an [35], siehe auch [89, 99].

4.1 Blockzerlegung des Gitters

4.1.1 Definition

Es sei $\{\mathcal{T}_1^1, \ldots, \mathcal{T}_{l_1}^1\}$ eine Partition von $\{1, \ldots, N_t\}$ in l_1 Blöcke zusammenhängender Punkte in der ersten Dimension, d. h. auf der Zeitachse, gilt:

$$\mathcal{T}_j^1 := \{t_{j-1} + 1, \ldots, t_j\}, \qquad j = 1, \ldots l_1, \quad 0 = t_0 < t_1 \cdots < t_{l_1} = N_t.$$

Analog zerlegen wir die Raumdimensionen in Blöcke $\{\mathcal{T}_1^\mu, \ldots, \mathcal{T}_{l_\mu}^\mu\}$, $\mu = 2, 3, 4$. Eine *Blockzerlegung* des gesamten Gitters $\mathcal{L}$ in $l = l_1 l_2 l_3 l_4$ Gitterblöcke $\mathcal{L}_i$ hat nun die Form

$$\mathcal{L}_i = \mathcal{T}_{j_1(i)}^1 \times \mathcal{T}_{j_2(i)}^2 \times \mathcal{T}_{j_3(i)}^3 \times \mathcal{T}_{j_4(i)}^4.$$

Darüber hinaus können nun alle $12n_\mathcal{L}$ Variablen aus $\mathcal{V} = \mathcal{L} \times \mathcal{C} \times \mathcal{S}$ in l Variablenblöcke $\mathcal{V}_i$ zerlegt werden, indem wir alle zugehörigen Spin- und Farbkomponenten des Gitterblocks $\mathcal{L}_i$ hinzunehmen:

$$\mathcal{V}_i = \mathcal{L}_i \times \mathcal{C} \times \mathcal{S}. \tag{4.1}$$

Eine *Vergröberung* des Gitters liegt vor, wenn für eine weitere Blockzerlegung $\{\mathcal{L}_i' \,:\, i = 1, \ldots, l'\}$ gilt: für jedes $\mathcal{L}_i'$ existiert ein $\mathcal{L}_j$ mit

$$\mathcal{L}_i' \subset \mathcal{L}_j. \qquad \Diamond$$

Zu den größten Herausforderungen im Bereich der Gitter-QCD gehört das wiederholte Lösen der diskretisierten DIRAC-Gleichung (siehe Abschnitt 2.2):

$$D\psi = \eta. \tag{4.2}$$

Die Systeme, die in Gitter-QCD-relevanten Berechnungen auftreten, haben hunderte von Millionen Unbekannte und sind daher in Computersimulationen nur mittels *Parallelisierung* in realistischer

Rechenzeit durchführbar. Hier spielt die Blockzerlegung eine entscheidende Rolle; jeder Prozess behandelt dabei einen Variablenblock $\mathcal{V}_i$ der Blockzerlegung des zugrundeliegenden Gitters und führt dort Berechnungen durch. Danach teilt er die relevanten Ergebnisse den Nachbarprozessen mit, denken wir hier beispielsweise an eine Matrix-Vektor-Operation, d. h., die Anwendung des Operators D auf einen Vektor ψ. Dieses Konzept ist auch erweiterbar auf das Invertieren des Operators. Effiziente Kommunikation zwischen den Prozessen vorausgesetzt, sind Gebietszerlegungsmethoden auf natürliche Weise kompatibel mit Parallelisierungsarchitekturen im Computer.

4.2 Additive und multiplikative Alternierende Verfahren von Schwarz

4.2.1 Definition

Sei $\mathcal{V}_i \subset \mathcal{V}$ ein Variablenblock. Wir definieren die *triviale Einbettung*

$$\mathcal{I}_{\mathcal{V}_i} : \mathcal{V}_i \to \mathcal{V}$$

als die Restriktion der Identität von $\mathcal{V}$ auf $\mathcal{V}_i$, d. h.,

$$I_{\mathcal{V}_i} := (\mathrm{id}_{\mathcal{V}})|_{\mathcal{V}_i}.$$

Hiermit sind die korrespondierenden *Block-Inversen* formal definiert durch

$$B_i := \mathcal{I}_{\mathcal{V}_i} D_i^{-1} \mathcal{I}_{\mathcal{V}_i}^H \quad \text{mit} \quad D_i := \mathcal{I}_{\mathcal{V}_i}^H D \mathcal{I}_{\mathcal{V}_i}. \qquad \Diamond$$

4.2.2 Lemma

Wir betrachten die Iteration

$$\psi^{(k+1)} = H\psi^{(k)} + L\eta \quad \text{mit} \quad H, L \in \mathbb{C}^{n\times n} \text{ und } \psi^{(k)}, \eta \in \mathbb{C}^n,\ k \in \mathbb{N}_0.$$

(i) Ist ψ^* ein Fixpunkt der Iteration, d. h., ψ^* erfüllt $\psi^* = H\psi^* + L\eta$, dann gilt für die Fehler $e^{(k)} := \psi^* - \psi^{(k)}$:

$$e^{(k+1)} = He^{(k)}.$$

Die Matrix H nennen wir den *Fehlerpropagator*.

(ii) Mit dem Startvektor $\psi^{(0)} = 0$ ist die k-te Iterierte gegeben durch

$$\psi^{(k)} = \sum_{i=0}^{k-1} H^i L\eta.$$

Beweis. (i) Es gilt

$$e^{(k+1)} = \psi^* - \psi^{(k)} = (H\psi^* + L\eta) - (H\psi^{(k)} + L\eta) = H(\psi^* - \psi^{(k)}) = He^{(k)}.$$

(ii) Es ist $\psi^{(1)} = H\psi^{(0)} + L\eta = H^0 L\eta$ und damit induktiv

$$\psi^{(k+1)} = H\left(\sum_{i=0}^{k-1} H^i L\eta\right) + L\eta = \sum_{i=1}^{k} H^i L\eta + H^0 L\eta = \sum_{i=0}^{k} H^i L\eta. \qquad \square$$

In unserem Kontext besteht eine Invertierung von D, bzw. das Lösen des Systems (4.2), aus dem Lösen aller Blocksysteme

$$D_i e_i = \mathcal{I}_{\mathcal{V}_i}^H r, \tag{4.3}$$

mit dem Residuum des Ausgangsystems $r = \eta - D\psi$, und korrigieren von

$$\psi \leftarrow \psi + B_i r \quad \text{mit } B_i r = \mathcal{I}_{\mathcal{V}_i} e_i \text{ für } i = 1, \ldots, l. \tag{4.4}$$

Das Residuum kann nach Bedarf zwischen den Iterationen via

$$r \leftarrow \eta - D\psi \tag{4.5}$$

aktualisiert werden.

Im Falle, dass die Residuenberechnung (4.5) nur *einmal* vor dem Lösen aller Blocksysteme durchgeführt wird, werden alle Iterationen (4.4) mit demselben Residuum r abgearbeitet und eine Iteration der Gebietszerlegungsmethode ist durch

$$\psi \leftarrow \psi + M(\eta - D\psi) = (I - MD)\psi + M\eta$$

gegeben, wobei die Abkürzung

$$M = \sum_{i=1}^{l} B_i$$

verwendet wurde. Der Fehlerpropagator ist dann (im Sinne von Lemma 4.2.2) gegeben durch

$$H = I - MD = I - \sum_{i=1}^{l} B_i D.$$

Andernfalls, wenn wir das Residuum (4.5) in *jeder* Iteration aktualisieren, hat der Fehlerpropagator die Form

$$H = \prod_{i=1}^{l} (I - B_i D).$$

Diese einfachsten Anwendungen von Gebietszerlegungsmethoden wurden im Rahmen der analytischen Theorie von partiellen Differentialgleichungen bereits im Jahre 1870 von SCHWARZ [94, 95] vorgeschlagen. Sie sind heutzutage unter den Namen *additive-* oder *multiplikative Alternierende Verfahren* von SCHWARZ (siehe Algorithmus 3 und 4) bekannt. Für eine tiefer gehende Auseinandersetzung mit diesem Themenkomplex verweisen wir exemplarisch auf [99].

Algorithmus 3: Additives Alternierendes Verfahren

Eingabe: ψ, η

Ausgabe: ψ

$r \leftarrow \eta - D\psi$

for $i = 1, \ldots, l$ **do**

  $\psi \leftarrow \psi + B_i r$

end for

Offenbar ist für Algorithmus 3 eine sehr einfache Parallelisierung möglich, da alle Blocksysteme unabhängig voneinander und gleichzeitig gelöst werden können.

Ganz anders in Algorithmus 4, welcher inhärent sequentiell ist, d. h., jeder folgende Schleifendurchlauf benötigt Information aus dem vorhergehenden Durchlauf.

Algorithmus 4: Multiplikatives Alternierendes Verfahren

Eingabe: ψ, η

Ausgabe: ψ

for $i = 1, \ldots, l$ **do**

  $r \leftarrow \eta - D\psi$

  $\psi \leftarrow \psi + B_i r$

end for

4.3 Rot-Schwarz-Ordnung und das multiplikative Alternierende Verfahren

Der Publikation [35] folgend, welche sich in diesem Teil auf [60] bezieht, stellen wir die multiplikative Alternierende Verfahren für die Rot-Schwarz-Ordnung (siehe Kapitel 2.3) vor, die im weiteren kurz SAP (*Schwarz Alternating Procedure*, vgl. [99]) genannt wird. Der Sinn ist, die ausschließlichen Nächste-Nachbar-Beziehungen des Wilson-Dirac-Operators auszunutzen. Hierfür werden die benachbarten Variablen Schachbrettartig abwechselnd mit zwei unterschiedlichen Farben koloriert. Algorithmus 5 zur Lösung von (4.2) fasst das Vorgehen zusammen.

Eine Iteration (d. h., $\nu = 1$) von Algorithmus 5 kann zusammengefasst notiert werden durch

$$\psi \leftarrow (I - KD)\psi + K\eta$$

mit der Abkürzung $B_{\mathsf{Farbe}} = \sum_{i \in \mathsf{Farbe}} B_i$ und

$$K = B_{\mathsf{schwarz}}(I - DB_{\mathsf{rot}}) + B_{\mathsf{rot}}.$$

Algorithmus 5: Rot-Schwarz Multiplikatives Alternierendes Verfahren (SAP)

Eingabe: ψ, η, ν
Ausgabe: ψ
for $k = 1, \ldots, \nu$ **do**
  $r \leftarrow \eta - D\psi$
  for all $i \in$ rot **do**
    $\psi \leftarrow \psi + B_i r$
  end for
  $r \leftarrow \eta - D\psi$
  for all $i \in$ schwarz **do**
    $\psi \leftarrow \psi + B_i r$
  end for
end for

Nach Lemma 4.2.2 ist der Fehlerpropagator der SAP-Methode gegeben durch

$$E_{\mathsf{SAP}} = I - KD = (I - B_{\mathsf{schwarz}} D)(I - B_{\mathsf{rot}} D)$$

und für den Startwert $\psi = 0$ erhalten wir nach ν Iterationen

$$M_{\mathsf{SAP}}^{(\nu)} \eta = \sum_{i=0}^{\nu-1} (I - KD)^i K \eta.$$

Offenbar ist $K = M_{\mathsf{SAP}} := M_{\mathsf{SAP}}^{(1)}$ und somit $E_{\mathsf{SAP}} = I - M_{\mathsf{SAP}} D$.

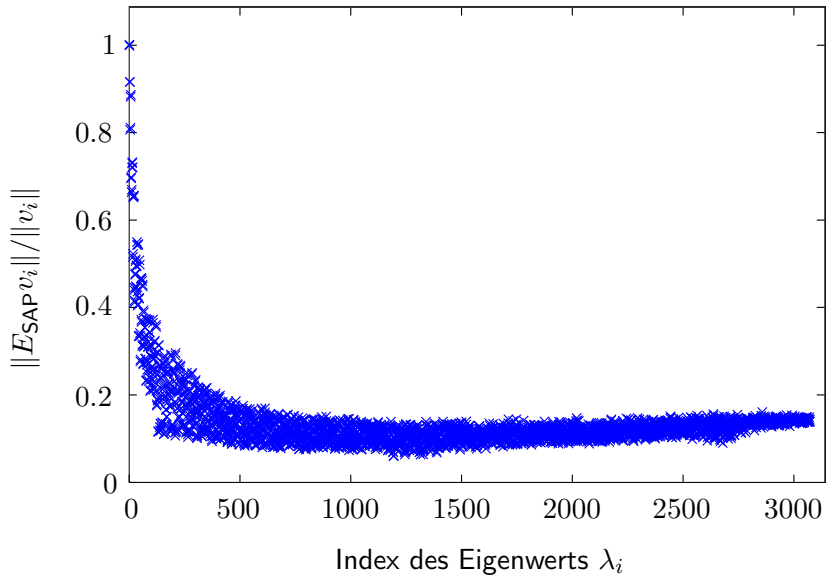


Abbildung 4.1: Fehlerreduktion von SAP bezüglich der Eigenmoden auf einem 4^4-Gitter mit Blöcken der Größe 2^4.

Die Lösungen der lokalen Blocksysteme (4.3), welche benötigt werden um $B_i r$ zu berechnen, sollten approximativ mit wenigen Iterationen eines KRYLOV-Unterraumverfahren bestimmt werden. Dies hat zur Folge, dass die SAP-Methode zu ein nicht-stationären iterativen Prozess wird. Wenn also diese nicht-stationäre SAP-Methode als Präkonditionierer fungieren soll, dann müssen flexible KRYLOV-Unterraumverfahren verwendet werden, vgl. Abschnitt 3.2 sowie [38, 60, 90].

Untersucht man die SAP-Methode als Präkonditionierer genauer, stellt man fest, dass sie nicht in der Lage ist, die Defizite, welche KRYLOV-Unterraumverfahren bei steigender Gittergröße oder abnehmenden Quarkmassen zeigen, auszugleichen. Dies lässt sich darauf zurückführen, dass die SAP-Methode Fehler, korrespondierend zu großen Eigenmoden, sehr gut reduziert, aber Fehler zu eher kleinen Eigenmoden praktisch unberührt lässt. Abbildung 4.1 illustriert den Umstand, wobei die horizontale Achse die Eigenvektoren v von D in aufsteigender Reihenfolge des Betrags des zugehörigen Eigenwertes repräsentiert. Die vertikale Achse beschreibt den Quotienten $\|E_{\mathsf{SAP}} v\|/\|v\|$. Dieser ist klein für die meisten (größeren) Eigenwerte, wird aber signifikant größer für betragsmäßig kleine Eigenwerte. Im Kontext von Glättern für Mehrgitterverfahren ist dies ein typisches Verhalten, was die Motivation begründet, SAP als Glätter für unser algebraisches Mehrgitterverfahren zu verwenden.

Anmerkungen*

[21] Hermann Amandus Schwarz (* 25. Januar 1843 in Hermsdorf, Schlesien; † 30. November 1921 in Berlin) war ein deutscher Mathematiker in Berlin. Unter Einfluss von Karl Weierstraß promovierte er 1864 bei Ernst Eduard Kummer.

*Alle Angaben aus der deutschen Wikipedia, stand 2017

5. Algebraische Mehrgitterverfahren

Mehrgitterverfahren bestehen immer aus zwei wesentlichen Komponenten: einem Glätter und einer Grobgitterkorrektur. Der Glätter ist meist ein Relaxierungsschema wie die JACOBI- oder GAUSS-SEIDEL-Verfahren sowie ihre Blockvarianten, welche äquivalent zu den in Kapitel 4 vorgestellten additiven und multiplikativen Alternierenden Verfahren von SCHWARZ sind. Ebenso kann auch ein KRYLOV-Unterraumverfahren verwendet werden. Vorerst fixieren wir als Glätter die im vorherigen Kapitel vorgestellte SAP-Methode.

Dieses Kapitel orientiert sich weiter an [35] (bzw. [89]) und konzentriert sich auf die Grobgitterkorrektur unseres Mehrgitterverfahrens.

Die Grobgitterkorrektur hat die Aufgabe, Fehlerkomponenten (auf einem gröberen Gitter, mit wenigen n_c Variablen) zu reduzieren, welche der Glätter schlecht oder gar nicht reduziert. Im Falle der Wahl von SAP als Glätter sind dies die Fehler, die zu betragsmäßig kleinen Eigenwerten korrespondieren. D. h., im Fokus stehen sog. *Nah-Kern*-Vektoren, also Eigenmoden, die von Eigenvektoren, zu betragsmäßig kleinen Eigenwerten, aufgespannt werden. Mit anderen Worten arbeitet die Grobgitterkorrektur mit einem Operator D_c, der D auf einem Unterraum repräsentiert und sowohl besonders im Nah-Kern-Bereich eine gute Approximation an D darstellt sowie gleichzeitig Eigenschaften wie Dünnbesetztheit und im besten Falle auch weitere Eigenschaften des Operators D (wie z. B. die Γ_5-Symmetrie) erhält. Letzteres ist besonders wichtig, wenn ein rekursiver Ansatz verfolgt wird, der es ermöglicht, nicht nur Zweigitter-, sondern auch echte Mehrgitterverfahren (mit mehr als 2 Gitterebenen) zu verwenden.

Um D_c zu konstruieren sind zwei wichtige Operatoren nötig.

5.0.1 Definition

Mit $n = 12n_{\mathcal{L}}$, $n_c < n$ und den Restriktions- und Prolongationsoperatoren*

$$R : \mathbb{C}^n \rightarrow \mathbb{C}^{n_c},$$
$$P : \mathbb{C}^{n_c} \rightarrow \mathbb{C}^n$$

definieren wir eine PETROV[22]-GALERKIN-Projektion von D bzw. den *Grobgitteroperator*

$$D_c := RDP$$

sowie die korrespondierende *Grobgitterkorrektur*

$$\psi \leftarrow \psi + PD_c^{-1}Rr$$

*D. h., R ist surjektiv und P ist injektiv. Konkreteres folgt in Kapitel 5.3.

mit dem Residuum $r = \eta - D\psi$. $\Diamond$

Die Abbildung R restringiert Informationen des ursprünglichen Raumes in einen „gröberen“ Unterraum und P transportiert (meist durch Interpolation) Informationen zurück in den ursprünglichen Raum. Die Grobgitterkorrektur zur aktuellen Iterierten ψ restringiert das aktuelle Residuum über R in den Unterraum, um dort

$$D_c e_c = Rr \tag{5.1}$$

zu lösen. Der *Grobgitterfehler* e_c wird dann via P zurück zum ursprünglichen Raum transportiert, um die Grobgitterkorrektur zu vollziehen. Eine Iteration der Grobgitterkorrektur kann zusammengefasst werden als

$$\psi \leftarrow (I - PD_c^{-1}RD)\psi + PD_c^{-1}R\eta.$$

Der zugehörige Fehlerpropagator ist gegeben durch

$$I - PD_c^{-1}RD.$$

Vorausgesetzt D_c ist bekannt, besteht ein Zweigitterverfahren aus wechselnder Anwendung des Glätters und der Grobgitterkorrektur:

Algorithmus 6: Zweigitterverfahren (V-Zykel mit Nachglättung)

Eingabe: ψ, η, ν

Ausgabe: ψ

1 $r \leftarrow \eta - D\psi$

2 $\psi \leftarrow \psi + PD_c^{-1}Rr$

3 $r \leftarrow \eta - D\psi$

4 $\psi \leftarrow \psi + M_{\mathsf{SAP}}^{(\nu)} r$

Algorithmus 6 zeigt die Vorgehensweise eines Zweigitterverfahrens mit ν-Schritten Nachglättung, in der Literatur auch V-Zykel genannt. Vorglätten ist ebenso möglich, resultiert aber wegen der Spektralgleichung

$$\sigma\big((I - M_{\mathsf{SAP}}D)(I - PD_c^{-1}RD)\big) = \sigma\big((I - PD_c^{-1}RD)(I - M_{\mathsf{SAP}}D)\big)$$

in keinem Vorteil (aber auch keinem Nachteil).

Dieses Zweigitterverfahren kann ebenso als Präkonditionierer für ein flexibles GMRES-Verfahren verwendet werden, wie die SAP-Methode (siehe Ende des vorherigen Kapitels).

Das Update der aktuellen Iterierten ψ ist in obigem Fall ein *multiplikatives* Update, da die Grobgitterkorrektur und die Anwendung des Glätters nacheinander mit dem jeweils aktuellsten Residuum $r = \eta - D\psi$ vollzogen wird. Wird andererseits auf Zeile 3 in Algorithmus 6 verzichtet und beide Schritte mit demselben Residuum durchgeführt, so spricht man von einem *additiven* Update. Eine Iteration des Verfahrens ist dann gegeben durch

$$\psi \leftarrow \psi + \Big(PD_c^{-1}R + M_{\mathsf{SAP}}^{(\nu)}\Big)(\eta - D\psi).$$

Diese additive Vorgehensweise erlaubt es, Glätter und Grobgitterkorrektur nebeneinander parallel ablaufen zu lassen, allerdings führt dies zu Effizienzeinbrüchen im FGMRES-Verfahren. Für mehr Details verweisen wir auf [99].

Algorithmus 6 können wir rekursiv zu einem echten Mehrgitterverfahren umwandeln, indem wir in Zeile 2 wiederum ein Zweigitterverfahren desselben Typs solange aufrufen, bis die Gleichung (5.1) aufgrund der kleinen Größe des Systems direkt gelöst werden kann.

Damit das gesamte Verfahren effizient ist, muss das Lösen von (5.1) wesentlich günstiger im Rechenaufwand sein, als die Originalgleichung $D\psi = \eta$. Insbesondere heißt das, dass D_c dünnbesetzt sein sollte. Um dies zu gewährleisten, müssen die Matrizen P und R, neben der Fähigkeit Links- und Rechtseigenvektoren von D möglichst gut zu approximieren, ebenfalls dünnbesetzt sein.

5.1 Aggregat-basierte Interpolation

Wir betrachten eine Blockzerlegung $\{\mathcal{L}_i \ : \ i = 1, \ldots, l\}$ des Gitters $\mathcal{L}$ (vgl. Definition 4.1.1). In einer Arbeit von Lüscher [61] beobachtete er, dass Eigenvektoren von D, die zu betragsmäßig kleinen Eigenwerten gehören, dazu tendieren, auf einer großen Anzahl von Gitterblöcken $\mathcal{L}_i$ nahezu konstant zu sein. Dieses Phänomen nannte er *lokale Kohärenz* (engl. *local coherence*). Lokale Kohärenz bedeutet insbesondere, dass viele Eigenvektoren zu kleinen Eigenmoden durch einige wenige dieser Vektoren darstellbar sind, indem sie über verschiedene Gitterblöcke zerlegt werden. Für eine tiefergehende quantitative Analyse der Beobachtung siehe [61]. Lokale Kohärenz ist der Kerngedanke hinter Aggregat-basierten Transferoperatoren in allgemeineren Problemstellungen, siehe z. B. [13, 18] und speziell für Gitter-QCD-Anwendungen [4, 17, 81].

Ähnlich zur Blockzerlegung definieren wir die Aggregate folgendermaßen.

5.1.1 Definition

Eine *Aggregation* $\{\mathcal{A}_1, \ldots, \mathcal{A}_s\}$ ist eine Partition der Variablenmenge $\mathcal{V} = \mathcal{L} \times \mathcal{C} \times \mathcal{S}$ (vgl. Definition 4.1.1). Wir bezeichnen sie als *Gitterblock*-Aggregation, falls jedes *Aggregat* $\mathcal{A}_i$ von der Form

$$\mathcal{A}_i := \mathcal{L}_{j(i)} \times \mathcal{W}_i$$

ist, wobei $\mathcal{L}_{j(i)}$ ein Gitterblock einer Blockzerlegung des Gitters $\mathcal{L}$ sowie $\mathcal{W}_i \subseteq \mathcal{C} \times \mathcal{S}$ ist. $\Diamond$

Aggregate für den Wilson-Dirac-Operator (2.7) werden typischerweise als Gitterblock-Aggregate realisiert. Im Unterschied zu einer „puren“ Blockzerlegung $\{\mathcal{V}_i \ : \ i = 1, \ldots, l\}$ des Gitters (wie z. B. bei der SAP-Methode) müssen Aggregate nicht alle Spin- und Farbvariablen enthalten. Darüber hinaus können mehrere Aggregate denselben Gitterblock $\mathcal{L}_i$ enthalten. Insbesondere müssen der SAP-Glätter und die Aggregat-basierten Transferoperatoren nicht auf derselben Blockzerlegung $\mathcal{L}$ basieren.

5.1.2 Definition

Wir betrachten eine gewisse Anzahl von *Testvektoren* $v_1, \ldots, v_N \in \mathbb{C}^n$ (welche möglichst den Nah-Kern von D repräsentieren, $N \ll n$) und eine Aggregation $\{\mathcal{A}_1, \ldots, \mathcal{A}_s\}$ ($s \ll n$). Der Aggregat-basierte Prolongations- oder *Interpolations*operator P ist dann anschaulich definiert durch das Zerlegen der Testvektoren über die verschiedenen Aggregate:

$$(v_1|\ldots|v_N) \quad = \quad \begin{bmatrix} \vdots & & \vdots \\ \vdots & & \vdots \\ \vdots & & \vdots \end{bmatrix}_{n\times N} \quad \longrightarrow \quad P \quad = \quad \begin{bmatrix} \vdots & & \\ & \ddots & \\ & & \vdots \end{bmatrix}_{n\times(N\cdot s)} \begin{matrix} \mathcal{A}_1 \\ \mathcal{A}_2 \\ \vdots \\ \mathcal{A}_s \end{matrix}. \tag{5.2}$$

Formal induziert jedes Aggregat $\mathcal{A}_i$ N Variablen mit den Indizes $(i-1)N+1, \ldots, iN$ in das grobe Gitter und wir definieren zeilenweise

$$Pe_{(i-1)N+j} := \mathcal{I}_{\mathcal{A}_i}\mathcal{I}_{\mathcal{A}_i}^H v_j, \quad \text{für} \quad i = 1, \ldots, s,\ j = 1, \ldots, N, \tag{5.3}$$

wobei $e_{(i-1)N+j}$ den $\big((i-1)N+j\big)$-ten Einheitsvektor in $\mathbb{C}^{N\cdot s}$ bezeichnet. $\Diamond$

Es wurden hierfür die trivialen Einbettungen aus Definition 4.2.1 verwendet, d. h., $\mathcal{I}_{\mathcal{A}_i}\mathcal{I}_{\mathcal{A}_i}^H v_j$ lässt alle Komponenten von v_j, welche zu $\mathcal{A}_i$ gehören, unverändert und setzt alle anderen auf Null (M. a. W., $\mathcal{I}_{\mathcal{A}_i}\mathcal{I}_{\mathcal{A}_i}^H$ ist eine Orthogonalprojektion auf $\mathcal{A}_i$). Aus Stabilitätsgründen werden die Testvektoren lokal orthonormalisiert, d. h., für jedes i ersetzen wir $\mathcal{I}_{\mathcal{A}_i}^H v_j$ in (5.3) durch den j-ten Basisvektor der Orthonormalbasis von $\mathrm{Spann}\{\mathcal{I}_{\mathcal{A}_i}^H v_1, \ldots, \mathcal{I}_{\mathcal{A}_i}^H v_N\}$. Dies ändert weder das Bild von P noch den Grobgitteroperator $I - P(RDP)^{-1}RD$, garantiert aber, dass $P^H P = I$.

Die Restriktion R wird analog zu P konstruiert: eine Menge von Testvektoren $\{\hat{v}_1, \ldots \hat{v}_N\}$ muss gewählt werden, die Aggregate von P können wieder verwendet werden.

Abbildung 5.1 zeigt eine auf Gitterblöcken basierte Aggregation bezüglich eines Raumzeit-Gitterknotens (reduziert auf zwei Dimensionen). Hierbei wurde für jedes Aggregat $\mathcal{A}_i$ als $\mathcal{W}_i$ die gesamte Variablenmenge $\mathcal{C} \times \mathcal{S}$ verwendet. Anschaulich formt die Aggregation ein neues, gröberes Gitter, wobei die Dünnbesetztheit und Kopplungsstruktur von $D_c = RDP$ die von D widerspiegeln, d. h., insbesondere haben wir wieder nur nächste-Nachbar-Beziehungen der einzelnen Gitterknoten. Jeder Gitterpunkt des groben Gitters (bzw. des Aggregats) fasst N Variablen zusammen.

5.2 Galerkin und Petrov-Galerkin Ansätze

Um die Struktur- und Spektraleigenschaften (siehe Lemma 2.2.13) des Wilson-Dirac-Operators D auf das grobe Gitter zu übertragen, bedarf es einer expliziten Abhängigkeit zwischen der Restriktion R und der Interpolation P. Die folgende Konstruktion von P (und damit auch von R)

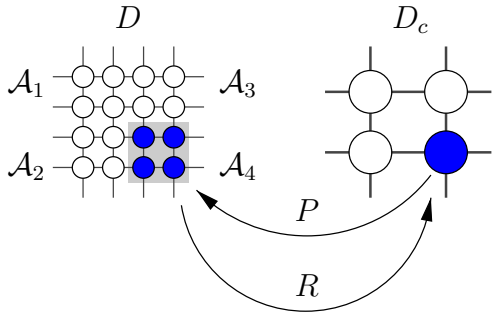


Abbildung 5.1: Aggregat-basierte Interpolation (Ansicht auf zwei Dimensionen reduziert).

ist ähnlich zu den Konstruktionen in [17, 81, 61, 4] mit einem Unterschied im Bereich der zu wählenden Testvektoren.

Der PETROV-GALERKIN Ansatz ist

$$R := (\Gamma_5 P)^H.$$

Wenn hierbei P aus den Testvektoren $v_1, \ldots, v_N$ konstruiert wurde, die die Rechtseigenvektoren zu betragsmäßig kleinen Eigenwerten von D approximieren, so liegen $R = (\Gamma_5 P)^H$ die Testvektoren $\hat{v}_i = \Gamma_5 v_i$ zugrunde, welche die Linkseigenvektoren zu kleinen Eigenwerten approximieren.

In [61] wird unter anderem festgestellt, dass es möglich ist, $R = P^H$ zu erreichen, solange Aggregate gewisse Regeln für Spin-Variablen einhalten, was auf folgende Definition führt.

5.2.1 Definition

Eine Aggregation $\{\mathcal{A}_i \ : \ i = 1, \ldots, s\}$ heißt Γ_5-*kompatibel*, falls jedes Aggregat $\mathcal{A}_i$ Raumzeit-Variablen ausschließlich entweder mit Spin-Variablen mit Index Null und Eins des feinen Gitters oder Spin-Variablen mit Index Zwei und Drei des feinen Gitters gruppiert. $\lozenge$

Angenommen, wir haben eine Γ_5-kompatible Aggregation und betrachten den Interpolationsoperator (5.2). Dann sehen wir, da Γ_5 trivial auf die Spins mit Index Null und Eins, und als Vorzeichenwechsel auf die Spins Zwei und Drei wirkt (vgl. (2.10) bzw. Lemma 2.2.7), dass beim Übergang von P zu $\Gamma_5 P$ jeder Block eines speziellen Aggregats entweder mit $+1$ oder -1 multipliziert wird. Mit anderen Worten

$$\Gamma_5 P = P \Gamma_5^c,$$

wobei Γ_5^c alle Spin-Null und -Eins Aggregate unverändert lässt und die Spin-Zwei und -Drei Aggregate mit -1 multipliziert.

5.2.2 Lemma

Gegeben sei eine Γ_5-kompatible Aggregation und P eine Aggregat-basierte Interpolation wie in (5.2) mit $R = (\Gamma_5 P)^H$. Wir unterscheiden die beiden Grobgitteroperatoren

$$D_c^{\mathsf{PG}} = RDP \qquad \text{und} \qquad D_c = P^H D P.$$

Dann gelten

(i) $D_c = \Gamma_5^c D_c^{\mathsf{PG}}$.

(ii) $I - PD_c^{-1}P^H D = I - P(D_c^{\mathsf{PG}})^{-1}RD$.

(iii) D_c^{PG} ist hermitesch, D_c ist Γ_5^c-symmetrisch.

(iv) Für den Wertebereich (siehe Definition 2.4.2) gilt $\mathcal{F}(D_c) \subseteq \mathcal{F}(D)$.

Beweis. Zunächst ist klar, dass Γ_5^c genau wie Γ_5 eine Diagonalmatrix mit Einträgen $+1$ oder -1 ist, d. h., $\Gamma_5^c = (\Gamma_5^c)^H = (\Gamma_5^c)^{-1}$. Teil (i) folgt nun aus

$$D_c^{\mathsf{PG}} = RDP = (\Gamma_5 P)^H DP = (P\Gamma_5^c)^H DP = \Gamma_5^c P^H DP = \Gamma_5^c D_c.$$

Daraus folgt ebenfalls unmittelbar

$$P(D_c^{\mathsf{PG}})^{-1}RD = P(\Gamma_5^c D_c)^{-1}(\Gamma_5 P)^H D = PD_c^{-1}\Gamma_5^c P^H \Gamma_5 D = PD_c^{-1}\Gamma_5^c\Gamma_5^c P^H D = PD_c^{-1}P^H D,$$

was Behauptung (ii) zeigt. Für Teil (iii) gilt wegen $D^H\Gamma_5 = \Gamma_5 D$ (vgl. Lemma 2.2.13):

$$(D_c^{\mathsf{PG}})^H = P^H D^H R^H = P^H D^H \Gamma_5 P = P^H \Gamma_5 DP = RDP = D_c^{\mathsf{PG}}.$$

Also ist D_c^{PG} hermitesch, was äquivalent dazu ist, dass $D_c = \Gamma_5^c D_c^{\mathsf{PG}}$ eine Γ_5^c-Symmetrie aufweist. Schließlich gilt, da P eine Isometrie ist (vgl. Definition 5.1.2 ff.), d. h., $P^H P = I$:

$$\begin{aligned}\mathcal{F}(D_c) = \{\psi_c^H D_c \psi_c \ : \ \psi_c^H \psi_c = 1\} &= \{(P\psi_c)^H D(P\psi_c) \ : \ (P\psi_c)^H(P\psi_c) = 1\} \\ &\subseteq \{\psi^H D\psi \ : \ \psi^H\psi = 1\} = \mathcal{F}(D).\end{aligned}$$

Dies zeigt Teil (iv) der Behauptung. □

Lemma 5.2.2 hat einige tiefgreifende Konsequenzen. Zunächst besagt Teil (ii), dass unabhängig davon, ob wir den Petrov-Galerkin-Ansatz D_c^{PG} mit $R = (\Gamma_5 P)^H$, oder den Galerkin-Ansatz D_c mit $R = P^H$ wählen, bei derselben Grobgitterkorrektur landen. Letzterer Ansatz erhält die Γ_5-Symmetrie von D auf dem gröberen Gitter und somit die Symmetrie des Spektrums (siehe nochmals Lemma 2.2.13). Falls $\mathcal{F}(D)$ in der rechten Halbebene liegt (wovon i. d. R. ausgegangen wird), befindet sich nach (iv) auch $\mathcal{F}(D_c)$ in der rechten Halbebene und somit auch das Spektrum von D_c. Betrachten wir den „symmetrisierten“ Wilson-Dirac-Operator $Q := \Gamma_5 D$, so wissen wir aus Kapitel 2, dass dieser annähernd maximal indefinit ist und wünschenswerter Weise ergeben sich ähnliche Beobachtungen bei numerischer Untersuchung des Operators $\Gamma_5^c D_c = D_c^{\mathsf{PG}}$.

Die Γ_5-Symmetrie impliziert eine weitere bemerkenswerte Eigenschaft, wenn es um Eigensysteme von Q und die Singulärwertzerlegung von D geht:

5.2.3 Proposition

Sei eine Eigendekomposition

$$Q = V\Lambda V^H, \quad \Lambda \text{ diagonal}, V^H V = I,$$

des hermiteschen Operators $Q = \Gamma_5 D$ gegeben. Dann ist

$$D = \big(\Gamma_5 V \operatorname{Sign}(\Lambda)\big)|\Lambda|V^H = U\Sigma V^H \tag{5.4}$$

eine Singulärwertzerlegung von D mit unitärem $U := \Gamma_5 V \operatorname{Sign}(\Lambda)$ und der Diagonalmatrix $\Sigma := |\Lambda|$. ◊

In Publikationen zum Thema algebraische Mehrgitterverfahren wie [18] wird vorgeschlagen, Restriktion und Interpolation mittels Links- und Rechts*singulär*vektoren (von zugehörigen kleinen Singulärwerten) zu konstruieren, anstelle von Eigenvektoren, um dann die obige Relation auszunutzen. Allerdings sind gute Approximationen an jene Singulärvektoren in der Praxis bei unserem Problem Q schwerer zu berechnen als Eigenvektoren zum Problem D. Dies liegt (vermutlich) letztendlich an der Spektralstruktur der Operatoren, da es beim Operator Q schwieriger ist an kleine Eigenwerte heran zu kommen, weil diese im Zentrum des Spektrums liegen. Im Falle von D, wo die kleinen Eigenwerte am Rand des Spektrums liegen und darüber hinaus in der rechten Halbebene $\mathbb{C}^+$ liegen (falls $\mathcal{F}(D) \subset \mathbb{C}^+$), ist deren Beschaffung einfacher. Verschiedene numerische Tests ergaben keinen Mehrwert im Verfolgen des Singulärwert-Ansatzes, sodass wir auf ein auf Eigenvektoren basierendes Mehrgitterverfahren setzen, was auch motiviert, D_c gegenüber D_c^{PG} als „korrekten“ Grobgitteroperator anzusehen, insbesondere um ein echtes Mehrgitterverfahren rekursiv auf D_c (mit identischen Attributen wie D) anwenden zu können.

Um dies möglichst universell machen zu können, verwenden wir eine spezielle Γ_5-kompatible Gitterblock-Aggregation:

5.2.4 Definition

Sei eine Blockzerlegung $\{\mathcal{L}_i \ : \ i = 1, \ldots, n_{\mathcal{L}_c}\}$ des Gitters $\mathcal{L}$ gegeben. Dann ist die *Standard-Aggregation* $\{\mathcal{A}_{i,\tau} \ : \ i = 1, \ldots, n_{\mathcal{L}_c}, \tau = 0, 1\}$ bezüglich dieser Blockzerlegung gegeben durch

$$\mathcal{A}_{i,0} := \mathcal{L}_i \times \{0, 1\} \times \mathcal{C} \quad \text{und} \quad \mathcal{A}_{i,1} := \mathcal{L}_i \times \{2, 3\} \times \mathcal{C}.$$ ◊

Diese Standard-Aggregation kombiniert immer zwei Spin-Freiheitsgrade in Γ_5-kompatibler Weise (vgl. Definition 5.2.1) mit allen drei Farb-Freiheitsgraden. Zu jedem i sind die Aggregate $\mathcal{A}_{i,0}$ und $\mathcal{A}_{i,1}$ die einzigen beiden Aggregate, die mit dem Gitterblock $\mathcal{L}_i$ assoziiert sind. Die Standard-Aggregation induziert hier ein grobes Gitter $\mathcal{L}_c$ mit $n_{\mathcal{L}_c}$ Punkten, wobei jeder Grobgitterpunkt zu einem Gitterblock $\mathcal{L}_i$ korrespondiert und $2N$ Variablen umfasst, mit N der Anzahl

der Testvektoren. Jeweils N Variablen gehören zu den Spin-Indizes Null und Eins (bzw. dem Aggregat $\mathcal{A}_{i,0}$) und weitere N zu den Spin-Indizes Zwei und Drei (bzw. dem Aggregat $\mathcal{A}_{i,1}$). Daraus ergibt sich eine Gesamtgröße des Grobgittersystems von $n_c = 2Nn_{\mathcal{L}_c}$. Standard-Aggregation und Konsequenzen aus Lemma 5.2.2, d. h., $D_c = P^H DP$, erhalten die Eigenschaften der Nächste-Nachbar-Kopplung, Γ_5-Kompatibilität, Spektraleigenschaften sowie Dünnbesetztheit.

Nun wenden wir uns den Testvektoren zu.

5.3 Adaptive Testvektorberechnung

Solange keine *a priori* Informationen über den Nah-Kern vorhanden ist, werden die Testvektoren $v_1, \ldots, v_N$ für unser auf Aggregation basierendes Mehrgitterverfahren in einer sog. *Setup-Phase* berechnet. Weiter [35] folgend, welches sich hierbei stark auf [18] bezieht, gehen wir folgendermaßen vor. Wir legen uns auf den GALERKIN-Ansatz fest, d. h.,

$$R = P^H.$$

Die fundamentale Idee von Mehrgitterverfahren ist es, Fehlerkomponenten zu finden, die vom Glätter nicht effektiv reduziert werden können, in unserem Kontext also den Nah-Kern. Demnach liefert ein Glätter nach einigen wenigen Iterationsschritten, angewendet auf die homogene Testgleichung

$$Du = 0$$

mit einem zufälligen Startvektor u, eine Näherung $\tilde{v}$ mit hohen Fehleranteilen in Eigenmoden, welche vom Glätter schlecht reduziert werden können. Nun könnten wir immer weitere Zufallsvektoren generieren und die homogene Gleichung lösen bis die gewünschte Anzahl von Testvektoren erreicht ist. Die nach (5.3) konstruierte Interpolation garantiert dann, dass auf dem groben Gitter genau diese Eigenmoden betont werden. Diese (vorläufige) Konstruktion von D_c kann dann verwendet werden, um die homogene Gleichung mittels eines Mehrgitterverfahrens zu lösen, was ein neueres, verbessertes Set an Testvektoren produziert mit hohen Fehleranteilen, welche vom Mehrgitterverfahren schlecht reduziert werden können. Damit wird wiederum ein verbesserter Grobgitteroperator D_c konstruiert, der sich immer spezieller auf jene Eigenmoden konzentriert. Das Iterieren dieses Prozesses führt ultimativ zu einem schnell konvergierenden Verfahren, allerdings mit womöglich unverhältnismäßig großem Aufwand, abhängig davon, wie oft der Prozess iterieren und wie viele Testvektoren verwendet werden sollen.

Den Aufwand, immer das gesamte Mehrgitterverfahren auf die homogene Gleichung anzuwenden, um eventuelle Mängel aufzuzeigen und diese dann zu beheben, gilt es im Zaum zu halten. Wir wollen uns eine Methode anschauen, die ihren Ursprung in den Arbeiten [14, 15] hat – den sog. „Bootstrap“-Ansatz. Der Ansatz beruht auf der folgenden fundamentalen Beobachtung.

5.3.1 Lemma

Gegeben ein Eigenpaar (λ_c, v_c) des Eigenwertproblems auf dem groben Gitter

$$D_c v_c = \lambda_c v_c,$$

so löst das Paar (λ_c, Pv_c) das bedingte Eigenwertproblem

$$\text{finde } (\lambda, v) \text{ mit } v \in \mathrm{Bild}(P) \text{ so, dass } P^H(Dv - \lambda v) = 0$$

auf dem feinen Gitter.

Beweis.

$$P^H(DPv_c - \lambda_c P v_c) = D_c v_c - \lambda_c P^H P v_c = 0,$$

wobei wegen $P^H P = I$ die Behauptung folgt. □

Es ist natürlich einfacher, Eigenvektoren zu betragsmäßig kleinen Eigenwerten auf dem groben Gitter zu berechnen, als auf dem feinen, und die Information des Grobgitteroperators wird so optimal genutzt. Einige Iterationen des Glätters angewandt auf Pv_c liefern dann brauchbare Testvektoren mit hohen Anteilen im Nah-Kern-Bereich des feinen Gitters. Der Ansatz in Lüschers Arbeit [61], d. h., der „inexakt deflation"-Ansatz, beruht ebenfalls auf dieser Idee, wobei das grobe Eigenwertproblem dort approximativ mit relativ grober Fehlertoleranz gelöst wird.

5.4 DD-αAMG

Wir haben nun alle Zutaten zur Beschreibung des auf Gebietszerlegung und Aggregation beruhenden adaptiven (die adaptive Komponente des Verfahrens liegt in der Generierung der Testvektoren, der Setup-Phase) algebraischen Mehrgitterverfahrens *DD-αAMG*[†] [35] zum Lösen der diskretisierten DIRAC-Gleichung mittels des WILSON-DIRAC-Operators mit Clover-Term (2.7).

Der Glätter ist standardmäßig $M_{\mathrm{SAP}}^{(\nu)}$, in späteren Anwendungen aber auch ein mit SCHUR-Komplement (siehe Abschnitt 2.3) präkonditioniertes FGMRES mit Neustarts (wie in der Implementierung des Verfahrens in der QOPQDP Software-Bibliothek, [82]).

Das verwendete Grobgittersystem ist $D_c = P^H D P$, wobei P der auf Aggregation basierende Interpolationsoperator ist, welcher in einer adaptiven Setup-Phase generiert wird. Algorithmus 7 gibt eine Übersicht des Vorgehens in Pseudocode. Die Konstruktionsphase in Zeile 6 wird, wie in Definition 5.1.2 beschrieben, vollzogen, inklusive lokaler Orthonormalisierung. Der Operator $C^{(\nu)}$ in Zeile 8 ist ein Platzhalter für *entweder* eine iterative Methode, um ein Eigenpaar (λ_c, Pv_c) des Grobgitteroperators D_c zu generieren und den „gelifteten" Vektor Pv_c mit ν Schritten zu glätten, *oder* aber einen V-Zykel des gesamten Lösers (ebenfalls mit ν Glättungsschritten). Bei letzterem gibt es wieder eine Wahl zwischen Anwendung des Lösers auf die homogene Gleichung

[†] Engl.: *Domain Decomposition-adaptive Algebraic MultiGrid*

mit Startvektor v_i *oder* Lösen der Gleichung $Dv_{\text{neu}} = v_i$ mit dem Nullvektor als Startvektor, wie in Lüschers Arbeit [61] vorgeschlagen. Wir werden im folgenden den ersten Ansatz verwenden. Für weitere Details und Vergleiche, sowohl analytischer als auch numerischer Natur, der beiden Methoden DD-αAMG und Lüschers „inexact deflation"-Methode verweisen wir auf [35].

Algorithmus 7: Zweigitter-Setup-Phase

Eingabe: $n_{\text{inv}}, \eta, \nu$

Ausgabe: $v_1, \ldots, v_n, P, D_c$

```
Generiere v_1, ..., v_N ∈ C^n Zufallsvektoren
for i = 1, ..., N do
    v_i ← M_SAP^(η) v_i
end for
for j = 1, ..., n_inv do
    (Re-)Konstruiere P und D_c aus aktuellen v_1, ..., v_N
    for i = 1, ..., N do
        v_i ← C^(ν)
    end for
end for
```

Zur Veranschaulichung zeigt Abbildung 5.2 die Wirkung der Grobgitterkorrektur auf die aufsteigend sortierten Eigenmoden. Die erhoffte Fehlerreduktion in den kleinen Eigenmoden tritt (zumindest in diesem kleinen Beispiel, $n_{\text{inv}} = 10,\ \eta = \nu = 3$) deutlich ein. Mit dem Fehlerpropagator der Zweigitterverfahren (mit Nachglättung)

$$E_{\text{2G}} := (I - M_{\text{SAP}} D)(I - P D_c^{-1} P^H D)$$

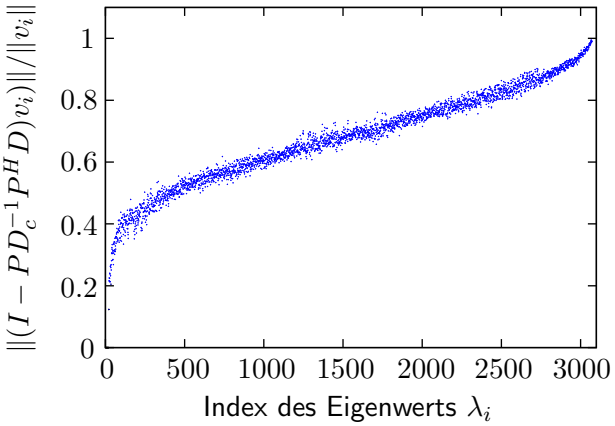


Abbildung 5.2: Fehlerreduktion der Grobgitterkorrektur bezüglich der Eigenmoden (4^4-Gitter).

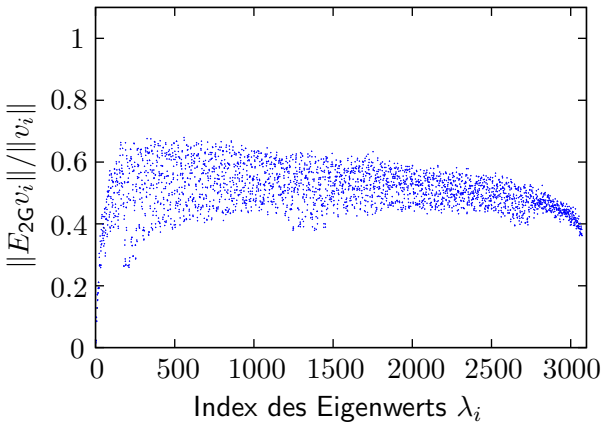


Abbildung 5.3: Fehlerreduktion der Zweigitterverfahrens bezüglich der Eigenmoden (4^4-Gitter).

zeigt Abbildung 5.3 die Wirkung des gesamten DD-αAMG-Verfahrens auf Fehlerreduktion in aufsteigend sortierten Eigenmoden am Beispiel eines Zweigitter-V-Zykels. Eine Fehlerreduktion entlang aller Eigenmoden ist zu beobachten.

Generell gibt es für ein (echtes) Mehrgitterverfahren verschiedene Zykel-Strategien, so wie den bereits angesprochenen V-Zykel (Algorithmus 6) oder den W-Zykel (ab drei Gitter). Als numerisch besonders stabil hat sich der sog. K-Zykel [79] erwiesen (K wie KRYLOV). Um diesen zu skizzieren, definieren wir Notationen im Zusammenhang mit echten Mehrgitterverfahren:

5.4.1 Definition

Sei L die Anzahl der Gitter, wobei das feinste Gitter das Erste ist, d. h., $D_1 := D$. Mit n_l, $l = 1, \ldots, L$, sei die Dimension des jeweils zugrundeliegenden Vektorraums auf jeder Gitterebene l bezeichnet. Die verschiedenen Interpolationen notieren wir dann mit

$$P_l : \mathbb{C}^{n_{l+1}} \to \mathbb{C}^{n_l}, \quad l = 1, \ldots, L-1.$$

Sie transportieren Informationen von Gitterebene $l+1$ nach Ebene l. Entsprechend transportieren die Operatoren P_l^H Informationen von Gitterebene l nach Ebene $l+1$. Die Grobgitteroperatoren sind rekursiv gegeben durch

$$D_l : \mathbb{C}^{n_l} \to \mathbb{C}^{n_l}, \quad D_l := P_{l-1}^H D_{l-1} P_{l-1},$$

für $l = 2, \ldots, L$. Die Glätter auf den verschiedenen Ebenen schreiben wir kurz als

$$M_l, \quad l = 1, \ldots, L-1.$$

Analog kennzeichnet ψ_l, dass der Vektor zur Gitterebene l bzw. zum Vektorraum $\mathbb{C}^{n_l}$ gehört. $\Diamond$

Das Vorgehen bei der K-Zykel-Strategie ist in Algorithmus 8 skizziert. Zum Verständnis muss betont werden, dass FGMRES stets selbst mit einem K-Zykel präkonditioniert wird, daher ist der Aufruf in Zeile 9 tatsächlich eine Rekursion, da eine neue Instanz von FGMRES mit Matrix D_{l+1} und rechter Seite η_{l+1} auf der nächst tieferen Gitterebene alsbald einen K-Zykel aufruft.

Die Setup-Phase für ein echtes Mehrgitterverfahren besteht nun ähnlich wie Algorithmus 7 aus zwei wesentlichen Phasen:

1. Eine Anfangsphase, gegeben durch Algorithmus 9, welche ausschließlich den Glätter benutzt um eine hierarchische Mehrgitterstruktur aufzubauen.

2. Eine iterative Phase, in der die Testvektoren der verschiedenen Gitterebenen durch Anwendung des aktuell verfügbaren Mehrgitterverfahrens verbessert werden. Algorithmus 10 skizziert hier das Vorgehen in Pseudocode.

Algorithmus 8: K-Zykel

Eingabe: l, η_l

Ausgabe: ψ_l

if $l = L$ **then**
  $\psi_l \leftarrow D_l^{-1}\eta_l$
else
  $\psi_l \leftarrow 0$
  for $i = 1, \ldots, \mu$ **do**
    $\psi_l \leftarrow M_l(\eta_l - D_l\psi_l)$ /* Vorglätten */
  end for
  $\eta_{l+1} \leftarrow P_l^H(\eta_l - D_l\psi_l)$
  $\psi_{l+1} \leftarrow \mathsf{FGMRES}(D_{l+1}, \eta_{l+1})$ /* Erläuterung siehe Text */
  $\psi_l \leftarrow \psi_l + P_l\psi_{l+1}$
  for $i = 1, \ldots, \nu$ **do**
    $\psi_l \leftarrow \psi_l + M_l(\eta_l - D_l\psi)$ /* Nachglätten */
  end for
end if

Algorithmus 9: Anfangs-Mehrgitter-Setup-Phase

Eingabe: l, N, η

Ausgabe: $v_j^{(1)}, \ldots, v_j^{(N)}$, P_j, D_{j+1} für $j = l, \ldots, L-1$

if $l = 1$ **then**
  Generiere N Zufallstestvektoren $v_1^{(1)}, \ldots, v_1^{(N)}$
else
  for $j = 1, \ldots, N$ **do**
    $v_l^{(j)} \leftarrow P_{l-1}^H v_{l-1}^{(j)}$ /* Restringiere Vektoren von feinerem Gitter */
  end for
end if
for $j = 1, \ldots, N$ **do**
  $v_l^{(j)} \leftarrow M_l^{(\eta)}(v_l^{(j)})$ /* η Glättungsschritte angewendet auf das homogene System $D_l x = 0$ mit Startvektor $v_l^{(j)}$ */
end for
Konstruiere P_l und setze $D_{l+1} := P_l^H D_l P_l$
if $l < L - 1$ **then**
  Rekursiver Aufruf von Algorithmus 9 auf Gitterebene l+1
end if

Es ist durchaus möglich, auf verschiedenen Gitterebenen eine unterschiedliche Anzahl von Testvektoren N_l zu verwenden. Aufgrund des verminderten Rechenaufwandes auf gröberen Gittern

Algorithmus 10: Iterative Mehrgitter-Setup-Phase

Eingabe: $l, N, n_{\text{inv}}, v_j^{(1)}, \ldots, v_j^{(N)}$, P_j, D_{j+1} für $j = l, \ldots, L-1$
Ausgabe: Verbesserte $v_j^{(1)}, \ldots, v_j^{(N_l)}$, P_j, D_{j+1} für $j = l, \ldots, L-1$

```
1  if l < L then
2      for i = 1, ..., n_inv do
3          for j = 1, ..., N do
4              for m = l, ..., L-1 do        /* Löser anw. auf allen Gitterebenen */
5                  ψ_m ← K-Zykel(l, v_{l-1}^(j))
6                  v_m^(j) ← ψ_m / ||ψ_m||
7              end for
8          end for
9          for m = l, ..., L-1 do
10             aktualisiere P_m, D_{m+1}
11         end for
12     end for
13     Rekursiver Aufruf von Algorithmus 10 auf Gitterebene l+1
14 end if
```

könnte eine wachsende Folge $N_l \leq N_{l+1}$ verwendet werden, aber die Auswirkungen sind wenig signifikant. Für weitere Details siehe [89].

DD-αAMG (Dreigitter)			
Setup-Schritte n_{inv}	Setup-Zeit	Löser-Zeit	Gesamtzeit
1	2.08s	6.42s	8.5s
2	3.06s	3.42s	6.48s
3	4.69s	2.37s	7.06s
4	7.39s	1.95s	9.34s
5	10.8s	1.82s	12.6s
6	14.1s	1.89s	16.0s
8	19.5s	2.02s	21.5s
10	24.3s	2.31s	21.6s

Tabelle 5.1: Vergleich Setup- und Löser-Zeiten, 48^4 Gitter, $\eta = 5$.[‡]

Bei der Setup-Phase gilt es generell, eine gute Balance zwischen Setup-Zeit und Löser-Zeit zu finden. Eine optimale Wahl für die Inversion von D bezüglich einer einzigen rechten Seite kann in Tabelle 5.1 bei $n_{\text{inv}} = 2$ ausgemacht werden. Soll dasselbe System mehrfach für verschiedene rechte Seiten gelöst werden, sind Setups im Bereich $n_{\text{inv}} = 5$ effizienter.

[‡]Daten aus [35]. Berechnung auf Juropa, Jülich Supercomputing Center (JSC).

Das Performancepotential des DD-αAMG-Verfahrens auf größeren Konfigurationen wird eindrucksvoll durch Abbildung 5.4 beschrieben. Die Abkürzung „mp oe“ steht hierbei für „mixed precision“ und „odd-even preconditioned“. Es beschreibt einen „state-of-the-art“ KRYLOV-Löser, welcher aufgebaut ist auf FGMRES(25) mit doppelter Maschinengenauigkeit, präkonditioniert mit 50 Iterationen BiCGStab in einfacher Maschinengenauigkeit. Das System ist darüber hinaus statisch präkonditioniert mit SCHUR-Komplement. $m_{\text{ud}} = -0.05294$ bezeichnet den physikalischen Massenparameter, der den thermischen Zustand der genutzten Konfiguration beschreibt und $m_{\text{crit}} = -0.05419$ bezeichnet die kritische Masse (Details dazu in [27, 28]; hieraus stammt auch die benutzte 64^4-BMW-c Konfiguration). Insbesondere wird hier zwischen den Massen des Up- und Down-Quark unterschieden. Weitere Details zu der Vielzahl an Parametern innerhalb der Löser, siehe [89, 35].

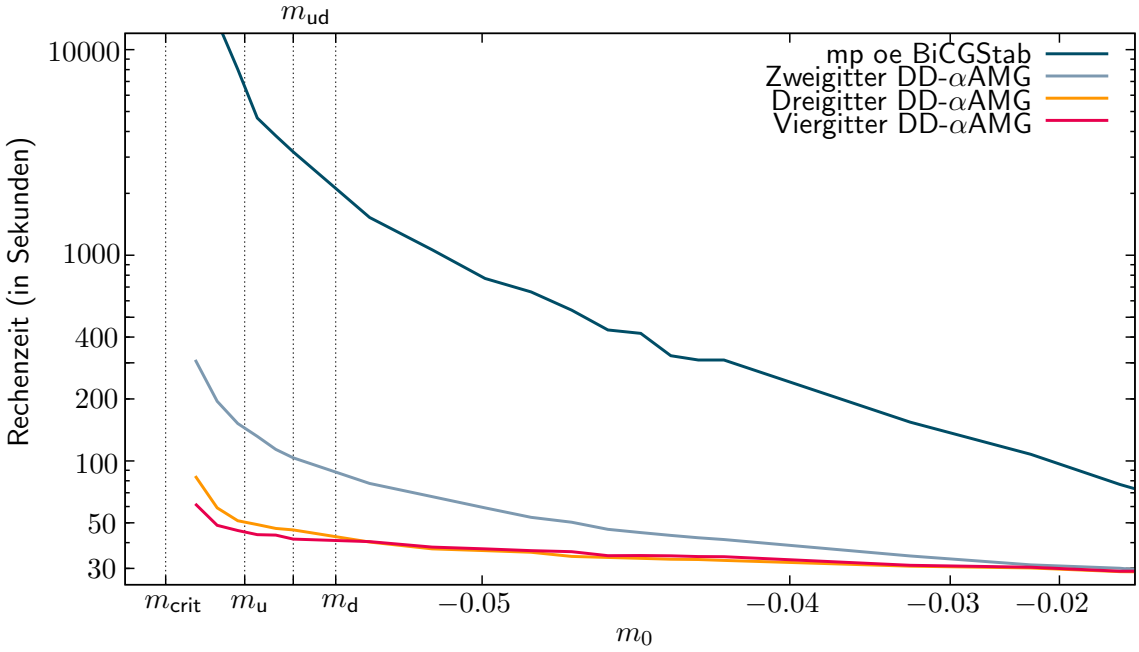


Abbildung 5.4: Skalierung von BiCGStab und DD-αAMG bezüglich des Massenparameters m_0 (64^4-Gitter, 128 Prozesse).[§]

Abbildung 5.5 zeigt einen Vergleich des DD-αAMG-Verfahrens mit dem (durch [35] inspirierten) aktuellen „inexact deflation“-Methode von Lüscher mit „inaccurate projection“, implementiert in der Programmbibliothek Open-QCD [62]. Sein Vorgehen unterscheidet sich von (Zweigitter-) DD-αAMG insbesondere durch die Konstruktion des Grobgitteroperators, welche die Γ_5-Symmetrie nicht erhält. Darüber hinaus konstruiert die „inexact deflation“-Methode den Prolongationsoperator so, dass nur halb so viele Variablen ins Grobgitter übernommen werden und konsequenterweise hat der Grobgitteroperator in Matrixdarstellung dann viermal weniger nicht-Null Einträge gegenüber unserem Verfahren, falls dieselbe Anzahl von Testvektoren verwendet wurde. In Tests stellte sich allerdings heraus, dass 30 Testvektoren (gegenüber 20 in DD-αAMG)

[§]Daten aus [35]. Berechnung auf Juropa, Jülich Supercomputing Center (JSC).

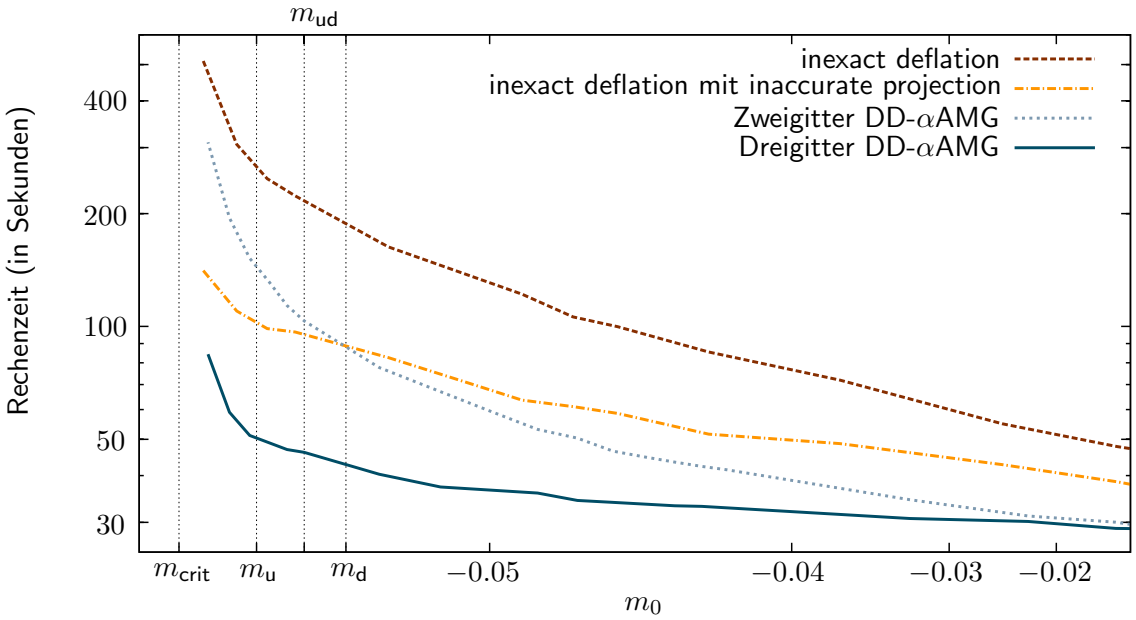


Abbildung 5.5: Skalierung der Lüscher-Methoden und DD-αAMG bezüglich des Massenparameters m_0 (64^4-Gitter, 128 Prozesse).¶

in den Lüscher-Methoden die beste Performance liefert. Obwohl mehr Testvektoren verwendet wurden, ist die Arbeit auf dem zweiten Gitter geringer im Vergleich zum Zweigitterverfahren. Sobald aber drei Gitter verwendet werden, sinkt die benötigte Arbeit um auf dem zweiten Gitter zu iterieren erheblich, was die Überlegenheit des Dreigitterverfahrens erklärt.

Die „Inexact deflation"-Methode mit „inaccurate projection" ist der Zweigitter-DD-αAMG bis zum Massenparameter m_{d} in Sachen Rechenzeiten unterlegen, die Skalierung von ersterem ist aber etwas besser. Ab Massenparametern kleiner als m_{d} ist die Skalierung des Zweigitterverfahren ähnlich zum (insgesamt schlechtestem) „inexact deflation"-Ansatz. Die Verbesserung der „inexact deflation"-Methode durch die Verwendung von inakkuraten Projektionen, führt zu einem verbesserten Skalierungsverhalten, ähnlich zum Verhalten des Dreigitterverfahrens. Insgesamt führt der Ansatz mehr Testvektoren zu nutzen (aber damit höheren Setup-Aufwand in Kauf zu nehmen), um dafür günstiger zu lösende Grobgittersysteme zu erhalten, generell zu einem gänzlich anderen Skalierungsverhalten als der des DD-αAMG-Ansatzes. Die insgesamt beste Performance (sowohl in Rechenzeit als auch Skalierung) wird durch Dreigitter-DD-αAMG erreicht. DreigitterDD-αAMG lohnt sich sogar wenn schwerere Massen als m_{d} verwendet werden. Voraussichtlich wird der Löser in Zukunft stark von seiner rekursiven Struktur profitieren, wenn immer größere Gitterkonfigurationen verwendet werden.

Ein Nachteil des robusten FGMRES-Verfahrens, welches als äußerer Löser für DD-αAMG dient, ist der sehr hohe Speicherbedarf (vgl. Abschnitt 3.2). Hier könnte dieses durch andere

¶Daten aus [89]. Berechnung auf Juropa, Jülich Supercomputing Center (JSC).

Verfahren ersetzt werden, wie zum Beispiel dem zu GMRES-ähnlichen FQMR-Verfahren (*Flexible Quasi-Minimal Residual*) von Szyld und Vogel [104] aus dem Jahre 2001. Das QMR-Verfahren basiert auf der „look-ahead"-Bi-LANCZOS-Methode, welcher „Breakdowns", also Verfahrensabbrüche ohne verwendbares Ergebnis, verhindert, an denen andere Bi-CG-Verfahren kränkeln. QMR benötigt, genau wie alle anderen Bi-CG-Verfahren und anders als GMRES, zwei Operator-Anwendungen pro Iteration. Es verfügt aber über kurze Rekursionen, kann also mit geringem Speicheraufwand betrieben werden. Praxistests zeigten beim vorliegenden Problem aber leider keine ernsthaften Verbesserungen bezüglich Konvergenzverhalten und Rechenzeit gegenüber FGMRES. Ebenfalls könnte DD-αAMG mit FQMRIDR (*Flexible QMR- Induced Dimension Reduction*) [107] und mit FBi-CGSTAB (*Flexible Bi-Conjugated Gradient STABilized*) [108] (Bi-CGSTAB wurde ursprünglich von van der Vorst [106], 1992 entwickelt) kombiniert werden. Beide Verfahren leiden bei vorliegendem Problem an stagnierenden Residuenverläufen. FGMRES ist also weiterhin die bessere Wahl.

Anmerkungen[‖]

[22] Georgi Iwanowitsch Petrow (wiss. Transliteration Georgij Ivanovič Petrov; * 18. Mai (jul.) / 31. Mai 1912 (greg.) in Pinega; † 13. Mai 1987) war ein sowjetischer Ingenieur. Von 1965 bis 1973 war Petrow Direktor des Instituts für Weltraumforschung der Akademie der Wissenschaften der UdSSR.

[‖] Alle Angaben aus der deutschen Wikipedia, stand 2017

6. Eigenlöser und eine physikalische Anwendung

In dem nun folgenden Hauptteil der Arbeit soll aufgezeigt werden, wie die Methoden aus dem vorherigen Kapitel angewendet werden können, um ein neues Verfahren zur einfacheren Berechnung von Eigenmoden des Operators $Q := \Gamma_5 D$ zu erhalten. Eine Variante der hier vorgestellten Methodik wurde gemeinsam mit der Arbeitsgruppe um A. Frommer an der Universität Wuppertal implementiert und in [7] publiziert.

Generell gibt es innerhalb der Gitter-QCD neben Lösungen der DIRAC-Gleichung auch großen Bedarf an Informationen über das Spektrum des WILSON-DIRAC-Operators D. Viele physikalische Eigenschaften sind insbesondere in betragsmäßig kleinen Eigenwerten und deren zugehörigen Eigenmoden dieses Operators kodiert und deshalb begehrt. Das komplexe Spektrum von D (vgl. Abbildung 2.5) ist vergleichsweise schwer zu fassen, weshalb in der Gitter-QCD eher die Eigenmoden des positiv definiten Operators $D^H D$ (reelles Spektrum) betrachtet werden, zum Beispiel um stochastisches Rauschen bei Kenngrößen wie den unverbundenen Fermionenschleifen [73] zu reduzieren. Neben diesen spielen in der Gitter-QCD die kleinen Eigenmoden des hermiteschen, maximal indefiniten, Operators Q (ebenfalls reelles Spektrum) für Anwendungen wie „low-mode averaging“ und andere [23, 10, 17] eine wohl noch größere Rolle.

Trotz erhöhter Symmetrieeigenschaft und reellem Spektrum gegenüber D ist die Eigenwertberechnung noch immer äußerst aufwendig. Nach Lüscher [61, 58] skaliert das Problem etwa mit VN_{eig}^2, wobei V dem Volumen des 4D-Raumzeitgitters und N_{eig} der Anzahl der zu berechnenden kleinen Eigenmoden entspricht. Die Anzahl der gesuchten kleinen Eigenmoden N_{eig} steigt dabei etwa in demselben Maße wie das Volumen V. Es gibt im Wesentlichen zwei Herangehensweisen, um dem Eigenwertproblem numerisch gegenüber zu treten: Zum einen sind das KRYLOV-Unterraumverfahren wie das ARNOLDI-Verfahren oder aber Verfahren die auf Shift-Invertierung des vorliegenden Operators basieren wie die RAYLEIGH[23]-Quotienten-Iteration. Im besten Fall kann beides kombiniert werden. Generell liegt es nahe, die hermitesche Struktur von Q auszunutzen, was für das ARNOLDI-Verfahren sehr einfach ist, denn dieses ist für hermitesche Operatoren äquivalent zum wesentlich weniger aufwendigen LANCZOS[24]-Algorithmus. Beim Einsatz der Shift-Invertierung bzw. des Gleichungsysstemlösers ist beispielsweise MINRES eines der bekanntesten Verfahren die für (und nur für) hermitesche Operatoren ausgelegt sind.

Um die Symmetrie des hermiteschen Operators auszunutzen, könnte DD-αAMG für den Operator Q derart angepasst werden, dass die Glättung via der SAP-Methode (vgl. Kapitel 4.3) symmetrisiert wird (durch entsprechendes Vor- und Nachglätten; da nicht-stationäre Verfahren vorliegen, ist dies nicht trivial). Ebenso muss der Grobgitteroperator Q_c hermitesch sein (dies ist vergleichsweise einfach zu realisieren), damit die Grobgittersysteme mit MINRES gelöst werden

können. Darüber hinaus ist es auch kein Problem die Schurkomplement-Präkonditionierung zu symmetrisieren. Erste vielversprechende Ergebnisse für kleinere Konfigurationen konnten leider im Größeren nicht bestätigt werden. Dies liegt nachweisbar daran, dass die Hermitizität der auftretenden Systeme sehr anfällig gegenüber Störungen ist, insbesondere sind die Grobgittersysteme aufgrund von Rundungsfehlern niemals hundertprozentig hermitesch. Weiter konvergiert auch die SAP-Methode nur sehr schlecht für Q, unabhängig davon, ob zum Lösen der Blocksysteme MINRES oder GMRES verwendet wird. Demnach müssen wir uns leider vorerst vom Gedanken verabschieden, die Hermitizität von Q ausnutzen zu wollen.

Wie bereits erwähnt, zeigten viele Tests, dass SAP nicht, oder nur unzureichend, als Glätter für Q funktioniert, selbst wenn Überrelaxationsvarianten verwendet werden. Das GMRES-Verfahren konvergiert bekanntlich für jedes eindeutig lösbare Gleichungssystem und mit passend gewählten Neustarts produziert das Verfahren auch für Q stabile Glättungsresultate (vgl. auch Abbildung 6.6). Durch diese Anpassung ähnelt DD-αAMG für Q (im Folgenden mit AMG bezeichnet) den in [4, 17, 81] vorgeschlagenen Mehrgitterverfahren für D und wird darüber hinaus etwa um den Faktor 2.5 langsamer als DD-αAMG für D. Da das adaptive Mehrgitterverfahren darauf basiert, kleine Eigenmoden von Q auf den gröberen Gittern zu behandeln, funktioniert dies auch für $Q - \sigma I$, solange σ betragsmäßig hinreichend klein ist. Demnach erlaubt dieses Vorgehen Shift-Invertier-Eigenlöser mit Mehrgitterverfahren zu beschleunigen. In aller Regel sind die Shifts σ innerhalb solcher Verfahren nahe an Eigenwerten λ von Q, daher wird $Q - \sigma I$ noch sehr viel schlechter konditioniert sein als bereits ohnehin. Dennoch werden wir zeigen, dass wenn wir auf Eigenlöser-Strategien setzen, die auf einer Invertierung des geshifteten Systems $Q - \sigma I$ beruhen, wir das in der Gitter-QCD verbreitete ARNOLDI-Verfahren (Programmbibliothek ARPACK und seine parallelisierte Variante PARPACK [100]) in Sachen Effizienz übertreffen können.

Eine der am weit verbreitetsten Shift-Invertier-Eigenlöser ist die RAYLEIGH[25]-Quotienten-Iteration, auf die wir uns nun konzentrieren.

6.1 Rayleigh-Quotienten-Iteration

Zum leichteren Verständnis der RAYLEIGH-Quotienten-Iteration betrachten wir zunächst die Potenzmethode nach VON MISES[26] (vgl. auch [44, 73, 93]). Sei hierfür im Beispiel $A \in \mathbb{R}^{n\times n}$ eine invertierbare Matrix mit reellen Eigenwerten, welche o. E. betragsmäßig aufsteigend geordnet werden können:

$$0 < |\lambda_1| < |\lambda_2| < \cdots < |\lambda_n|.$$

Angenommen wir haben zu jedem λ_i den zugehörigen Eigenvektor v_i, $i = 1, \ldots, n$, mit $\|v_i\|_2^2 := v_i^H v_i = 1$, so können wir jeden Vektor $x \in \mathbb{R}^n$ in die Eigenbasis von A,

$$x = \sum_{i=1}^{n} \alpha_i v_i,$$

entwickeln. Hieraus erhalten wir

$$A^k x = \sum_{i=1}^{n} \lambda_i^k \alpha_i v_i, \qquad k \in \mathbb{N},$$

und stellen fest, dass sich für große k auf der rechten Seite der Summand durchsetzt mit dem dominanten Eigenwert λ_n (o. E. $\alpha_n \neq 0$, sonst wähle anderes x). Wir können also mittels

$$A^k x \approx \lambda_n^k \alpha_n v_n$$

aus $A^k x$ den Eigenvektor zum betragsmäßig größten Eigenwert von A berechnen. $\|A^k x\|$ konvergiert demnach gegen $|\lambda_n|$ (unabhängig von der gewählten Norm). Das bestimmen des Vorzeichens des Eigenwerts benötigt einige Tricks, im komplexen Fall ist die Berechnung des Eigenwerts via $v^H A v$, mit $v := A^k x$, k hinreichend groß, aber sehr einfach, zumindest wenn keine vielfachen Eigenwerte auftreten (die *Diagonalisierbarkeit* von A spielt hier eine Rolle). Algorithmus 11 beschreibt das Verfahren (auch für komplexe Matrizen, ohne vielfache Eigenwerte), wobei die Normierung in jeder Iteration aus Stabilitätsgründen erfolgt.

Algorithmus 11: Potenzmethode nach VON MISES (bzgl. der EUKLID-Norm)

Eingabe: Startvektor $v^{(0)}$ mit $\|v^{(0)}\|_2 = 1$

Ausgabe: Approximatives Eigenpaar $(\lambda^{(k)}, v^{(k)})$ zum betragsgrößten Eigenwert von A

1 **for** $k = 1, 2, \ldots$ **do**
2 $\quad \tilde{v}^{(k)} \leftarrow A v^{(k-1)}$
3 $\quad v^{(k)} \leftarrow \tilde{v}^{(k)} / \|\tilde{v}^{(k)}\|_2$
4 $\quad \lambda^{(k)} \leftarrow (v^{(k-1)})^H \tilde{v}^{(k)}$
5 **end for**

In der vorliegenden Form kann die Potenzmethode nur verwendet werden um den betragsmäßig größten Eigenwert und den zugehörigen Eigenvektor zu bestimmen. Zur Berechnung anderer Eigenwerte kann die Matrix aber passend transformiert werden:

(i) Ersetzten wir Zeile 2 von Algorithmus 11 durch $\tilde{v}^{(k)} \leftarrow A^{-1} v^{(k-1)}$, so konvergiert das Verfahren gegen den Eigenvektor zum kleinsten Eigenwert λ_1. Insbesondere hat A^{-1} dieselben Eigenvektoren wie A.

(ii) Falls λ eine Näherung an einen beliebigen Eigenwert von A ist, selbst aber nicht im Spektrum $\sigma(A)$ liegt, so konvergiert Algorithmus 11 mit $(A - \lambda I)^{-1}$ gegen einen Eigenvektor zum Eigenwert nahe an λ, denn $(A - \lambda I)^{-1}$ besitzt die Eigenwerte $(\lambda_i - \lambda)^{-1}$, $i = 1, \ldots, n$. Diese Variante des Algorithmus 11 ist bekannt als die *gebrochene Iteration von* WIELANDT[27].

(iii) Passen wir schließlich in jeder Iteration λ durch die aktuelle Eigenwertnäherung $\lambda^{(k)}$ an, so führt dies auf die RAYLEIGH-Quotienten-Iteration (RQI), welche lokal quadratisch, für hermitesche Matrizen sogar lokal kubisch konvergiert (für einen Beweis siehe z. B. [83]).

Beste Voraussetzungen also für das Eigenwertproblem mit hermiteschem Q, wobei statt einer echten Invertierung $(Q - \sigma I)^{-1}$ selbstverständlich unser Mehrgitterverfahren angewendet auf $(Q - \sigma I)x = v$ zum Einsatz kommt.

Wir verbinden RQI auf folgende Weise mit dem Mehrgitterverfahren: Initial werden N_{eig} orthonormale Startvektoren $v_1, \ldots, v_{N_{\text{eig}}}$ korrespondierend zu Eigenwertstartwerten $\lambda_1 = \cdots = \lambda_{N\text{eig}} = 0$ fixiert. Unter Verwendung des Mehrgitterverfahrens werden alle Vektoren v_i durch eine Shift-Inversion $v_i \leftarrow (Q - \lambda_i I)^{-1} v_i$ aktualisiert. Anschließend werden die Vektoren $v_1, \ldots, v_{N_{\text{eig}}}$ re-orthonormalisiert und die Eigenwertapproximationen mittels $\lambda_i = v_i^H Q v_i$ angepasst. Diesen Prozess wiederholen wir solange, bis die Norm des Eigenvektorresiduums $\|Qv_i - \lambda_i v_i\|_2$ kleiner als eine vorgegebene Fehlertoleranz ε ist.

Das beschriebene Vorgehen ist im Algorithmus 12 zusammengefasst. In der Praxis starten wir mit den Testvektoren $v_1, \ldots, v_{N_{\text{eig}}}$, die in der Setup-Phase des Mehrgitterverfahrens generiert wurden, welche approximativ schon Eigenvektoren zu kleinen Eigenwerten darstellen. In diversen Tests stellte sich heraus, dass nicht immer *alle* N_{eig} kleinsten Eigenpaare (λ_i, v_i) berechnet werden. Zwar sind diejenigen λ_i, die sehr nahe an Null liegen immer dabei, bei größerem Abstand zum Ursprung fehlen jedoch einige der übrigen kleinsten Eigenwerte. Dies passiert insbesondere dann, wenn der zufällige Startvektor nur wenig mit der Richtung des gewünschten neuen Eigenvektors gemein hat oder wenn die aktuelle Eigenwert-Iterierte zu groß wird. Um die Häufigkeit dieses Effekts zu reduzieren, wurde in Zeile 4 eine Dämpfung eingebaut, welche die Reichweite der Shifts begrenzt.

Für die folgenden numerischen Ergebnisse mit $N_{\text{eig}} = 20$ wurde AMG mit der in Algorithmus 12 beschriebenen RAYLEIGH-Quotienten-Iteration kombiniert. Dabei wurden alle Berechnun-

Algorithmus 12: Rayleigh-Quotienten Iteration + AMG

Eingabe: Orthonormale Startvektoren $v_1, \ldots, v_{N_{\text{eig}}}$, Fehlertoleranz ε

Ausgabe: Eigenpaare $(\lambda_1, v_1), \ldots, (\lambda_{N_{\text{eig}}}, v_{N_{\text{eig}}})$

1 Setze $\lambda_i = 0,\ \varepsilon_i = 1,\ \forall\, i = 1, \ldots, N_{\text{eig}}$
2 **while** $\exists\, \varepsilon_i\ :\ \varepsilon_i > \varepsilon$ **do**
3 **for all** $i = 1, \ldots, N_{\text{eig}}$ `mit` $\varepsilon_i > \varepsilon$ **do**
4 $\sigma \leftarrow \lambda_i \cdot \max(1 - \varepsilon_i, 0)$
5 $v_i \leftarrow (Q - \sigma I)^{-1} v_i$ `/* GLS lösen via Mehrgitterverfahren */`
6 $v_i \leftarrow v_i - \sum_{j=1}^{i-1} (v_j^H v_i) v_j$
7 $v_i \leftarrow v_i / \|v_i\|_2$
8 `Aktualisiere` v_i `in Interpolation` P
9 $\lambda_i \leftarrow v_i^H Q v_i$
10 $\varepsilon_i \leftarrow \|Qv_i - \lambda_i v_i\|_2$
11 **end for**
12 **end for**

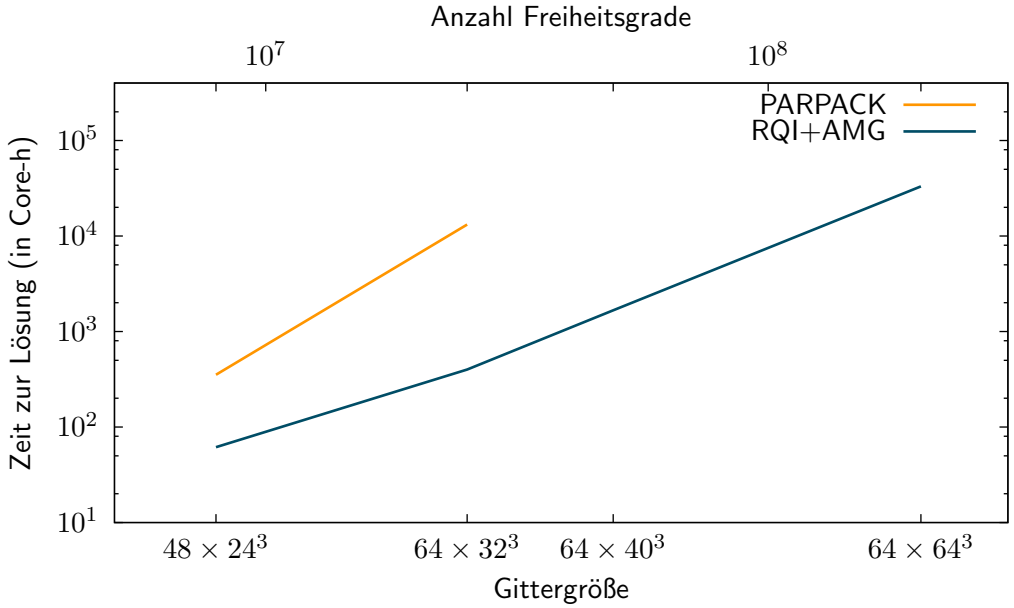


Abbildung 6.1: Vergleich zwischen PARPACK und RQI+AMG, benötigte Zeit um $N_{\text{eig}} = 20$ kleinste Eigenvektoren zu berechnen.

gen innerhalb der RQI bis auf die Invertierung mit doppelter Maschinengenauigkeit durchgeführt. Jede Shift-Inversion wurde mit FGMRES in doppelter Maschinengenauigkeit ausgeführt, welches wiederum flexibel mit dem algebraischen Mehrgitterverfahren (AMG) in einfacher Maschinengenauigkeit präkonditioniert wurde. Alle Ergebnisse wurden auf dem Juropa Rechencluster des Jülich Supercomputing Center (JSC) berechnet. Dieser Rechner besitzt 2208 Nodes, jeweils mit zwei Intel Xeon X5570 (Nehalem-EP) Quad-Core-Prozessoren. Er lässt ein Maximum von 8192 Kernen pro Job zu und der verwendete `ICC`-Compiler benutzte die Optimierungs-Flags `-O3`, `-ipo`, `-axSSE4.2` und `-m64`.

In Abbildung 6.1 ist ein Vergleich zwischen der RAYLEIGH-Quotienten-Iteration, kombiniert mit dem algebraischen Mehrgitterverfahren (RQI+AMG) und dem PARPACK [100] dargestellt. Letzteres ist die parallelisierte Implementierung der Open-Source-Programmbibliothek ARPACK, basierend auf dem *implicit restarted*-ARNOLDI-Algorithmus, welcher (nicht nur*) in der Gitter-QCD verbreitet ist. Das Verfahren (vgl. Algorithmus 1) baut zunächst einen KRYLOV-Unterraum mit fest gewählter Dimension N_{kw} (die Größe, bei der neu gestartet wird) auf und approximiert dort N_{eig} Eigenpaare von Q durch die der HESSENBERG-Matrix $H_{N_{\text{kw}}}$. Der Neustart innerhalb des Verfahrens behält die N_{eig} berechneten Eigenpaarapproximationen und verbessert diese mit einem neuen KRYLOV-Unterraum, bestehend aus N_{eig} alten Vektoren und $N_{\text{kw}} - N_{\text{eig}}$ neuen Eigenvektoren aus neuerlichen ARNOLDI-Iterationen. Mit der Neustartlänge von $N_{\text{kw}} = 100$ wurden hier die besten Resultate erzielt. Wir stellen fest, RQI+AMG schlägt PARPACK bezüglich der Rechenzeit bereits bei kleineren 48×24^3-Konfigurationen um eine Zehnerpotenz. Für Gitter mit Volumen 64×40^3 überschreitet PARPACK bereits das 24-Stunden Job-Limit bei 1024 Kernen,

*Die Matlab-Funktion `eigs` basiert ebenfalls auf ARPACK.

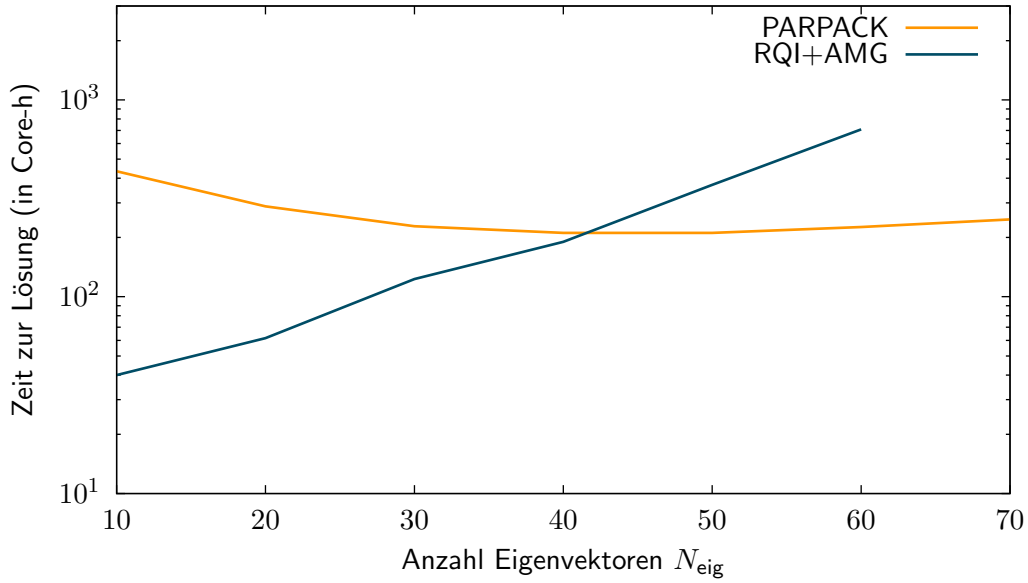


Abbildung 6.2: Vergleich zwischen PARPACK und RQI+AMG, Skalierung bezügl. der Anzahl kleinster Eigenmoden N_{eig}.

dennoch lässt sich an den Kurven ablesen, dass die Skalierung bezüglich der Gittergröße bei RQI wesentlich besser ist als bei PARPACK. Dies ist ein großer Vorteil, wenn es um aktuelle großvolumige Gittersimulationen geht. Die betrachteten Konfigurationen sind zwei-Flavour-Simulationen mit $m_{ud} \sim 290$ MeV und Gitterabstand $a \sim 0.071$ fm (weitere Details zu den verwendeten Konfigurationen sind in [8] zu finden).

Etwas anders ist die Situation leider, wenn wir die Skalierung bezüglich N_{eig} untersuchen. In den Rechnungen für Abbildung 6.2 verwenden wir konstante $N_{\text{kw}} = 200$, da Tests gezeigt haben, dass es kaum Einfluss auf die Laufzeit hat N_{kw} in Abhängigkeit zu N_{eig} zu setzen. Wir merken an, dass bei allen diesbezüglichen Tests immer $N_{\text{eig}} < \frac{1}{2}N_{\text{kw}}$ eingehalten wurde. Die Rechenzeit für RQI+AMG wächst rapide je mehr Eigenwerte berechnet werden sollen. PARPACK hingegen weißt nahezu konstante Laufzeiten bei wachsendem N_{eig} auf. Grundsätzlich skaliert der Orthogonalisierungsprozess des ARNOLDI-Verfahrens in der Größenordnung $\mathcal{O}(N_{\text{eig}}^2)$. In den vorliegenden Berechnungen dominieren aber, wegen der Neustarts, eher die Matrix-Vektor-Produkte und nicht der Orthogonalisierungsprozess. RQI+AMG andererseits verwendet alle N_{eig} berechneten Eigenvektorapproximationen für den Interpolationsoperator P des Mehrgitterverfahrens. Der Grobgitteroperator $Q_c = P^H Q P$ hat dann die Komplexität $\mathcal{O}(N_{\text{eig}}^2)$, da jeder Grobgitterknoten $2N_{\text{eig}}$ Variablen hält, die alle benachbarten Grobgitterknoten über eine (weniger dünnbesetzte) $2N_{\text{eig}} \times 2N_{\text{eig}}$-Matrix koppelt. Das Lösen des Grobgittersystems skaliert dadurch mindestens mit $\mathcal{O}(N_{\text{eig}}^2)$.

Abbildung 6.3 zeigt schließlich den Einfluss von Fluktuationen in acht verschiedenen, stochastisch unabhängigen Konfigurationen für zwei verschiedene Gittergrößen. Die Einflüsse sind wenig signifikant.

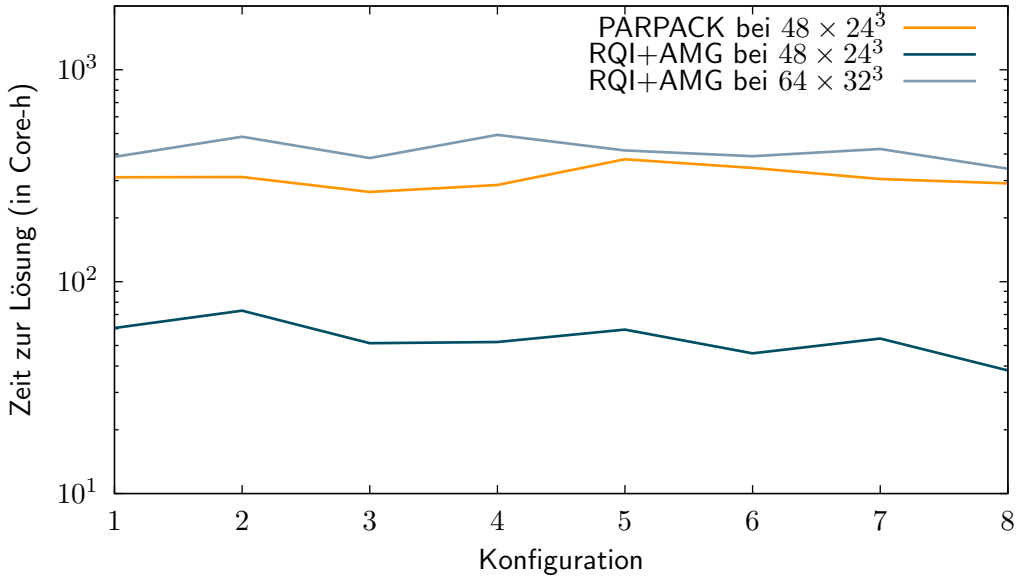


Abbildung 6.3: Vergleich zwischen PARPACK und RQI+AMG, Fluktuation über 2×8 stochastisch unabhängige Konfigurationen.

Um die Probleme bei der Skalierung bezüglich N_{eig} in den Griff zu bekommen, käme ein Ansatz in Frage, der nicht immer alle Eigenvektorapproximationen verwendet, um die Interpolation P zu verbessern. Tatsächlich ist es möglich, die Shift-Inversionen bezüglich Q durch Systeme mit dem Operator D zu ersetzen, um die Schwächen des algebraischen Mehrgitterverfahrens bezüglich Q zu umgehen. Genauer gilt aufgrund der Γ_5-Symmetrie des WILSON-DIRAC-Operators

$$(Q - \sigma I)^{-1} = (D - \sigma\Gamma_5)^{-1}\Gamma_5 \tag{6.1}$$

und numerische Tests zeigen, dass eine Anpassung in Zeile fünf von Algorithmus 12 zu

$$v_i \leftarrow (D - \sigma\Gamma_5)^{-1}\Gamma_5 v_i$$

den verloren gegangenen Laufzeitfaktor von 2.5 wieder nahezu gutmacht, auch weil es wieder möglich wird, SAP sinnvoll als Glätter zu verwenden.

Um die Skalierung bezüglich N_{eig} bei RQI+AMG weiter zu verbessern, soll im Folgenden ein neuer, alternativer Ansatz vorgestellt werden, der RQI durch ein Verfahren ersetzt, das auf Shift-Invertierung *und* Unterraumprojektionen basiert und zwar zusätzlich und anders zu den KRYLOV-Unterraumprojektionen und den Projektionen innerhalb des Mehrgitterverfahrens.

6.2 Jacobi-Davidson

Das JACOBI[28]-Davidson-Verfahren (JD) wurde 1996 von Sleijpen und van der Vorst [97] vorgeschlagen. Es kombiniert Ideen des Davidson-Verfahrens von Ernest R. Davidson [22] aus dem Jahre 1975 und des JACOBI-Verfahrens [49, 50] von 1845 (!). Ursprünglich für nur reell-symmetrische

Matrizen konzipiert, gibt es auch Versionen für hermitesche Operatoren [3] wie unser Q. Vereinfacht können wir uns das JD-Verfahren als ein Unterraumverfahren in Kombination mit RQI vorstellen, daher funktioniert der „Γ_5-Trick“ hier ebenso (vgl. Gleichung (6.1) und (6.2)).

Das Verfahren teilt sich im Wesentlichen in zwei Schritte, erstens einen Informationsgewinn aus dem aktuellen Ansatzraum und zweitens eine Erweiterung dieses Raumes. Ersteres wird durch ein Standard-RAYLEIGH-RITZ-Verfahren oder mittels harmonischer- (z. B. [98]) bzw. „refined“-RAYLEIGH-RITZ-Vektoren [31] realisiert. Zweiteres, die Erweiterung des Ansatzraumes, erfolgt durch approximatives Lösen der sog. *Korrekturgleichung*

$$\left(I - uu^H\right)(Q - \theta I)\left(I - uu^H\right)v = -r. \tag{6.2}$$

Um das JD-Verfahren im Detail zu verstehen, beginnen wir mit einem kurzen Überblick zum Davidson-Verfahren. Sei dazu $A \in \mathbb{R}^{n\times n}$ regulär und $U_k := [u_1, \ldots, u_k] \in \mathbb{R}^{n\times k}$ eine Matrix mit orthonormalen Spalten. Sei weiter (θ, w) ein Eigenpaar der projizierten Eigenwertgleichung

$$U_k^H A U_k w = \theta w.$$

Davidson schlug nun vor den Ansatzraum $\mathrm{Spann}\{u_1, \ldots, u_k\}$ mit der Suchrichtung

$$t := (D_A - \theta I)^{-1} r$$

zu erweitern, wobei $r := Au - \theta u$ das Residuum bezüglich des sog. RITZ-Paares (θ, u) mit $u := U_k w$ und D_A die Diagonale von A bezeichnet. u_{k+1} ist dann festgelegt durch Orthonormalisieren von t gegen $\mathrm{Spann}\{u_1, \ldots, u_k\}$. Algorithmus 13 beschreibt das Vorgehen in Pseudocode.

Algorithmus 13: Davidson-Verfahren

Eingabe: A, u_1

Ausgabe: U_{k+1}

for $j = 1, \ldots, k$ **do**
- $B \leftarrow U_j^H A U_j$
- `Berechne Eigenpaar` (θ, w) `von` B
- $u \leftarrow U_j w$
- $r \leftarrow Au - \theta u$
- $t \leftarrow (D_A - \theta I)^{-1} r$
- `Orthogonalisiere` t `gegen` $u_1, \ldots, u_j$
- $u_{j+1} \leftarrow t/\|t\|_2$
- $U_{j+1} \leftarrow [U_j, u_{j+1}]$

end for

Bemerkenswert ist die Tatsache, dass das Davidson-Verfahren scheitert, falls A selbst eine Diagonalmatrix ist, denn dann gilt

$$t = (D_A - \theta I)^{-1} r = u \in \mathrm{Spann}\{u_1, \ldots, u_k\},$$

d. h., der Ansatzraum wird nicht erweitert.

JACOBIs Ansatz (vgl. [97]), approximative Eigenpaare von diagonaldominanten Matrizen zu berechnen beruht auf folgender Überlegung: Ist $\alpha := a_{11}$ das größte Diagonalelement, dann ist α ein approximativer Eigenwert zum approximativen Eigenvektor $e_1 := [1, 0, \ldots, 0]^T \equiv [1, z_0^T]^T$. Um diese Approximationen zu verbessern, können wir die Gleichung

$$A \begin{bmatrix} 1 \\ z \end{bmatrix} \equiv \begin{bmatrix} \alpha & c^T \\ b & F \end{bmatrix} \begin{bmatrix} 1 \\ z \end{bmatrix} = \lambda \begin{bmatrix} 1 \\ z \end{bmatrix}$$

nach dem unbekannten (verbesserten) Eigenvektorteil z und dem unbekannten (verbesserten) Eigenwert λ lösen†. Dies ist äquivalent zum Lösen des Gleichungssystems

$$\begin{aligned} \lambda &= \alpha + c^T z, \\ (F - \lambda I)z &= -b. \end{aligned} \tag{6.3}$$

JACOBI schlug hier vor, dieses Gleichungssystem iterativ zu lösen:

$$\begin{aligned} \theta_k &= \alpha + c^T z^{(k)}, \\ (D_F - \theta_k I)z^{(k+1)} &= (D_F - F)z^{(k)} - b, \end{aligned}$$

mit D_F der Diagonalen von F.

Tatsächlich können die Eigenwertverbesserungen $z^{(k+1)}$ des JACOBI-Verfahrens auch derart interpretiert werden, dass sie einer Orthogonalitätsbedingung genügen, d. h., das JACOBI-Verfahren kann als Projektionsmethode angesehen werden. Betrachten wir hierfür einen simpleren Fall mit $Az = b$ (vgl. Gleichung (6.3)), dann gilt für die Iterierten komponentenweise

$$z_i^{(k+1)} = \frac{1}{a_{ii}} \left(b_i - \sum_{\substack{j=1 \\ j \neq i}}^{n} a_{ij} z_i^{(k)} \right).$$

Dies ist mit $L := \mathrm{Spann}\{e_i\}$ und $z_0 := [z_1^{(k)}, \ldots, z_{i-1}^{(k)}, 0, z_{i+1}^{(k)}, \ldots, z_n^{(k)}]^T$ äquivalent zu

$$\begin{aligned} & b_i - (Az^{(k+1)})_i = 0 \quad \text{mit} \quad z^{(k+1)} \in z_0 + L \\ \Leftrightarrow \quad & b - Az^{(k+1)} \perp L \quad\quad \text{mit} \quad z^{(k+1)} \in z_0 + L. \end{aligned}$$

Sleijpen und van der Vorst kombinierten nun das Davidson-Verfahren und die Projektions-Idee von JACOBI zu einer neuen iterativen Projektionsmethode (vgl. [97]). Genauer soll zu einem gegebenen RITZ-Paar (θ, u) der gegebene Ansatzraum $\mathcal{U}$ um eine Suchrichtung erweitert werden, die orthogonal zu u steht und approximativ die Korrekturgleichung (6.2) erfüllt.

Sei u eine Approximation an einen Eigenvektor von A und sei θ ein RITZ-Wert bezüglich u. Ähnlich zum JACOBI-Ansatz bestimmen wir eine Verbesserung des approximativen Eigenvektors u

† Jeder Eigenvektor v (der nicht orthogonal zu e_1 ist) lässt sich durch entsprechende Skalierung auf die Gestalt $v = (1, \tilde{v})^T$ bringen.

durch einen Anteil, der orthogonal zu u steht. Falls $\|u\|_2 = 1$ ist, sieht eine Orthogonalprojektion von A an $u^\perp$ ($\equiv \mathrm{Spann}\{u\}^\perp$) folgendermaßen aus:

$$B = (I - uu^H)A(I - uu^H).$$

Wegen $\theta = u^H Au$ gilt

$$A = B + Auu^H + uu^H A - \theta uu^H. \tag{6.4}$$

Sei (λ, x) ein Eigenpaar von A mit $x = u + v$, wobei v unbekannt mit der Eigenschaft $v \perp u$. Diese Darstellung ist möglich solange $x \neq \alpha u$, $\alpha \in \mathbb{C}$, und $x \notin u^\perp$. Falls $x = \alpha u$ haben wir einen Eigenvektor gefunden und sind fertig. Der Fall $x \in u^\perp$ hat Wahrscheinlichkeit Null. Für dieses Eigenpaar gilt

$$A(u + v) = \lambda(u + v).$$

Mit $Bu = 0$ und $u \perp v$ folgt mit Hilfe von Gleichung (6.4)

$$A(u + v) = Bv + Au + \lambda u - \theta u$$

und demnach

$$(B - \lambda I)v = -r,$$

wobei $r = Au - \theta u$. Da λ unbekannt ist, ersetzen wir den Wert durch den RITZ-Wert θ (oder falls vorhanden, durch eine andere Approximation an einen gesuchten Eigenwert) und erhalten die Korrekturgleichung

$$\left(I - uu^H\right)(A - \theta I)\left(I - uu^H\right)v = -r.$$

Die Lösbarkeit der Korrekturgleichung mit $v \in u^\perp$ ist äquivalent zur Existenz eines $\alpha \in \mathbb{C}$ mit

$$(A - \theta I)v = -r + \alpha u.$$

Dabei geht ein, dass $u^H r = 0$. Sofern $\theta \notin \sigma(A)$ gilt folgende Rechnung

$$\begin{aligned} v &= -(A - \theta I)^{-1} r + \alpha (A - \theta I)^{-1} u \\ &= -u + \alpha (A - \theta I)^{-1} u \end{aligned}$$

und α lässt sich so bestimmten, dass $v \perp u$, außer wenn $(A - \theta I)^{-1} u \in u^\perp$, was mit Wahrscheinlichkeit Null eintritt.

Darüber hinaus heißt das, dass der Ansatzraum, welcher um v erweitert wird und u bereits enthält, auch den Vektor $t := (A - \theta I)^{-1} u$ enthält. Mit anderen Worten: t ist eine Verbesserung des RITZ-Paares (θ, u) generiert durch einen Schritt RQI mit Shift θ und Startvektor u. Demnach kann das JD-Verfahren als eine Variante der RQI angesehen werden, und wir können Konvergenzraten erwarten, die mindestens so hoch sind wie die der RQI, d. h. quadratisch oder im hermiteschen Fall sogar kubisch [1].

Nachdem die Korrekturgleichung approximativ nach v gelöst wurde, wird v Anschließend gegen den Ansatzraum $\mathcal{U}$ orthonormalisiert und um diesen Vektor erweitert. Dann bestimmen wir

Algorithmus 14: Jacobi-Davidson-Verfahren (vereinfachte Darstellung)

Eingabe: Startvektor u

Ausgabe: Eigenwerte von A auf Diagonale von $U^H AU$

for $j = 1, \ldots$ **do**
- Berechne Eigenpaare (θ, w) von $U^H(A - \theta I)U$
- Wähle RITZ-Paar $(\theta, u := Uw)$
- Residuum $r \leftarrow Au - \theta u$ /* Abbruch, wenn $\|r\|_2$ klein genug */
- Löse Korrekturgleichung mit Mehrgitterverfahren:
- $\left(I - uu^H\right)(A - \theta I)\left(I - uu^H\right) v = -r$
- Orthonormalisiere v gegen bisherigen Ansatzraum
- Erweitere Ansatzraum $U \leftarrow [U, v]$

end for

darin das nächste RITZ-Paar. Diesen Prozess iterieren wir solange bis ein Abbruchkriterium erfüllt wird (d. h., die berechnete Eigenpaarapproximation ist hinreichend gut).

Damit ist das JD-Verfahren von Sleijpen und van der Vorst im Groben beschrieben. Algorithmus 14 beschreibt das Vorgehen (vereinfacht) im Pseudocode. Für die konkrete C-Implementierung des JD-Verfahrens verwenden wir Neustarts [109], harmonische RITZ-Werte [98] und optionale Polynomfilterung [114].

Für weiterführende Details und Analysen des JD-Verfahrens siehe [80] oder [109]. Tatsächlich

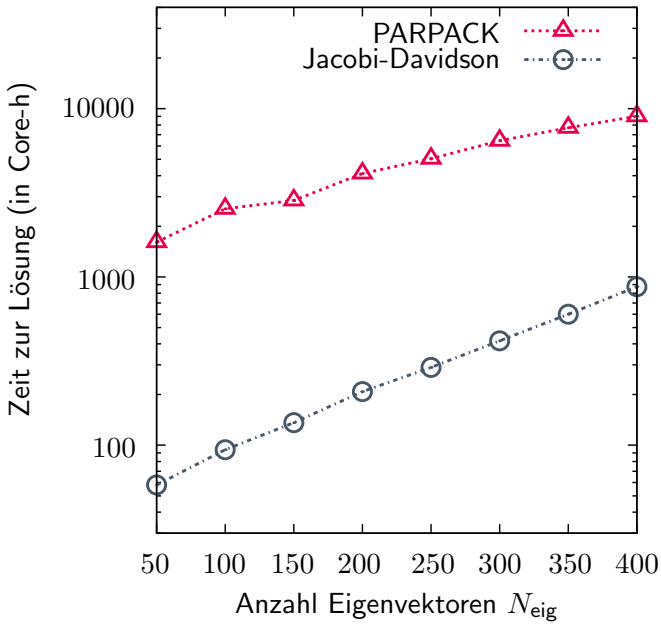


Abbildung 6.4: Vergleich zwischen PARPACK und JACOBI-Davidson, Skalierung bzgl. Anzahl kleinster Eigenmoden N_{eig}.

wird in den Quellen auch gezeigt, dass die Ansatzraumerweiterung im JD-Verfahren auf die robusteste Art geschieht, in dem Sinne, dass die Erweiterung möglichst unempfindlich gegenüber dem Fehler der approximativen Lösung der Korrekturgleichung ist.

Abbildung 6.4 zeigt das Ergebnis erster Tests mit bis zu 400 kleinsten Eigenvektoren eines 48×24^3 Gitters bei $m_0 = 350$ MeV. Das resultierende Verfahren ist über zehn Mal schneller als PARPACK.

PARPACK, bzw. bereits das zugrundeliegende ARNOLDI-Verfahren (im hermiteschen Fall das LANCZOS-Verfahren), kann bei vorliegenden Problem durch *Polynomfilterung* teilweise wesentlich beschleunigt werden [23], auch wenn z. B. die MATLAB-Funktion `eigs` entgegen früherer Versionen [88] stattdessen vorschlägt, mit A^{-1} oder $(A-\sigma I)^{-1}$ zu arbeiten. Zu diesem Vorschlag konnte in zahlreichen Tests keine Laufzeitverbesserungen festgestellt werden (ganz im Gegenteil).

6.3 Polynomfilter

Polynomfilterung ist in der Gitter-QCD eine verbreitete Methode und ermöglicht es, Bereiche des Spektrums hervorzuheben und gleichzeitig weniger gewünschte Bereiche zu dämpfen. Abbildung 6.5 und 6.6 veranschaulichen den Sachverhalt mit TSCHEBYSCHEFF[29]-Polynomen der Grade acht und zehn, sowie Polynomen konstruiert von Zhou und Saad [114]. Ausgenutzt wird bei der Polynomfilterung insbesondere, dass die Eigenvektoren einer Matrix Q identisch sind mit den Eigenvektoren von $p(Q)$, wobei $p(x)$ ein beliebiges nicht-konstantes Polynom ist. Die Eigenwerte von Q erhalten wir leicht aus den Eigenvektoren von $p(Q)$ zurück, indem vQv^H für jeden Eigenvektor v von $p(Q)$ berechnet wird. Durch die rekursive Definition der TSCHEBYSCHEFF-Polynome

$$T_0(x) := 1, \quad T_1(x) := x, \quad T_{k+1}(x) = 2T_k(x)x - T_{k-1}(x), \quad k = 2, 3, \ldots,$$

können die entsprechenden Matrixpolynome durch diverse weitere Matrix-Vektor-Operationen ausgerechnet werden. Mittels a-priori-Informationen über das Spektrum von Q via kleineren Konfigurationen (vgl. Kapitel 2.2), liefert z. B. $T_m(Q-10I)T_m(Q+10I)$ via Algorithmus 15 teils deutliche Beschleunigung im ARNOLDI-Verfahren, vgl. [23]. Es gibt auch Varianten der Rekursion, sodass

Algorithmus 15: TSCHEBYCHEFF-Polynomfilter

Eingabe: Matrix Q, Vektor v, Shift σ und Polynomgrad $m \geq 2$

Ausgabe: $v \leftarrow T_m(Q + \sigma)v$

1 $v_0 \leftarrow v$
2 $v_1 \leftarrow Qv + \sigma v$
3 **for** $k = 2, \ldots, m$ **do**
4 $\quad v \leftarrow 2Qv_1 - v_0$
5 $\quad v_0 \leftarrow v_1$
6 $\quad v_1 \leftarrow v$
7 **end for**

$T_m^{[a,b]}(x)$ das Spektrum auf einem beliebigen Intervall $[a, b]$ (statt auf$[-1, 1]$) dämpft, vgl. [115]. Verbreitet ist auch $T_m(Q^2)$ für m bis zu 20 zu verwenden. Der Clou ist nun aber, dass auch das JD-Verfahren mit Polynomfiltern beschleunigt werden kann [114].

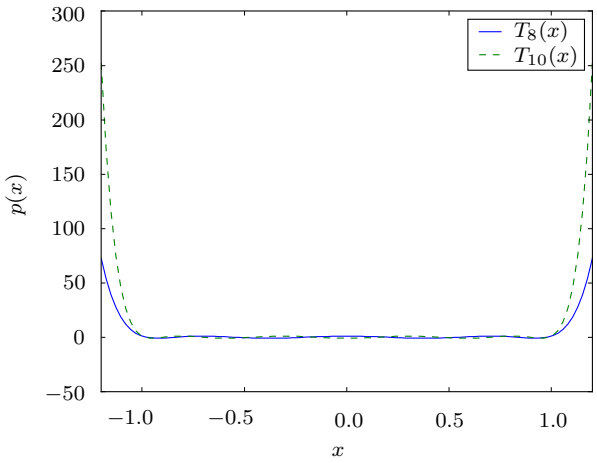


Abbildung 6.5: Zwei TSCHEBYCHEFF-Polynome.

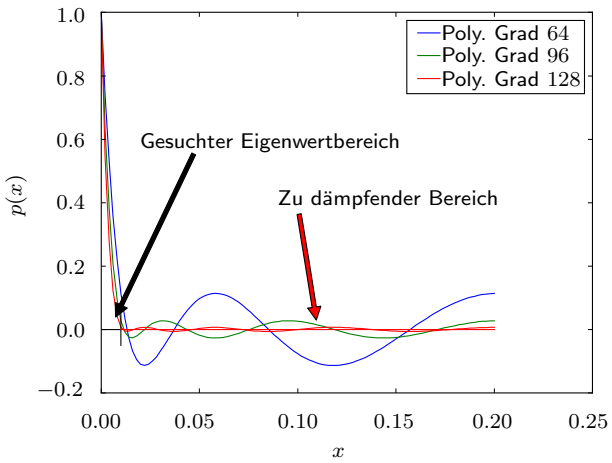


Abbildung 6.6: Polynome von Zhou und Saad [114].

Als Ausblick ist das JD-basierte PRIMME-Verfahren (*PReconditioned Iterative MultiMethod Eigensolver*) von Stathopoulos et al. [113] von 2016 interessant. Darüber hinaus arbeitet auch der FEAST-Algorithmus von Polizzi et al. [105] von 2014 mit approximativen Inversionen, die mit unserem hochperformanten DD-αAMG-Löser harmonieren könnten.

Zum Abschluss der Arbeit betrachten wir im folgenden Abschnitt eine physikalische Anwendung von Eigenlösern für die Gitter-QCD, das approximative „low-mode averaging“.

6.4 Approximative Eigenmoden und deren physikalische Anwendung

Wir wollen die beschriebenen Eigenlöser für das „low-mode averaging“ benutzen, welches verwendet wird um stochastisches Rauschen in zusammenhängenden (engl. *connected*) [25, 40] und unzusammenhängenden (engl. *disconnected*) [73, 10, 32, 9] Hadron-Observablen zu reduzieren. Dabei konzentrieren wir uns auf sog. Pion- und η-Meson-Korrelatoren.

Diese Art Rauschunterdrückungstechniken sind besonders wichtig bei fermionischen n-Punkt-Funktionen von sog. Flavour-Singulett-Größen. Für η-Mesonen der Zwei-Flavour-Theorie, ist beispielsweise ein Interpolator gegeben durch

$$O_x^\eta = \frac{1}{\sqrt{2}} \left(\bar{u}_x \Gamma_5 u_x + \bar{d}_x \Gamma_5 d_x \right),$$

wobei $\bar{u}_x, u_x, \bar{d}_x, d_x$ Spinore der verschiedenen Pion-Flavours bezüglich des Raumzeitpunkts x bezeichnen. Für weitere Details zur Notation siehe [39]. Mittels WICK[30]-Rotationen angewandt

auf Zwei-Punkt-Funktionen erhalten wir für den Fall von entarteten Quarkmassen den η-Korrelator

$$C_\eta = \langle O_x^\eta \bar{O}_y^\eta \rangle \propto \mathrm{Spur}\left(D_{x,y}^{-1}\Gamma_5 D_{y,x}^{-1}\Gamma_5\right) - 2\,\mathrm{Spur}(D_{x,x}^{-1}\Gamma_5)\,\mathrm{Spur}(D_{y,y}^{-1}\Gamma_5), \tag{6.5}$$

mit verschiedenen Quarkpropagatoren. Bspw. transportiert $D_{x,y}^{-1}$ einen Up- oder Down-Quark von Raumzeitpunkt x nach y (insbesondere unterscheiden wir nicht zwischen D_u und D_d, d. h., wir betrachten eine sog. exakte Isospin-Geometrie). Der erste Term auf der rechten Seite von Gleichung (6.5), dem zusammenhängenden Anteil, ist auf einer einzelnen Quelle y_0 unter Verwendung der Γ_5-Symmetrie und der Translationsinvarianz günstig zu berechnen:

$$\mathrm{Spur}\left(D_{x,y_0}^{-1}\Gamma_5 D_{y_0,x}^{-1}\Gamma_5\right) = \mathrm{Spur}\left(D_{x,y_0}^{-1}(D_{x,y_0}^{-1})^H\right),$$

wobei sich die Spur wie oben auf die Spin- und Farbindizes bezieht. Für den komplizierteren, unzusammenhängenden Anteil startet und endet der Propagator an demselben Raumzeitknoten und die Berechnung der „Schleife“ $D_{x,x}^{-1}\Gamma_5$ würde die Inversion der vollen Matrix D benötigen, was viel zu aufwändig wäre. Stattdessen werden stochastische Methoden verwendet (vgl. [9]), d. h.,

$$Q^{-1} = D^{-1}\Gamma_5 = \frac{1}{N_{\text{stoch}}}\sum_{i=1}^{N_{\text{stoch}}} s_i \eta_i^H + \mathcal{O}\left(\frac{1}{\sqrt{N_{\text{stoch}}}}\right) \tag{6.6}$$

mit hinreichend großem N_{stoch} und einer approximativen Lösung des linearen Gleichungssystems

$$Qs_i = \eta_i, \tag{6.7}$$

wobei η_i ein zufällig verrauschter Vektor mit Eigenschaften

$$\frac{1}{N_{\text{stoch}}}\sum_{i=1}^{N_{\text{stoch}}} \eta_i = \mathcal{O}\left(\frac{1}{\sqrt{N_{\text{stoch}}}}\right) \quad \text{und} \quad \frac{1}{N_{\text{stoch}}}\sum_{i=1}^{N_{\text{stoch}}} \eta_i \eta_i^H = I + \mathcal{O}\left(\frac{1}{\sqrt{N_{\text{stoch}}}}\right)$$

ist. Eine verbreitete Wahl, welche wir hier auch verwenden, ist die Einträge von η_i mit zufälligen Werten aus $\mathbb{Z}/2\mathbb{Z} + i\mathbb{Z}/2\mathbb{Z}$ zu füllen.

In Gleichung (6.6) akkumulieren sich mehrere Quellen stochastischen Rauschens, die sich zu dem inhärenten stochastischen Verhalten der Eichfelder aufaddieren. Genauer heißt das, N_{stoch} muss so groß gewählt werden, dass das stochastische Rauschen der Eichfelder im gesamten Fehlerrauschen dominiert. Dies benötigt weitere Lösungen der Gleichung (6.7) und wird bei kleinen Pionmassen und großem Gittervolumen schnell sehr rechenintensiv, selbst wenn moderne Mehrgitterverfahren verwendet werden.

Um das stochastische Rauschen zu reduzieren, gibt es eine Fülle an Techniken wie z. B. die Partitionierung [32, 11, 108], die „truncated solver“-Methode [9] oder das „low-mode averaging“ (in diesem Zusammenhang bekannt als „truncated eigenmode acceleration“ [10, 32]), und viele mehr. Welche Kombination dieser Methoden am besten funktioniert, hängt im Allgemeinen nicht nur von der Effizienz der benutzten Löser ab, sondern auch von der betrachteten Observable. Der η-Korrelator ist bekannt dafür, dominant in den kleinen Eigenmoden zu sein [73], daher ist er eine ideale Größe, um die von unserem Eigenlöser generierten, approximativen Eigenpaare zu testen.

6.4.1 Low-Mode Averaging

Die Grundidee des „low-mode averaging“ (LMA) ist es, die inverse des Operators Q additiv in zwei Teile zu trennen:

$$Q^{-1} = Q^{-1}_{\mathsf{low}} + Q^{-1}_{\mathsf{high}},$$

wobei Q^{-1}_{low} den Beitrag der N_{eig} kleinsten Eigenmoden von Q enthält:

$$Q^{-1}_{\mathsf{low}} = \sum_{i=1}^{N_{\mathsf{eig}}} \frac{1}{\lambda_i} v_i v_i^H. \tag{6.8}$$

Entsprechend ist $Q^{-1}_{\mathsf{high}} = Q^{-1} - Q^{-1}_{\mathsf{low}}$, vgl. (6.6).

Für den η-Korrelator (vgl. Gleichung 6.5) wenden wir LMA sowohl auf den verbundenen (Pion-) Korrelator

$$C_{\mathsf{con}}(x,y) = \mathrm{Spur}\left(Q^{-1}_{x,y} Q^{-1}_{y,x}\right)$$

als auch auf den unverbundenen Beitrag

$$C_{\mathsf{dis}}(x,y) = \mathrm{Spur}\left(Q^{-1}_{x,x}\right) \mathrm{Spur}\left(Q^{-1}_{y,y}\right)$$

an, jeweils wiederum durch Aufteilen der Terme. Für den ersteren, verbundenen Teil mitteln wir über die Raumdimensionen und dämpfen über die EUKLIDische Zeitdimension t, d. h.,

$$C_{\mathsf{con}}(t) = C^{\mathsf{low}}_{\mathsf{con}}(t) + C^{\mathsf{high}}_{\mathsf{con}}(t) = C^{\mathsf{low}}_{\mathsf{con}}(t) + \left(C^{\mathsf{p2a}}_{\mathsf{con}}(t) - C^{\mathsf{low,p2a}}_{\mathsf{con}}(t)\right), \tag{6.9}$$

wobei die einzelnen Terme mit $x = (\tilde{x}, t_0 + t)$, $y = (\tilde{y}, t_0)$, $y_0 = (\tilde{y}_0, t_0)$ gegeben sind durch

$$\begin{aligned}
C^{\mathsf{low}}_{\mathsf{con}}(t) &= \frac{1}{V} \sum_{\tilde{x},\tilde{y},t_0} \mathrm{Spur}\left((Q^{-1}_{\mathsf{low}})_{x,y} (Q^{-1}_{\mathsf{low}})_{y,x}\right), \\
C^{\mathsf{low,p2a}}_{\mathsf{con}}(t) &= \sum_{\tilde{x}} \mathrm{Spur}\left((Q^{-1}_{\mathsf{low}})_{x,y_0} (Q^{-1}_{\mathsf{low}})_{y_0,x}\right), \\
C^{\mathsf{p2a}}_{\mathsf{con}}(t) &= \sum_{\tilde{x}} \mathrm{Spur}\left((D^{-1})_{x,y_0} (D^{-1})^H_{y_0,x}\right).
\end{aligned}$$

Die konkrete Berechnung ist wie folgt: Zuerst berechnen wir die kleinen Eigenmoden von $C^{\mathsf{low}}_{\mathsf{con}}$, welcher die gesamte („all-to-all“) Information des Operators Q benötigt. Die Korrektur über die restlichen, hohen Eigenmoden (die Terme in der Klammer von Gleichung (6.9)) wird aus den Eigenmoden der „point-to-all“-Zweipunkt-Funktionen $C^{\mathsf{2pa}}_{\mathsf{con}}$ und $C^{\mathsf{low,p2a}}_{\mathsf{con}}$ am Punkt y_0 gewonnen.

Für den unzusammenhängenden Teil korrelieren wir zwei Schleifen zu den Zeitpunkten t_0 und $t_0 + t$:

$$C_{\mathsf{dis}}(t) = \frac{1}{N_t} \sum_{t_0} L(t_0 + t) L(t_0),$$

wobei die Information zu den kleinen Eigenmoden aus wiederum aufgeteilten Einzelschleifen stammt:

$$L(t) = \sum_{\tilde{x}} \mathrm{Spur}\left(Q^{-1}_{x,x}\right) = L^{\mathsf{low}}(t) + L^{\mathsf{high}}(t).$$

Entsprechend Gleichung (6.8) werden die kleinen Eigenmoden via

$$L^{\mathsf{low}}(t) = \sum_{\tilde{x}} \mathrm{Spur}\left((Q_{\mathsf{low}}^{-1})_{x,x}\right)$$

berechnet. Um die restlichen Eigenmoden von Q in diesem Fall zu extrahieren benutzen wir den Orthogonalprojektor

$$\mathcal{P} = I - \sum_{i=1}^{N_{\mathsf{eig}}} v_i v_i^H,$$

und mitteln

$$L^{\mathsf{high}}(t) = \sum_{\tilde{x}} \mathrm{Spur}\left((\mathcal{P}Q)_{x,x}^{-1}\right).$$

Wir merken an, dass in diesem Fall $\mathcal{P}Q = Q\mathcal{P}$ gilt.

Wenn nun, wie im Fall des η-Korrelators, die kleinen Eigenmoden dominieren, so ist die Berechnung von Q_{high}^{-1} sehr günstig. Die größte Rechenzeit liegt indes beim Berechnen der Eigenmoden von Q, weshalb sich LMA umso mehr lohnt, desto öfter die Eigenmoden verwendet werden können.

6.4.2 Approximatives LMA

Neben dem Verbessern der Eigenlöser können auch die Rechenkosten für LMA selbst reduziert werden, indem die Fehlertoleranz der einzelnen Eigenmoden heruntergeschraubt wird und danach mit eben dieser Fehlertoleranz auftretendes Rauschen reduziert wird. Sei dazu mit $(\tilde{\lambda}_i, \tilde{v}_i)$, $i = 1, \ldots, N_{\mathsf{eig}}$, das i-te approximative Eigenpaar zum exakten Eigenpaar (λ_i, v_i) von Q bezeichnet. Dann ist die Eigenmoden-Fehlertoleranz durch

$$\varepsilon_i = \|Q\tilde{v}_i - \tilde{\lambda}_i \tilde{v}_i\|_2,$$

gegeben, wobei wir annehmen, dass die $\tilde{v}_i$ orthonormalisiert sind. Seien weiter

$$A_{ij} := \tilde{v}_i^H Q \tilde{v}_j, \qquad i, j = 1, \ldots, N_{\mathsf{eig}},$$

die Einträge einer Matrix A, die wir zum Messen der Inexaktheit der Eigenmoden verwenden können: Mit

$$\tilde{v}_i = v_i + v_i^\delta$$

und dem KRONECKER-Delta δ_{ij} können wir die Einträge A_{ij} schreiben als

$$A_{ij} = \lambda_j \delta_{ij} + \lambda_i v_i^H v_j^\delta + \lambda_j (v_i^\delta)^H v_j + (v_i^\delta)^H Q v_j^\delta.$$

Insbesondere ist A eine Diagonalmatrix, für den Fall, dass alle Eigenpaare exakt wären, d. h., $v_i^\delta = 0$, $i = 1, \ldots, N_{\mathsf{eig}}$. Wenn wir nun in Gleichung (6.8) alle inversen Eigenwerte durch die Einträge der Inverse von A ersetzen, ergibt sich

$$\tilde{Q}_{\mathsf{low}}^{-1} := \sum_{i,j=1}^{N_{\mathsf{eig}}} (A^{-1})_{ij} \tilde{v}_i v_j^H, \tag{6.10}$$

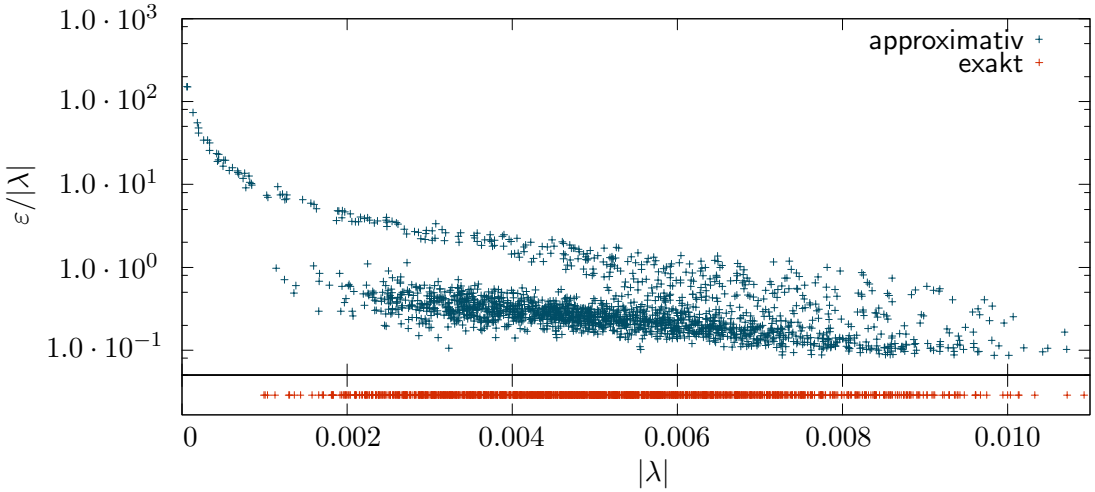


Abbildung 6.7: Jeweils 30 kleinste Eigenwerte von Q zu 64 Konfigurationen mit 64×40^3 Knoten, $m_{\text{ud}} \approx 290$ MeV. Der untere Bereich zeigt das „exakte“ Spektrum ($\varepsilon \leq 10^{-8}$). Die Eigenwerte im oberen Bereich sind mittels der Testvektoren der Setup-Phase berechnet mit dem relativen Fehler $\frac{\varepsilon}{|\lambda|}$.

wobei $\tilde{Q}^{-1}_{\text{low}} = Q^{-1}_{\text{low}}$ für exakte Eigenpaare gilt. Ersetzen aller Q^{-1}_{low} im vorherigen Abschnitt 6.4.1 durch $\tilde{Q}^{-1}_{\text{low}}$ bewirkt eine gleichmäßige Einflussnahme aller inexakten Eigenmoden und wir erhalten trotz geringerer Fehlertoleranzen ähnliche, stochastisch unverzerrte Resultate.

6.4.3 Verwenden von Testvektoren für LMA

Um weitere Rechenkosten für LMA einzusparen, können wir noch einen Schritt weitergehen und die Testvektoren, die in der Setup-Phase des Mehrgitterverfahrens anfallen, als approximative Eigenvektoren verwenden, da diese bereits ebenfalls kleine Eigenmoden approximieren. Tatsächlich hat sich in numerischen Tests bei vorliegender Gitter-Konfiguration gezeigt, dass sich nach 30 Setup-Iterationen und Fehlern der Größenordnung $\varepsilon \approx 10^{-1}$ durchaus gute Ergebnisse mit LMA erzielen lassen.

Da innerhalb der Setup-Phase die Testvektoren als Zufallsvektoren initialisiert werden und der Mittelwert über alle $r^H Q r$ verschwindet (wobei r hier einen Zufallsvektor bezeichnet), hat dies den Effekt, dass einige Eigenwerte unterrepräsentiert sind und die Eigenapproximationen spiegeln die sog. Massenlücke nicht wider. Dies hat störenden Einfluss auf die Qualität von $\tilde{Q}^{-1}_{\text{low}}$.

Glücklicherweise können solche Abweichungen durch Untersuchen der relativen Toleranz $\varepsilon/|\tilde{\lambda}|$ der Eigenmoden erkannt werden, vgl. Abbildung 6.7.

Es stellt sich heraus, dass die Eigenmoden, welche die Massenlücke nicht widerspiegeln, genau die sind mit großem $\varepsilon/|\tilde{\lambda}|$. Für unsere Berechnungen schneiden wir daher die Menge der verwendeten Eigenpaare ab:

$$\left\{(\tilde{\lambda}, \tilde{v}, \varepsilon)\right\} \rightsquigarrow \left\{(\tilde{\lambda}, \tilde{v}, \varepsilon) \;:\; \frac{\varepsilon}{|\tilde{\lambda}|} \leq C\right\}$$

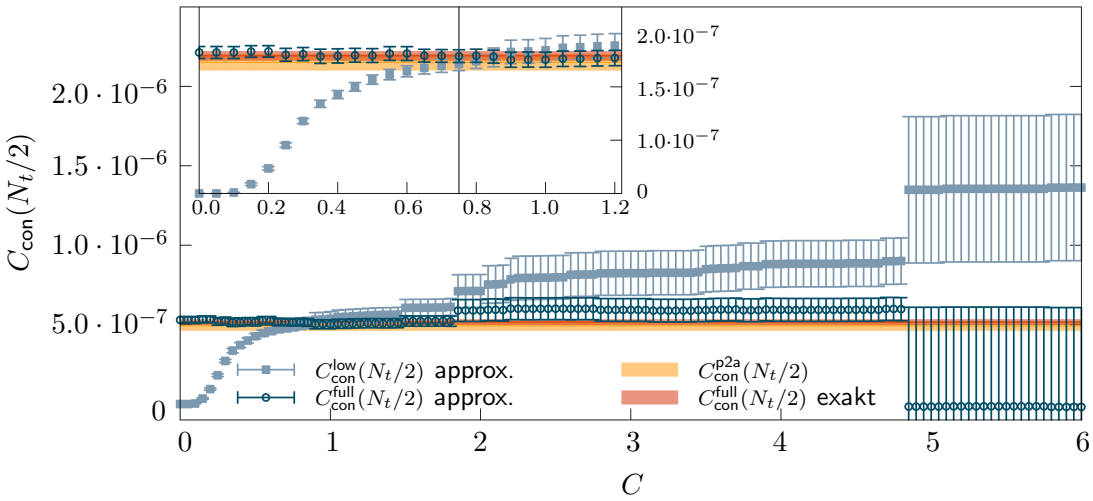


Abbildung 6.8: Der pseudoskalare verbundene Zwei-Punkt-Korrelator zum Zeitpunkt $t = N_t/2$, berechnet mit den abgeschnittenen Eigenmoden bezüglich C. Die Kästchen zeigen den Beitrag der inexakten kleinen Eigenmoden, die Kreise zeigen den vollen Beitrag der Zwei-Punkt-Funktion. Zum Vergleich zeigt der rote Streifen die berechneten Werte unter Verwendung von 20 exakten Eigenmoden ($\varepsilon \leq 10^{-8}$), sowie der gelbe Streifen die konventionellen „point-to-all“ Resultate.

Zu beachten ist, dass nach gefundener Schranke C, diese unabhängig vom Gittervolumen oder der Pionmasse ist, da sich jede Normalisierung herauskürzt.

Ein guter Test zur Bestimmung der Schranke C ist es, den Beitrag der kleinen Eigenmoden der unverbundenen Zweipunkt-Funktion zu untersuchen, da diese ohne stochastische Abschätzung berechnet werden kann. Abbildung 6.8 zeigt das Verhalten von $C^{\mathrm{low}}_{\mathrm{con}}$ in Abhängigkeit von C. Die Daten stammen aus einer zentralen Zeitebene, wo die relativen Beiträge der kleinen Eigenmoden am größten sind und demnach der Effekt der inexakten Eigenmoden am besten gesehen werden kann. Falls C zu groß gewählt wird, treten große Fehler in den Funktionen $C^{\mathrm{low}}_{\mathrm{con}}$ und $C^{\mathrm{full}}_{\mathrm{con}}$ auf, denn viele irrelevante Richtungen dominieren die Eigenmoden. Wird C zu klein gewählt, können wir vom Effekt von LMA nicht mehr profitieren. Nachdem die Korrektur bezüglich der großen Eigenmoden (vgl. Gleichung (6.9)) vollzogen wird, sind die Ergebnisse unabhängig von C bezüglich der Fehlertoleranz korrekt, wie an den horizontalen Streifen im Fall des „point-to-all“ und dem LMA-Fall zu erkennen ist. Tatsächlich scheint $C = 0.75$ ein guter Kompromiss zu sein.

Wir betonen, dass, obwohl $\tilde{Q}^{-1}_{\mathrm{low}}$ sowohl von der Anzahl als auch der Genauigkeit der Eigenmoden beeinflusst ist, die Korrektur der größeren Eigenmoden stabil und stochastisch unverzerrt ist. Variieren des Abschneideparameters C zeigt dies empirisch.

6.4.4 Numerische Ergebnisse

Für den ersten Praxistest benutzen wir dieselbe Gitterkonfiguration wie im vorhergehenden Abschnitt: ein moderat großes Gitter mit Volumen $V = 64{\times}40^3$ mit zwei Seequark-Flavours generiert von QCDSF [8] bei Pionmasse $m_{\mathrm{ud}} \approx 290$ MeV und Invers-Kopplung $\beta = 5.29$, d. h. Gitterab-

stand $a \approx 0.071$ fm. Um die Beiträge angeregter Zustände zu reduzieren, verwenden wir 400 Schritte *Wuppertal-Smearing* [42] mit Parameter $\delta = 0.25$ für die Quark-Quellen und -Senken sowie für die Eigenvektoren. Für die Eichfelder verwenden wir APE-Smearing [30] mit Gewichtung $\alpha = 0.25$.

Insgesamt wurden 64 stochastisch unabhängige Konfigurationen verwendet. Für jede werden 30 approximative Eigenmoden mit Hilfe der Setup-Phase des DD-αAMG-Lösers für Q berechnet. Die Inexaktheit beläuft sich bei diesen Eigenmoden auf etwa $\varepsilon \approx 10^{-1}$. Im Mittel werden durch $C = 0.75$ etwa drei der 30 Eigenmoden verworfen. Zum Vergleich und Verifizierung wurden ebenfalls die 20 kleinsten Eigenmoden mittels RQI+AMG (siehe Abschnitt 6.1) mit einer Fehlertoleranz $\varepsilon = 10^{-8}$ berechnet. Wir bezeichnen diese als „exakt“.

Abbildung 6.9 zeigt den verbundenen (Pion-) Korrelator und seinen relativen Fehler. Auf-

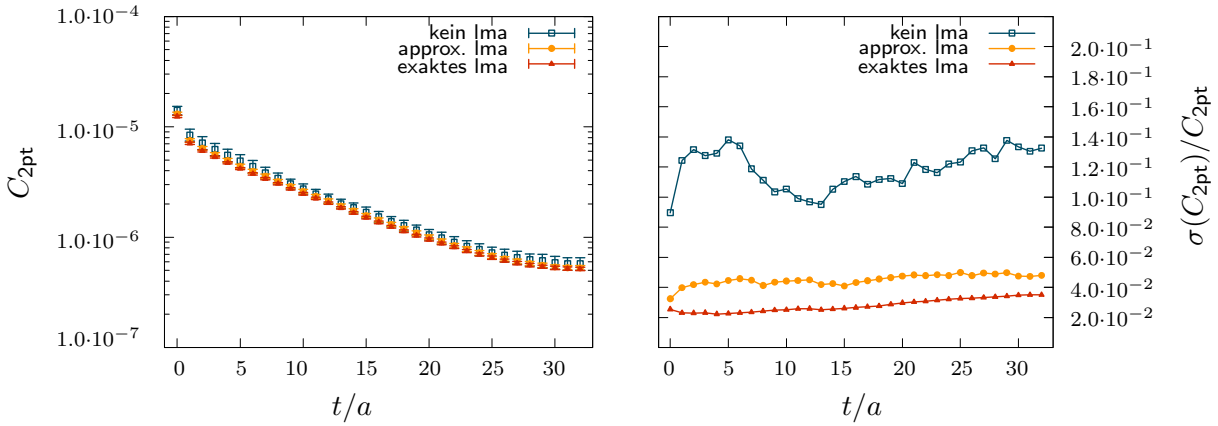


Abbildung 6.9: Der pseudoskalare verbundene (Pion-) Korrelator (links) und sein relativer Fehler zu jeder Zeitebene (rechts), berechnet mit exaktem (rote Dreiecke), approximativem (orangene Kreise) und ganz ohne (blaue Boxen) LMA.

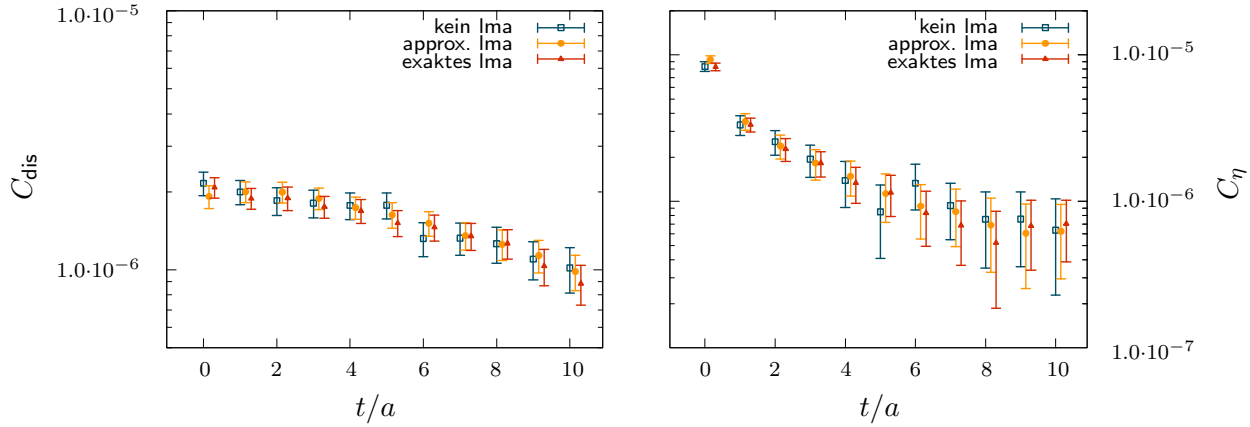


Abbildung 6.10: Der unverbundene Beitrag (links) und der volle η-Korrelator (rechts), berechnet mit exakten (rote Dreiecke), approximativen (orangene Kreise) und ohne (blaue Boxen) LMA. Für alle Berechnungen wurde $N_{\text{stoch}} = 20$ verwendet. (Die Datenpunkte sind zur besseren Lesbarkeit geringfügig horizontal verschoben.)

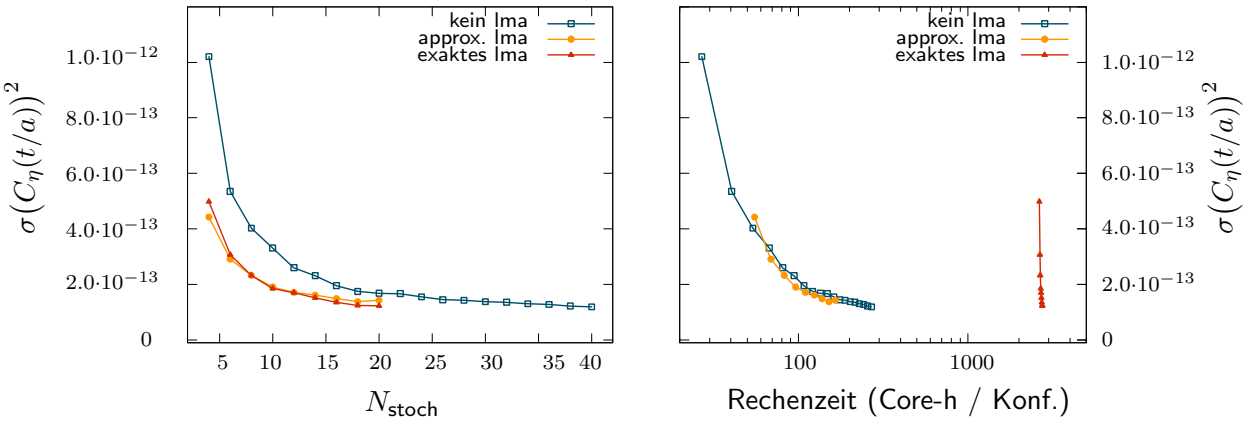


Abbildung 6.11: Der durchschnittliche quadratische Fehler der ersten zehn Zeitebenen des η-Korrelators berechnet mit exaktem (rote Dreiecke), approximativem (orangene Kreise) und ohne (blaue Boxen) LMA. Der linke Plot zeigt wie viele stochastische Schätzer N_{stoch} gebraucht werden, um eine gewisse Fehlerschranke zu erreichen (gemittelt über 64 stoch. unabh. Konfigurationen). Der rechte Plot vergleicht die tatsächlich anfallenden gesamt Rechenkosten.

grund der Mittelung über die Raumkoordinaten funktioniert LMA in diesem Fall sehr gut. Der verbundene Beitrag liefert erste Hinweise darauf, dass unsere Verbesserungen funktionieren: Die Fehler für das approximative LMA sind fast gleich wie bei der Verwendung von exakten Eigenmoden. Für den unverbundenen Anteil wurde in allen Fällen „time dilution“ [108, 81, 6] verwendet mit $\Delta t = 4a$. Wie in Abbildung 6.10 zu sehen, stimmen die Datenpunkte des approximativen LMA sowie exakten LMA mit denen überein, für die kein LMA verwendet wurde. Die blauen Werte in der Mitte der beiden Plots schwanken etwas stärker, hier scheint LMA glättend zu wirken. Um beim Kombinieren der verbundenen und unverbundenen Korrelator-Anteile den η-Korrelator C_η zu erhalten, akkumulieren sich die Fehler bei größeren Zeiteinheiten und die Daten zeigen nicht das zu erwartende exponentielle Abfallverhalten. Dies beruht vermutlich auf den zu kleinen Stichproben der gewählten Konfigurationen (vgl. auch [6]).

An den Plots erkennen wir, dass sowohl exaktes als auch approximatives LMA tendenziell Fehler reduzieren. Alternativ führt sicherlich das Erhöhen der Anzahl stochastischer Vektoren N_{stoch} zu ähnlichen Fehlertoleranzen ohne LMA zu benutzen. Die linke Abbildung 6.11 zeigt hierzu den quadratischen Fehler gemittelt über die ersten zehn Zeitabschnitte (danach erhöht sich das Rauschen rapide) in Abhängigkeit von N_{stoch}. Dieser Vergleich verdeutlicht den positiven Effekt von LMA: In allen Fällen sind ohne LMA knapp doppelt so viele Inversionen nötig um dieselbe Fehlertoleranz zu erreichen. Exaktes und approximatives LMA zeigen nahezu gleiches Verhalten. Schließlich zeigt der rechte Plot von Abbildung 6.11 den wohl interessantesten Teil: Berechnungszeit des η-Korrelators in Abhängigkeit zu gewissen Fehlerschranken. Hier zeigt sich, die benötigte Rechenzeit kann mit approximativem LMA um etwa einen Faktor zehn gegenüber exaktem LMA reduziert werden. Insbesondere ist approximatives LMA kosteneffizient und zuverlässig anwendbar bei größeren Gitterkonfigurationen, sogar dann, wenn nur eine kleine Anzahl verschiedener n-Punkt-Funktionen berechnet werden müssen. Es bleibt zu betonen, dass die Möglichkeiten des

LMA sowie die benötigte Anzahl an Eigenmoden stark von der untersuchten Observable, sowie noch stärker vom Volumen der Gitterkonfiguration und der Pionmasse abhängt.

Die Berechnungen in diesem Abschnitt wurden gemeinsam mit der Arbeitsgruppe G. Bali an der Universität Regensburg auf SuperMUC im Leibniz-Rechenzentrum in Garching ausgeführt und in [7] publiziert.

Anmerkungen[‡]

[23] John William Strutt, 3. Baron Rayleigh (* 12. November 1842 in Langford Grove, Maldon, England; † 30. Juni 1919 in Terlins Place bei Witham, England), war ein englischer Physiker. Er erhielt 1904 den Nobelpreis für Physik.

[24] Carl Gustav Jacob Jacobi (eigentlich Jacques Simon; * 10. Dezember 1804 in Potsdam; † 18. Februar 1851 in Berlin) war ein deutscher Mathematiker.

[25] Cornelius Lanczos (auch Kornél Lőwy, Kornél Lánczos; * 2. Februar 1893 in Székesfehérvár, Österreich-Ungarn; † 25. Juni 1974 in Budapest) war ein ungarischer Mathematiker und Physiker.

[26] Richard Edler von Mises (* 19. April 1883 in Lemberg, Galizien, Österreich-Ungarn; † 14. Juli 1953 in Boston, Massachusetts, Vereinigte Staaten) war ein österreichischer Mathematiker. Er ist der Bruder des Wirtschaftswissenschaftlers Ludwig von Mises.

[27] Helmut Wielandt (* 19. Dezember 1910 in Niedereggenen; † 14. Februar 2001 in Schliersee) war ein deutscher Mathematiker. Sein Hauptarbeitsgebiet war die Gruppentheorie, speziell die Theorie der Permutationsgruppen.

[28] Pafnuti Lwowitsch Tschebyscheff (wiss. Transliteration Pafnutij L'vovič Čebyšëv; * 4. (jul.)/ 16. Mai 1821 (greg.) in Okatowo im Kreis Borowsk (heute in der Oblast Kaluga); † am 26. November (jul.)/ 8. Dezember 1894 (greg.) in Sankt Petersburg) war ein russischer Mathematiker. Tschebyscheff gilt zusammen mit Nikolai Iwanowitsch Lobatschewski als der bedeutendste russische Mathematiker des 19. Jahrhunderts.

[29] Gian-Carlo Wick (* 15. Oktober 1909 in Turin; † 20. April 1992 ebenda) war ein italienischer Physiker, der wichtige Beiträge zur Quantenfeldtheorie leistete.

[‡]Alle Angaben aus der deutschen Wikipedia, Stand 2017

Abbildungsverzeichnis

Tabellenverzeichnis

Liste der Algorithmen

Literaturverzeichnis

[1] K. Aishima, *Global vonvergence of the restarted Lanczos and Jacobi-Davidson methods for symmetric eigenvalue problems.* Numer. Math. 131: 405-423, 2015.

[2] M. Albanese, F. Costantini, G. Fiorentini, F. Flore, M. P. Lombardo, P. Bacilieri R. Tripiccione, L. Fonti, E. Remiddi, M. Bernaschi, N. Cabibbo, L. A. Fernandez, E. Marinari, G. Parisi, G. Salina, S. Cabasino, F. Marzano, P. Paolucci, S. Petrarca, F. Rapuano, P. Marchesini, P. Giacomelli und R. Rusack, *Glueball masses and string tension in lattice QCD.* Phys. Lett. B192: 163–169, 1987.

[3] P. Arbenz und M. E. Hochstenbach, *A Jacobi-Davidson method for solving complex symmetric eigenvalue problems.* SIAM J. Sci. Comput. 25(5): 1655-1673, 2004.

[4] R. Babich, J. Brannick, R. C. Brower, M. A. Clark, T. A. Manteuffel, S. F. McCormick, J. C. Osborn und C. Rebbi, *Adaptive multigrid algorithm for the lattice Wilson-Dirac operator.* Phys. Rev. Lett. 105:201602, 2010.

[5] G. S. Bali, P. C. Bruns, S. Collins, M. Deka, B. Gläßle, M. Göckeler, L. Greil, T. R. Hemmert, R. Horsley, J. Najjar, Y. Nakamura, A. Nobile, D. Pleiter, P. E. L. Rakow, A. Schäfer, R. Schiel, G. Schierholz, A. Sternbeck und J. Zanotti, *Nucleon mass and sigma term from lattice QCD with two light fermion flavors.* Nucl. Phys. B866: 1–25, 2013.

[6] G. S. Bali, S. Collins, S. Dürr und I. Kanamori, $D_s \to \eta, \eta'$ *semileptonic decay form factors with disconnected quark loop contributions.* Phys. Rev. D91, 014503, 2015. arXiv:1406.5449

[7] G. Bali, S. Collins, A. Frommer, K. Kahl, I. Kanamori, B. Müller, M. Rottmann und J. Simeth, *(Approximate) low-mode averaging with a new multigrid eigensolver.* PoS, LATTICE2015: 350, 2015.

[8] G. S. Bali, S. Collins, B. Gläßle, M. Göckeler, J. Najjar, R. H. Rödl, A. Schäfer, R. W. Schiel, A. Sternbeck und W. Söldner, *Moments of structure functions for* $N_f = 2$ *near the physical point.* Phys. Rev. D90, 074510, 2014. arXiv:1408.6850

[9] G. S. Bali, S. Collins und A. Schäfer, *Effective noise reduction techniques for disconnected loops in Lattice QCD.* Comput. Phys. Commun. 181, 1570, 2010. arXiv:0910.3970

[10] G. S. Bali, H. Neff, T. Duessel, T. Lippert und K. Schilling (SESAM Kollaboration), *Observation of String Breaking in QCD.* Phys. Rev. D71, 114513, 2005. hep-lat/0505012

[11] S. Bernardson, P. McCarty und C. Thron, *Monte Carlo methods for estimating linear combinations of inverse matrix entries in lattice QCD* Comput. Phys. Commun. 78, 256, 1994.

[11] F. Berruto, R. Narayanan und H. Neuberger, *Exact local fermionic zero modes.* Phys. Lett. B489: 243–250, 2000.

[12] C. Bonati und M. D'Elia, *A comparison of the gradient flow with cooling in SU(3) pure gauge theory.* Phys. Rev., D89:105005, 2014.

[13] D. Braess, *Towards algebraic multigrid for elliptic problems of second order.* Computing, 55(4): 379–393, 1995.

[14] A. Brandt, *Multiscale scientific computation: Review 2001. In Multiscale and Multiresolution Methods.* Volume 20 of Lecture Notes in Computational Science and Engineering, 3–95, Springer Berlin Heidelberg, 2002.

[15] A. Brandt, J. Brannick, K. Kahl und I. Livshits, *Bootstrap AMG.* SIAM J. Sci. Comput. 33(2): 612–632, 2011.

[16] J. Brannick, A. Frommer, K. Kahl, B. Leder, M. Rottmann und A. Strebel, *Multigrid preconditioning for the overlap operator in lattice QCD.* Numer. Math., 2015.

[17] J. Brannick, R. C. Brower, M. A. Clark, J. C. Osborn und C. Rebbi, *Adaptive multigrid algorithm for lattice QCD.* Phys. Rev. Lett., 100:041601, 2007.

[18] M. Brezina, R. Falgout, S. MacLachlan, T. Manteuffel, S. McCormick, und J. Ruge, *Adaptive smoothed aggregation (αSA) multigrid.* SIAM Review 47: 317–346, 2005.

[19] R. Brower, E. Myers, C. Rebbi und K. Moriarty, *The multigrid method for fermion calculations in quantum chromodynamics.* Technical Report Print-87-0335, IAS, Princeton, 1987.

[20] S. Capitani, S. Durr und C. Hoelbling. *Rationale for UV-filtered clover fermions.* JHEP, 0611 028, 2006.

[21] M. A. Clark, Bálint Joó, Alexei Strelchenko, Michael Cheng, Arjun Gambhir und Richard C. Brower, *Accelerating Lattice QCD Multigrid on GPUs Using Fine-Grained Parallelization.*, 2016. arXiv:1612.07873v1

[22] E. R. Davidson, *The iterative calculation of a few of the lowest eigenvalues and corresponding eigenvectors of large real-symmetric matrices.* J. Comput. Phys. 17: 87-94, 1975.

[23] C. Davies, C. DeTar, C. McNeile und A. Vaquero, *Numerical experiments using deflation with the HISQ action.* 2017. arXiv:1710.07219v1

[24] T. DeGrand und C. E. Detar. *Lattice Methods for Quantum Chromodynamics.* World Scientific, 2006.

[25] T. A. DeGrand und S. Schaefer, *Improving meson two-point functions in lattice QCD.* Comput. Phys. Commun. 159, 185, 2004. hep-lat/0401011

[26] A. Duffy, *Creating and Using a Red-Black Matrix.* `http://computationalmathematics.org/topics/files/red_black.pdf`, 2010.

[27] S. Dürr, Z. Fodor, C. Hoelbling, S. D. Katz, S. Krieg, et al, *Lattice QCD at the physical point: Simulation and analysis details.* JHEP, 08 148, 2011.

[28] S. Dürr, Z. Fodor, C. Hoelbling, S. D. Katz, S. Krieg, T. Kurth, L. Lellouch, T. Lippert, K. K. Szabo und G. Vulvert, *Lattice QCD at the physical point: Light quark masses.* Phys. Lett. B701: 265–268, 2011.

[29] R. G. Edwards, U. M. Heller und T. R. Klassen, *Effectiveness of nonperturbative $\mathcal{O}(a)$ improvement in lattice QCD.* Phys. Rev. Lett. 80-3448, 1998. hep-lat/9711052.

[30] M. Falcioni, M. L. Paciello, G. Parisi und B. Taglienti, *Again On Su(3) Glueball Mass* Nucl. Phys. B251, 624, 1985.

[31] S. Feng und Z. Jia, *A refined Jacobi-Davidson method and its correction equation.* Comput. Math. Appl. 49(2-3): 417-427, 2005.

[32] J. Foley, K. Jimmy Juge, A. O'Cais, M. Peardon, S. M. Ryan und J.-I. Skullerud, *Practical all-to-all propagators for lattice QCD.* Comput. Phys. Commun. 172, 145, 2005. hep-lat/0505023

[33] A. Frommer, S. Güttel und M. Schweitzer, *Convergence of restarted Krylovsubspace methods for Stieltjes functions of matrices.* SIAM J. Matrix Anal. Appl. 35: 1602–1624, 2014.

[34] A. Frommer, S. Güttel und M. Schweitzer, *Efficient and stable Arnoldi restarts for matrix functions based on quadrature.* SIAM J. Matrix Anal. Appl. 35: 661–683, 2014.

[35] A. Frommer, K. Kahl, S. Krieg, B. Leder und M. Rottmann, *Adaptive aggregation based domain decomposition multigrid for the lattice Wilson Dirac operator.* SIAM J. Sci. Comput. 36(4): A1581–A1608, 2014.

[36] A. Frommer, K. Kahl, S. Krieg, B. Leder und M. Rottmann, *Aggregationbased multilevel methods for lattice QCD.* Proceedings of Science, http://pos.sissa.it, volume LATTICE2011:046, 2011.

[37] A. Frommer, K. Kahl, S. Krieg, B. Leder und M. Rottmann, *An adaptive aggregation based domain decomposition multilevel method for the lattice Wilson Dirac operator: Multilevel results.*, 2013. arXiv:1307.6101

[38] A. Frommer, A. Nobile und P. Zingler, *Deflation and flexible SAP-preconditioning of GMRES in lattice QCD simulation.*, 2012. arXiv:1204.5463

[39] C. Gattringer und C. B. Lang, *Quantum Chromodynamics on the Lattice.* Volume 788 of Lect. Notes Phys. Springer, 2009.

[40] L. Giusti, P. Hernandez, M. Laine, P. Weisz und H. Wittig, *Low-energy couplings of QCD from current correlators near the chiral limit.* JHEP 04, 013, 2004. hep-lat/0402002

[41] I. Gohberg, P. Lancaster und L. Rodman, *Indefinite Linear Algebra and Applications.* Birkhäuser, Basel, 2005.

[42] S. Güsken, U. Löw, K. H. Mütter, R. Sommer, A. Patel und K. Schilling, *Non-singlet axial vector couplings of the baryon octet in lattice QCD.* Phys. Lett. B 227(2): 266-269, 1989.

[43] E. Hairer, C. Lubich und G. Wanner, *Geometric Numerical Integration.* Volume 31 of Springer Series in Computational Mathematics. Springer, Heidelberg, 2010.

[44] M. Hanke-Bourgeois, *Grundlagen der Numerischen Mathematik und des Wissenschaftlichen Rechnens.* Vieweg+Teubner, Wiesbaden, 3. Auflage, 2009.

[45] A. Hasenfratz, R. Hoffmann und S. Schaefer, *Hypercubic smeared links for dynamical fermions.* JHEP, 0705-029, 2007.

[46] A. Hasenfratz, R. Hoffmann und S. Schaefer, *Localized eigenmodes of the overlap operator and their impact on the eigenvalue distribution.* JHEP, 0711-071, 2007.

[47] M. Hestenes und E. Stiefel, *Methods of conjugate gradients for solving linear systems.* Journal of Research of the National Bureau of Standards 49: 409–436, 1952. doi:10.6028/jres.049.044

[48] M. E. Hochstenbach, *Jacobi-Davidson Gateway.* Webseite: `http://www.win.tue.nl/casa/research/scientificcomputing/topics/jd/`, 2014.

[49] C. G. J. Jacobi, *Über eine neue Auflösungsart der bei der Methode der kleinsten Quadrate vorkommende linearen Gleichungen.* Astronom. Nachr., 297-306, 1845.

[50] C. G. J. Jacobi, *Über ein leichtes Verfahren, die in der Theorie der Säcularstörungen vorkommenden Gleichungen numerisch aufzulösen.* J. Reine und Angew. Math., 51-94, 1846.

[51] K. Jansen und R. Sommer (ALPHA), $\mathcal{O}(a)$ *improvement of lattice QCD with two flavors of Wilson quarks.* Nucl. Phys. B530-185, 1998. hep-lat/9803017

[52] T. Kalkreuter, *Multigrid methods for propagators in lattice gauge theories.* J. Comput. Appl. Math. 63: 57–68, 1995.

[53] S. Krieg und T. Lippert, *Tuning lattice QCD to petascale on Blue Gene/P.* NIC Symposium 2010, 155–164, 2010.

[54] A. N. Krylov, *О численном решении уравнения, которым в технических вопросах определяются частоты малых колебаний материальных систем.* Izvestija AN SSSR (News of Academy of Sciences of the USSR), Otdel. mat. i estest. nauk, 1931, VII, Nr.4,

491-539 (auf Russisch). Übersetzung: *On the Numerical Solution of Equation by Which are Determined in Technical Problems the Frequencies of Small Vibrations of Material Systems*, oder *On the numerical solution of the equation by which in technical questions frequencies of small oscillations of material systems are determined.* Nach Grigorian (2008) bzw. Botchev (2002).

[55] H. B. Lawson und M. L. Michelsohn, *Spin Geometry.* Princeton Mathematical Series, Princeton Mathematical Series, Princeton University Press, 1989.

[56] N. N. Lebedev, *Special Functions and Their Applications.* Prentice-Hall, Englewood Cliffs, 1965.

[57] M. Lüscher, *DD-HMC algorithm for two-flavour lattice QCD.* Version: DD-HMC-1.2.2, September 2008. `http://luscher.web.cern.ch/luscher/DD-HMC`,

[58] M. Lüscher, *Deflation acceleration of lattice QCD simulations.* JHEP 12, 011, 0710.5417, 2007.

[59] M. Lüscher, *Exact chiral symmetry on the lattice and the Ginsparg-Wilsonrelation.* Phys. Lett., B428:342–345, 1998.

[60] M. Lüscher, *Lattice QCD and the Schwarz alternating procedure.* CERN-TH/2003-088, 2003.

[61] M. Lüscher, *Local coherence and deflation of the low quark modes in lattice QCD.* JHEP 07, 081, 0706.2298, 2007.

[62] M. Lüscher, *openQCD simulation program for lattice QCD with open boundary conditions.* Version: openQCD-1.4, 2013. `http://luscher.web.cern.ch/luscher/openQCD/`,

[63] M. Lüscher, *Properties and uses of the Wilson flow in lattice QCD.* JHEP, 1008 071, 2010.

[64] M. Lüscher, *Schwarz-preconditioned HMC algorithm for two-flavour lattice QCD.* CERN-PH-TH/2004-177, 2004.

[65] M. Lüscher, *Solution of the Dirac equation in lattice QCD using a domain decomposition method.* Comput. Phys. Commun. 156: 209–220, 2004. hep-lat/0310048v1

[66] M. Lüscher, *Trivializing maps, the Wilson flow and the HMC algorithm.* Commun. Math. Phys. 293: 899–919, 2010.

[67] M. Lüscher, S. Sint, R. Sommer, P. Weisz und U. Wolff, *Non-perturbative $\mathcal{O}(a)$ improvement of lattice QCD.* Nucl. Phys. B491-323, 1997. hep-lat/9609035

[68] A. Meister, *Numerik linearer Gleichungssysteme.* Vieweg+Teubner, Wiesbaden, 4. Auflage, 2011.

[69] I. Montvay und G. Münster, *Quantum Fields on a Lattice.* Cambridge Monographs on Mathematical Physics. Cambridge University Press, 1994.

[70] R. B. Morgan, *GMRES with defltated restarting.* SIAM J. Sci. Comput. 24(1): 20–37, 2002.

[71] K. Morikuni, L. Reichel und K. Hayami, *FGMRES for linear discrete ill-posed problems.* APNUM, 2013.

[72] C. Morningstar und M. J. Peardon, *Analytic smearing of SU(3) link variables in lattice QCD.* Phys. Rev., D69:054501, 2004.

[73] H. Neff, N. Eicker, Th. Lippert, J. W. Negele und K. Schilling, *On the low fermionic eigenmode dominance in QCD on the lattice.* Phys. Rev. D64, 114509, 2001. hep-lat/0106016

[74] J. W. Negele, *Instantons, the QCD vacuum, and hadronic physics.* Nucl. Phys. Proc. Suppl. 73: 92–104, 1999.

[75] H. Neuberger, *Bounds on the Wilson Dirac operator.* Phys. Rev., D61:085015, 2000.

[76] F. Niedermayer, *Exact chiral symmetry, topological charge and related topics.* Nucl. Phys. Proc. Suppl. 73: 105–119, 1999.

[77] H. B. Nielsen und M. Ninomiya, *Absence of neutrinos on a lattice (I). Proof by homotopy theory.* Nucl. Phys. B185-20, 1981.

[78] H. B. Nielsen und M. Ninomiya, *Absence of neutrinos on a lattice. (II). Intuitive topological proof*, Nucl. Phys. B193-173, 1981.

[79] Y. Notay und P. S. Vassilevski, *Recursive Krylov-based multigrid cycles.* Numerical Linear Algebra with Applications 15: 473–487, 2008.

[80] Y. Notay, *Combination of Jacobi-Davidson and conjugate gradients for the partial symmetric eigenproblem.* Numer. Lin. Alg. Appl. 9: 21-44, 2002.

[81] A. O'Cais, K. J. Juge, M. J. Peardon, S. M. Ryan und J.-I. Skullerud (TrinLat), *Improving Algorithms to Compute All Elements of the Lattice Quark Propagator.* Lattice field theory: Proceedings, 22nd International Symposium: 844–849, Lattice 2004, Batavia, USA, 2004. hep-lat/0409069, `http://dx.doi.org/10.1016/j.nuclphysbps.2004.11.286`

[81] J. C. Osborn, R. Babich, J. Brannick, R. C. Brower, M. A. Clark, S. D. Cohen und C. Rebbi, *Multigrid solver for clover fermions.* PoS, LATTICE2010:037, 1011.2775, 2010.

[82] J. C. Osborn, *Multigrid solver for clover fermions, implementation within QOPQDP.* Version: QOPQDP 0.19.0, 2013.
`http://usqcd.jlab.org/usqcd-docs/qopqdp/`

[83] B. N. Parlett, *The symmetric eigenvalue problem.* SIAM, Classics in Applied Mathematics, 1998.

[84] C. C. Paige und M. A. Saunders, *Solution of sparse indefinite systems of linear equations.* SIAM Journal on Numerical Analysis. 12(4), 1975.

[85] M. Peardon et al., *A novel quark-field creation operator construction for hadronic physics in lattice QCD.* Phys. Rev., D80:054506, 2009. arXiv:0905.2160

[86] M. E. Peskin und D. V. Schroeder, *An Introduction To Quantum Field Theory (Frontiers in Physics).* Advanced Book Program. Westview Press, Boulder Colorado, 1995.

[87] PRACE Annual Report 2016.
`http://www.prace-ri.eu/IMG/pdf/PRACE2016Annual-Report_2017_LOWRES.pdf`

[88] R. J. Radke, *A Matlab implementation of the implicitly restarted Arnoldi method for solving large-scale eigenvalue problems.* Dissertation, Rice University, 1996.

[89] M. Rottmann, *Adaptive Dopmain Decomposition Multigrid for Lattice QCD.* Dissertation, Bergische Universtität Wuppertal, urn:nbn:de:hbz:468-20160304-125225-9, 2016.
`http://nbn-resolving.de/urn/resolver.pl?urn=urn%3Anbn%3Ade%3Ahbz%3A468-20160304-125225-9`

[90] Y. Saad, *Iterative Methods for Sparse Linear Systems.* SIAM, Philadelphia, PA, USA, 2nd edition, 2003.

[91] Y. Saad und M. H. Schultz, *GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems.* SIAM J. Sci. Comput. 7: 856-869, 1986.

[92] Y. Saad, *A flexible inner-outer preconditioned GMRES algorithm.* SIAM J. Sci. Comput. 14: 461–469, 1993.

[93] Y. Saad, *Numerical Methods for Large Eigenvalue Eigenvalue Problems: Revised Edition.* Classics in Applied Mathematics SIAM, 2011.

[94] H. Schwarz, *Gesammelte mathematische Abhandlungen.* Vierteljahrschrift Naturforsch. Ges. Zürich 272–286, 1870.

[95] H. Schwarz, *Über einen Grenzübergang durch alternierendes Verfahren.* Vierteljahrsschrift der Naturforschenden Gesellschaft in Zürich, 15: 272–286, JFM 02.0214.02, 1870.

[96] B. Sheikholeslami und R. Wohlert, *Improved continuum limit lattice action for QCD with Wilson fermions.* Nucl. Phys. B259-572, 1985.

[97] G. L. G. Sleijpen, H. A. van der Vorst, *A Jacobi-Davidson iteration method for linear eigenvalue problems.* SIAM J. Matrix Anal. Appl. 17(2): 401-425, 1996.

[98] G. L. G. Sleijpen und J. van den Eshof, *On the use of harmonic Ritz pairs in approximating internal eigenpairs.* Lin. Alg. Appl. 358(1-3): 115-137, 2003.

[99] B. F. Smith, P. E. Bjørstad und W. D. Gropp, *Domain decomposition: Parallel Multilevel methods for elliptic partial differential equations.* Cambridge University Press, New York, 1996.

[100] D. Sorensen, R. Lehoucq, C. Yang und K. Maschhoff, *PARPACK.* Version: 2.1, September 1996. `http://http://www.caam.rice.edu/software/ARPACK`

[88] A. Stathopoulos, Y. Saad, *Restarting techniques for the (Jacobi-)Davidson symmetric eigenvalue methods.* Electron. Trans. Numer. Anal. 7: 163-181, 1998.

[102] A. Stathopoulos, *Nearly optimal preconditioned methods for hermitian eigenproblems under limited memory. Part II: Seeking many eigenvalues.* SIAM J. Sci. Comput. 29(5): 2162-2188, 2007.

[103] A. Stathopoulos und J. R. McCombs, *PRIMME: PReconditioned Iterative MultiMethod Eigensolver: Methods and software description.* ACM Transaction on Mathematical Software 37(2): 21:1-21:30, 2010.

[104] D. B. Szyld und J. A. Vogel, *FQMR: A flexible quasi-minimal residual method with inexact preconditioning*, SIAM J. Sci. Comput. 23(2): 363-380, 2001.

[105] P. T. P. Tang und E. Polizzi, *FEAST as a subspace iteration eigensolver accelerated by approximate spectral projection.* SIAM Journal on Matrix Analysis and Applications (SIMAX) 35: 354-390, 2014.

[106] H. A. van der Vorst, *Bi-CGSTAB: A fast and smoothly converging variant of Bi-CG for the solution of nonsymmetric linear systems.* SIAM J. Sci. Stat. Comput. 13: 631-644, 1992.

[107] M. B. van Gijzen, G. L. G. Sleijpen und J.-P. M. Zemke, *Flexible and multi-shift induced dimension reduction algorithms for solving large sparse linear systems.* Report 11-06, DIAM Delf, 2011.

[108] J. Viehoff, N. Eicker, S. Güsken, H. Hoeber, P. Lacock, T. Lippert, K. Schilling, A. Spitz und P. Überholz (TXL), *Improving stochastic estimator techniques for disconnected diagrams.* Nucl. Phys. Proc. Suppl. 63, 269, 1998. hep-lat/9710050.

[108] J. A. Vogel, *Flexible BiCG and flexible Bi-CGSTAB for nonsymmetric linear systems.* Applied Mathematics and Computation 188: 226-233, 2007.

[109] H. Voss, *A new justification of the Jacobi–Davidson method for large eigenproblems.* Linear Algebra and its Applications 424: 448–455, 2007.

[110] K. G. Wilson, *Confinement of quarks*, Phys. Rev. D10:2445–2459, 1974.

[111] K. G. Wilson, *Quarks and strings on a lattice.* In A. Zichichi, editor, New Phenomena in Subnuclear Physics. Part A. Proceedings of the First Half of the 1975 International School of Subnuclear Physics, Erice, Sicily, July 11 - August 1, 1975, volume 321 of CLNS Reports, pages 69–138, New York, 1977. Plenum Press.

[112] R. Wohlert, *Improved continuum limit lattice action for quarks*, DESY 87-069, 1987.

[113] L. Wu, E. Romero und A. Stathopoulos, *PRIMME_SVDS: A high-performance preconditioned SVD solver for accurate large-scale computations.*, 2016. arXiv:1607.01404

[114] Y. Zhou und Y. Saad, *A Chebyshev–Davidson algorithm for large symmetric eigenproblems.* SIAM Journal on Matrix Analysis and Applications 29(3): 954-971, 2007.

[115] Y. Zhou, Y. Saad, M. L. Tiago, J. R. Chelikowsky, *Self-consistent-field calculations using Chebyshev-filtered subspace iteration.* Journal of Computational Physics 219: 172–184, 2006.