Kurze Rekursionen bei präkonditionierten CKS-Verfahren

Diplomarbeit
von
Boris Wagner

9. Dezember 1994

Angefertigt am
Rechenzentrum
der Universität Karlsruhe

Referent:	Prof. Dr. Willi Schönauer
Korreferent:	Prof. Dr. Wilhelm Niethammer

Abstract

Ziel dieser Arbeit ist es, bei schrittabhängig präkonditionierten CKS-Verfahren durch geeignete Wahl der Präkonditionierungsmatrizen ein Verfahren zu konstruieren, das eine kurze Rekursion hat *und* eine Minimierungseigenschaft in der A^{-1}-Norm erfüllt.

Dieses Ziel wird auch erreicht. Es werden zwei Verfahren angegeben, die eine zweistufige Rekursion haben und die im k-ten Schritt eine Minimierungseigenschaft für das Residuum in der A^{-1}-Norm besitzen.

Gleichzeitig sind die zwei Verfahren aber auch äquivalent zu zwei speziellen *abgeschnittenen* verallgemeinerten cg-Verfahren. Damit ist es zum ersten Mal möglich, für abgeschnittene cg-Verfahren eine Minimierungseigenschaft über einem k-dimensionalen Raum anzugeben.

Es wird gezeigt, daß die konstruierten Verfahren nicht nur von theoretischem Interesse sind, sondern sich auch in der Praxis bewähren.

Introduction

In numerischen Anwendungen stellt sich oft die Aufgabe, ein lineares Gleichungssystem

$$Ax = b$$

näherungweise zu lösen. Falls die zugehörige Matrix A sehr groß und dünn besetzt ist, bieten sich iterative Verfahren zur Lösung dieser Aufgabe an. Eine große Anzahl dieser sogenannten *iterativen Gleichungslöser* sind *Methoden der konjugierten Krylowunterräume* (CKS[1], siehe [11]). Bekannte Beispiele sind das Verfahren des steilsten Abstiegs, das klassische cg-Verfahren, ORTHORES, GMRES, BCG, QMR und CGS.

Das klassische cg-Verfahren besitzt für symmetrisch positiv definite Matrizen A die beiden folgenden Eigenschaften.

1. Das Verfahren läßt sich mit einer kurzen Rekursion schreiben, d. h. es wird nur eine beschränkte Anzahl von zurückliegenden Iterationsvektoren für die Berechnung der neuen Iterierten benötigt.

2. Das Verfahren erfüllt im k-ten Schritt für das k-te Residuum $r_k = Ax_k - b$ eine Minimierungseigenschaft in der A^{-1}-Norm über dem Krylowraum $K_{k-1}(A, r_0) = \text{span}\{r_0, Ar_0, \ldots, A^{k-1}r_0\}$.

Das Ziel der Forschung ist seither, ein Verfahren zu konstruieren, das auch für nichtsymmetrische Matrizen diese beiden oder ähnliche Eigenschaften besitzt. Dabei sind folgende Resultate erzielt worden.

- Faber und Manteuffel [3, 4] zeigen, daß es für nichtsymmetrische Matrizen kein verallgemeinertes cg-Verfahren[2] vom Typ ORTHORES, ORTHOMIN oder ORTHODIR gibt, das beide Anforderungen erfüllt.

- Für CKS-Verfahren läßt sich eine kurze Rekursion durch geeignete Wahl der Orthogonalitätsmatrizen Z_k erzwingen (siehe Weiß [11]). Dadurch erhält man dann Abschätzungen in der Z_kA^{-1}-Norm mit einer schrittabhängigen Matrix Z_k.

[1]CKS = *conjugate Krylov subspace*

[2]Die verallgemeinerten cg-Verfahren sind eine Unterklasse der CKS-Verfahren.

- Barth und Manteuffel [1, 2] beweisen für die CKS-Verfahren BCG und QMR, die beide eine kurze Rekursion haben, daß sie eine Minimierungseigenschaft in der ZA^{-1}-Norm besitzen. Die Matrix Z hängt zwar nicht vom Iterationsindex k, dafür aber vom Startresiduum $r_0 = Ax_0 - b$ ab.

Inhaltsverzeichnis

Einleitung

Bei vielen numerischen Anwendungen auf Rechnern kommen lineare Gleichungssysteme der Form

$$Ax = b$$

vor, wobei die Matrix $A \in \mathbb{R}^{n \times n}$ und der Vektor $b \in \mathbb{R}^n$ gegeben sind und $x \in \mathbb{R}^n$ der gesuchte Vektor ist. Die Matrix A muß nicht symmetrisch oder positiv definit sein, aber sie wird als regulär vorausgesetzt.

Hier interessieren sehr große lineare Gleichungssysteme, bei denen die Matrix A dünn besetzt ist, d. h. nur ein geringer Anteil der Matrixelemente ist ungleich Null. Solche linearen Gleichungssysteme erhält man z. B. als Zwischenschritte bei der numerischen Lösung von partiellen Differentialgleichungen. Direkte Lösungsmethoden für lineare Gleichungssysteme wie etwa der Gauß-Algorithmus sind hier ausgeschlossen, denn sie verursachen ein Auffüllen der Matrix, wodurch der Speicher- und Rechenzeitbedarf über die Möglichkeiten der Rechner wächst. Da im allgemeinen auch nur eine Näherungslösung gesucht ist, kann man sich mit Iterationsverfahren begnügen. Wenn das verwendete Verfahren eine Fehlerschranke liefert, so muß man die Iteration nur so lange fortführen, bis die gewünschte Genauigkeit erreicht ist.

Die einzigen Operationen, die die hier behandelten Iterationsverfahren mit der Matrix A ausführen, sind Matrix-Vektor-Multiplikationen. Wenn die Matrix Diagonalgestalt hat, so lassen sich diese auf Parallel- und Vektorrechnern sehr effizient ausführen (siehe [13]). Das ist ein sehr wichtiger Aspekt für die Alltagstauglichkeit, weswegen sehr viel Forschung betrieben wird, um Iterationsverfahren, die nicht nur auf Matrix-Vektor-Multiplikationen beruhen, für die heutigen Supercomputer zu optimieren. Als Beispiel seien die Arbeiten von Fujino [5, 6] erwähnt, wo die ICCG-Methode[3] durch aufwendige Modifikationen für Vektorrechner optimiert wird. Bei den hier behandelten cg-Verfahren ist durch die Beschränkung auf die Matrix-Vektor-Multiplikation als existentielle Operation ein solcher Aufwand nicht nötig. Man muß lediglich

[3] ICCG = *incomplete Cholesky conjugate gradient.* Dieses Verfahren basiert auf der unvollständigen Cholesky-Zerlegung.

noch sicherstellen, daß der Algorithmus nur einen beschränkten Speicherbedarf hat. Verfahren, bei denen der Speicherbedarf linear in der Anzahl der Iterationen wächst, sind in der Praxis nur beschränkt tauglich, da man oft eine große Zahl von Iterationen braucht, um eine ausreichende Genauigkeit zu erreichen. Aus diesem Grund wird bei den hier behandelten Orthogonalisierungsverfahren versucht, kurze Rekursionen zu erreichen, d. h. nur eine beschränkte Anzahl von alten Iterierten zur Berechnung der neuen zu verwenden.

Notation

Die folgenden Notationen werden in dieser Arbeit verwendet: für positiv definite bzw. für positiv reelle Matrizen $Z \in \mathbb{R}^{n\times n}$ sei $\|y\|_Z$ für beliebige Vektoren $y \in \mathbb{R}^n$ definiert als $\|y\|_Z = \sqrt{y^T Z y}$. $K_k(B, y) = \operatorname{span}\{y, By, \ldots, B^k y\}$ sei der durch die Matrix B und den Vektor y aufgespannte Krylowraum. Der symmetrische Anteil einer quadratischen Matrix B ist $\frac{1}{2}(B + B^T)$, der schiefsymmetrische Anteil ist $\frac{1}{2}(B - B^T)$. Mit $\mathrm{GL}_n(\mathbb{R})$ ist die „general linear group“ der Dimension n über $\mathbb{R}$ gemeint, d. h. die Gruppe der invertierbaren rellen $(n \times n)$-Matrizen.

Kapitel 1

Präkonditionierte CKS-Verfahren

Wir wenden uns nun wieder schrittabhängig präkonditionierten CKS-Verfahren zu. Die im CKS-Verfahren bereits vorgesehene konstante rechtsseitige Präkonditionierungsmatrix P wird hier nicht verwendet, denn sie kann auf die $P_{R,k}$ aufmultipliziert werden.

Definition 1 *Ein* präkonditioniertes CKS-Verfahren *(pCKS) ist ein CKS-Verfahren, welches mit schrittabhängigen links- bzw. rechtsseitigen Präkonditionierungsmatrizen $P_{L,k}$ bzw. $P_{R,k}$ präkonditioniert wird.*

Wir verwenden nun die Ergebnisse aus Kapitel **??**, indem wir die präkonditionierten CKS-Verfahren als Spezialfall der Orthogonalisierungsverfahren darstellen.

Wie bereits in der Bemerkung nach der Definition der CKS-Verfahren auf Seite **??** erwähnt, gibt es für die Iterierten x_k und die zugehörigen Residuen r_k und Fehler e_k von CKS-Verfahren verschiedene äquivalente Darstellungen, für die jede durch geeignete Wahl von Z_k eine kurze Rekursion erzwungen werden kann. Für zwei dieser Darstellungen wollen wir im folgenden versuchen, exakte Verfahren mit kurzer Rekursion zu konstruieren.

Korollar 1
Für jedes $k \geq 1$ und jedes feste $l_k \in \{1, \ldots, k\}$ hat man für die Vektoren x_k, r_k bzw. e_k eines CKS-Verfahrens folgende zwei Arten von Darstellungen:

$$x_k = \gamma_{1,k} P r_{k-1} + \sum_{i=1}^{k-1} \gamma_{i+1,k}(x_{k-i} - x_{k-i-1}) + x_{k-l_k}\,, \tag{1.1a}$$

$$r_k = \gamma_{1,k} A P r_{k-1} + \sum_{i=1}^{k-1} \gamma_{i+1,k}(r_{k-i} - r_{k-i-1}) + r_{k-l_k}\,, \tag{1.1b}$$

$$e_k = \gamma_{1,k} PAe_{k-1} + \sum_{i=1}^{k-1} \gamma_{i+1,k}(e_{k-i} - e_{k-i-1}) + e_{k-l_k} \tag{1.1c}$$

und

$$x_k = \sum_{i=1}^{k} \mu_{i,k} Pr_{k-i} + x_{k-l_k}\,, \tag{1.2a}$$

$$r_k = \sum_{i=1}^{k} \mu_{i,k} APr_{k-i} + r_{k-l_k}\,, \tag{1.2b}$$

$$e_k = \sum_{i=1}^{k} \mu_{i,k} PAe_{k-i} + e_{k-l_k}\,. \tag{1.2c}$$

Beweis. Die Darstellungen (1.1a) bzw. (1.2a) für x_k unterscheiden sich von den Darstellungen (**??**) bzw. (**??**) aus Satz **??** nur durch eine andere Wahl des Fußpunktes. Die Äquivalenz folgt somit wegen $l_k \geq 1$ aus

$$x_{k-l_k} \overset{(\mathbf{??})}{\in} x_0 + K_{k-l_k-1}(PA, Pr_0) \subseteq x_0 + K_{k-1}(PA, Pr_0)\,. \tag{1.3}$$

Die Darstellungen von r_k bzw. e_k ergeben sich aus denen von x_k mittels der jeweiligen Definitionen. □

Definition 2 *Wir definieren als* Verfahren 1 *bzw.* Verfahren 2 *diejenigen auf das ursprüngliche System zurückformulierten schrittabhängig präkonditionierten CKS-Verfahren, die die Iterierten y_k des präkonditionierten Systems aus dem Ansatz (1.1) bzw. (1.2), jeweils mit $P = I$, berechnen.*

CKS-Verfahren sind spezielle exakte Orthogonalisierungsverfahren. Für diese haben wir im Kapitel **??** gezeigt, wie präkonditionierte Verfahren auf das ursprüngliche System zurückformuliert werden können.

Für die präkonditionierten Verfahren 1 und 2 leiten wir nun aus Satz **??** das folgende Korollar her.

Korollar 2
Das Verfahren 2 ist äquivalent zu einem exakten Orthogonalisierungsverfahren mit

$$q_{k-i,k} = P_{R,k} P_{L,k} r_{k-i}\,, \tag{1.4}$$

$$\tilde{x}_k = x_{k-l_k}\,, \tag{1.5}$$

$$Z_k = P_{L,k}^T \tilde{Z}_k P_{R,k}^{-1} \tag{1.6}$$

für $i = 1, \ldots, k$. In beiden Fällen sind $P_{L,k}, P_{R,k}, \tilde{Z}_k \in \mathbb{R}^{n\times n}$ beliebige, nichtsinguläre Matrizen.

BEWEIS. Trivial. □

Aus Korollar 2 ergeben sich die folgenden Darstellungen für die Iterationsvektoren von Verfahren 2.

Satz 1
Die Iterierten x_k, die Residuen r_k und die Fehler e_k des Verfahrens 2 haben die Darstellung

$$x_k = x_{k-l_k} + \sum_{i=1}^{k} \mu_{i,k} P_{R,k} P_{L,k} r_{k-i} \,, \tag{1.7a}$$

$$r_k = r_{k-l_k} + \sum_{i=1}^{k} \mu_{i,k} A P_{R,k} P_{L,k} r_{k-i} \,, \tag{1.7b}$$

$$e_k = e_{k-l_k} + \sum_{i=1}^{k} \mu_{i,k} P_{R,k} P_{L,k} A e_{k-i} \,. \tag{1.7c}$$

Die $\mu_{i,k}$ ergeben sich als Lösung des Gleichungssystems

$$r_k^T P_{L,k}^T \tilde{Z}_k P_{L,k} r_{k-i} = 0 \tag{1.8}$$

für $i = 1, \ldots, k$.

BEWEIS. Die Darstellungen von x_k, r_k bzw. e_k folgen aus den Aussagen von Korollar 2 zusammen mit (**??**). Die Orthogonalitätsbedingung (1.8) folgt aus Korollar 2 zusammen mit (**??**). □

Wie man an diesem Satz ablesen kann, unterscheiden sich die Darstellungen für die Iterierten x_k für die beiden Verfahren ab dem zweiten Schritt. So gilt z. B. für $k = 2$ für Verfahren 1

$$\begin{aligned} x_2 &= x_1 + \gamma_{1,2} P_{R,2} P_{L,2} r_1 + \gamma_{2,2}(x_1 - x_0) \\ &= x_1 + \gamma_{1,2} P_{R,2} P_{L,2} r_1 + \gamma_{2,2} \gamma_{1,1} P_{R,1} P_{L,1} r_0 \end{aligned}$$

und für Verfahren 2

$$x_2 = x_1 + \mu_{1,2} P_{R,2} P_{L,2} r_1 + \mu_{2,2} P_{R,2} P_{L,2} r_0$$

Da die Präkonditionierungsmatrizen $P_{R,k}$ bzw. $P_{L,k}$ beliebig gewählt werden können, sind das i. a. verschiedene Ansätze.

Dies ist insofern verwunderlich, als doch die beiden Ansätze (1.1) bzw. (1.2), die für die Iterierten y_k des präkonditionierten Systems von den Verfahren 1 bzw. 2 verwendet werden, für CKS-Verfahren identisch sind. Der Grund liegt darin, daß die Iterierten $y_{k-i}^{P_k}$ und die zugehörigen Residuen $r_{k-i}^{P_k}$, die im

k-ten Schritt zur Berechnung von y_k verwendet werden, nicht aus einem CKS-Verfahren herrühren, sondern nur die auf das k-te Gleichungssystem „umgerechneten" Iterationsvektoren von zurückliegenden Iterationsschritten auf jeweils verschiedenen Gleichungssystemen sind. Die „umgerechneten" Vektoren $y_{k-i}^{P_k}$ und $r_{k-i}^{P_k}$ genügen nicht mehr allen verschiedenen Darstellungen aus (**??**), die für CKS-Verfahren äquivalent sind.

1.1 Schrittunabhängige Normen

An den Aussagen von Satz **??** stören die schrittabhängigen Normen. Um Abschätzungen in einer festen Norm zu erhalten, wählt man im folgenden die Präkonditionierungsmatrizen $P_{L,k}$ bzw. $P_{R,k}$ geeignet.

Definition 3 *Wir definieren als* Verfahren 1^* *bzw.* 2^* *das Verfahren 1 bzw. 2 mit der speziellen Wahl der Präkonditionierungsmatrizen*

$$P_{L,k} = I \quad \textit{und} \tag{1.9a}$$
$$P_{R,k} = \tilde{Z}_k\,. \tag{1.9b}$$

Wir leiten aus den Sätzen 1 und **??** die entsprechenden Aussagen für das Verfahren 2^* her.

Korollar 3
Die Iterierten x_k und die Residuen r_k des Verfahrens 2^* haben die Darstellung

$$x_k = x_{k-l_k} + \sum_{i=1}^{k} \mu_{i,k} \tilde{Z}_k r_{k-i}\,, \tag{1.10}$$

$$r_k = r_{k-l_k} + \sum_{i=1}^{k} \mu_{i,k} A \tilde{Z}_k r_{k-i}\,. \tag{1.11}$$

Die Koeffizienten $\mu_{i,k}$ ergeben sich aus

$$r_k^T \tilde{Z}_k r_{k-i} = 0 \tag{1.12}$$

für $i = 1, \ldots, k$.

Aus Satz **??** erhält man für das Verfahren 2^* das folgende Korollar.

Korollar 4
*Falls A positiv reell ist, d. h. der symmetrische Anteil von A positiv definit ist, so gilt für Verfahren 2**

$$\|r_k\|_{A^{-1}} \leq \sqrt{\frac{1+\varrho^2(R)}{\mu_m^2}} \min_{\eta_1,\dots,\eta_k} \left\| r_{k-l_k} + \sum_{i=1}^{k} \eta_i A \tilde{Z}_k r_{k-i} \right\|_{A^{-1}} ,$$

$$\|e_k\|_{A^T} \leq \sqrt{\frac{1+\varrho^2(R)}{\mu_m^2}} \min_{\eta_1,\dots,\eta_k} \left\| e_{k-l_k} + \sum_{i=1}^{k} \eta_i \tilde{Z}_k A e_{k-i} \right\|_{A^T} .$$

Dabei ist $\varrho(R)$ der Spektralradius des schiefsymmetrischen Anteils R von A^{-1}. μ_m ist der kleinste Eigenwert von M, dem symmetrischen Anteil von A^{-1}. Falls insbesondere A symmetrisch ist, so gilt

$$\|r_k\|_{A^{-1}} = \min_{\eta_1,\dots,\eta_k} \left\| r_{k-l_k} + \sum_{i=1}^{k} \eta_i A \tilde{Z}_k r_{k-i} \right\|_{A^{-1}} ,$$

$$\|e_k\|_{A^T} = \min_{\eta_1,\dots,\eta_k} \left\| e_{k-l_k} + \sum_{i=1}^{k} \eta_i \tilde{Z}_k A e_{k-i} \right\|_{A^T} .$$

Dieses Korollar liefert Abschätzungen für $\|r_k\|_{A^{-1}}$ und $\|e_k\|_{A^T}$. Man hat also mit Verfahren 2^* ein exaktes Verfahren konstruiert, das für positiv reelle Matrizen A im k-ten Schritt für r_k eine Minimierungseigenschaft in einer festen Norm über einem k-dimensionalen affinen Raum erfüllt.

Das gilt aber für viele exakte Verfahren. Was man hier gewonnen hat, ist die freie Wahl der Matrix $\tilde{Z}_k$. Mit ihrer Hilfe kann man nun versuchen, ein exaktes Verfahren mit kurzer Rekursion aufzustellen.

Kapitel 2

Kurze Rekursionen

In [12] wird im Theorem 9 eine sehr allgemeine hinreichende Bedingung für l-stufige Rekursionen bei Orthogonalisierungsverfahren angegeben. In unserem Fall begnügen wir uns mit einem Speziallfall [12, Korollar 10], den wir auf den Fall der präkonditionierten CKS-Methoden *Verfahren 1** und *Verfahren 2** übertragen. Für den freien Parameter l_k muß im folgenden immer $l_k \leq l$ gelten.

Lemma 1
Das lineare Gleichungssystem

$$\begin{pmatrix} B_{1,1} & B_{1,2} \\ 0 & B_{2,2} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} b_1 \\ 0 \end{pmatrix} \tag{2.1}$$

mit $B_{1,1} \in \mathrm{GL}_j(\mathbb{R})$, $B_{1,2} \in \mathbb{R}^{j\times(k-j)}$, $B_{2,2} \in \mathrm{GL}_{k-j}(\mathbb{R})$, $y_1, b_1 \in \mathbb{R}^j$ *und* $y_2 \in \mathbb{R}^{k-j}$ *hat die Lösung*

$$\begin{aligned} y_1 &= B_{1,1}^{-1} b_1\,, \\ y_2 &= 0\,. \end{aligned}$$

Beweis. Trivial. □

Satz 2
Das exakte präkonditionierte CKS-Verfahren 2 läßt sich mit l-stufiger Rekursion darstellen, d. h. $\mu_{l+1,k} = \cdots = \mu_{k,k} = 0$ für alle $k > l$, falls gilt*

$$r_{k-l_k}^T \tilde{Z}_k r_{k-i} = 0 \quad \text{für } i = l+1, \ldots, k, \tag{2.2}$$

$$r_{k-j}^T \tilde{Z}_k^T A^T \tilde{Z}_k r_{k-i} = 0 \quad \text{für } i = l+1, \ldots, k;\ j = 1, \ldots, l. \tag{2.3}$$

Beweis. Die Bedingungen (2.2) und (2.3) sind so gewählt, daß das Gleichungssystem (1.12) die spezielle Form (2.1) aus Lemma 1 mit $j = l$ hat. (O.K.?) □

Dieser Satz ist die Ausgangsbasis der folgenden Kapitel, in denen es darum geht, wie man Orthogonalisierungmatrizen $\tilde{Z}_k$ so konstruieren kann, daß die Verfahren 1^* und 2^* eine kurze Rekursion erhalten. Wir wählen hier eine zweistufige Rekursion in Analogie zum klassischen cg-Verfahren und zu GMRES, die beide für symmetrische Matrizen eine zweistufige Rekursion haben.

Kapitel 3

Zweistufige Rekursion für Verfahren 2*

In diesem Kapitel konstruieren wir die Orthogonalitätsmatrizen $\tilde{Z}_k$ so, daß das exakte Verfahren 2^* eine zweistufige Rekursion erhält. Im folgenden gilt immer

$$l = 2\,, \tag{3.1}$$

$$\sigma_k = \min(2, k)\,. \tag{3.2}$$

Für die Wahl von l_k hat man wegen $l_k \leq l$ nur die zwei Möglichkeiten $l_k = 1$ oder $l_k = \sigma_k$.

Die Bedingungen aus Satz (2) für eine zweistufige Rekursion für das Verfahren 2^* lauten ausgeschrieben für $k \geq 3$

$$r_{k-l_k}^T \tilde{Z}_k r_{k-j} = 0\,, \tag{3.3a}$$

$$r_{k-1}^T \tilde{Z}_k^T A^T \tilde{Z}_k r_{k-j} = 0\,, \tag{3.3b}$$

$$r_{k-2}^T \tilde{Z}_k^T A^T \tilde{Z}_k r_{k-j} = 0 \tag{3.3c}$$

für $j = 3, \ldots, k$.

Die Gleichungen (3.3b) und (3.3c) bereitet dabei die meisten Probleme, da sie quadratische Gleichungen in $\tilde{Z}_k$ sind.

Um die Forderungen (3.3) zu erfüllen, muß man ab dem dritten Schritt die Matrix $\tilde{Z}_k$ geeignet wählen. In den ersten beiden Iterationsschritten setzt man

$$\tilde{Z}_1 = \tilde{Z}_2 = I\,. \tag{3.4}$$

Damit die Berechnung eines $\tilde{Z}_k \in \mathbb{R}^{n \times n}$ keinen Aufwand benötigt, der über dem Aufwand für das eigentliche Problem liegt, verwendet man für $k \geq 3$ als $\tilde{Z}_k$ eine Rank-3-Update-Matrix. Das ist eine Matrix, die sich von einer festen, nichtsingulären Matrix Z nur um einen Anteil vom Rang 3 unterscheidet, d.h. sie läßt sich darstellen in der Form

$$\tilde{Z}_k = Z + b_{1,k}c_{1,k}^T + b_{2,k}c_{2,k}^T + b_{3,k}c_{3,k}^T\,, \tag{3.5}$$

mit $b_{i,k}, c_{i,k} \in \mathbb{R}^n$ für $i = 1, 2, 3$ geeignet [9, § 2.7]. Oft wählt man für Z die Einheitsmatrix I. Dadurch und durch die Rank-3-Update-Form versucht man zu erreichen, daß $\tilde{Z}_k$ in einem gewissen Sinne „nahe" an der Einheitsmatrix liegt.

Definiert man für $k \geq 3$ die $(n \times 3)$-Matrizen

$$B_k = (b_{1,k}, b_{2,k}, b_{3,k})\,, \tag{3.6a}$$
$$C_k = (c_{1,k}, c_{2,k}, c_{3,k})\,, \tag{3.6b}$$

so erhält man für (3.5) die abkürzende Schreibweise

$$\tilde{Z}_k = Z + B_k C_k^T \tag{3.7}$$

für $k \geq 3$. Wie löst man nun mit der Rank-3-Update-Matrix $\tilde{Z}_k$ die Gleichungen (3.3)?

Definition 4 *Wir definieren als*

$$S_k = \operatorname{span}\{r_0, r_1, \ldots, r_k\}$$

den von den Residuen $r_0, \ldots, r_k$ aufgespannten linearen Unterrraum vom $\mathbb{R}^n$. Mit $S_k^\perp$ bezeichnen wir den Orthogonalraum von S_k bezüglich $\mathbb{R}^n$.

Desweiteren definieren wir mit

$$T_k = \operatorname{span}\{x_1 - x_0, x_2 - x_1, \ldots, x_k - x_{k-1}\}$$

den von den Differenzen von je zwei aufeinanderfolgenden Iterationsvektoren aufgespannten linearen Unterrraum vom $\mathbb{R}^n$.

Wir setzen in diesem Kapitel voraus, daß die Residuen r_k bzw. die Differenzen von zwei aufeinanderfolgenden Iterationsvektoren $x_k - x_{k-1}$ in allgemeiner Lage sind, d. h. daß

$$\dim S_{k-1} = k\,, \tag{3.8}$$
$$\dim T_k = k \tag{3.9}$$

für alle $k = 1, \ldots, n$ gilt.

Mit Hilfe von Definition 4 lassen sich die Forderungen (3.3) schreiben als

$$\tilde{Z}_k^T r_{k-l_k}, \tilde{Z}_k^T A\tilde{Z}_k r_{k-1}, \tilde{Z}_k^T A\tilde{Z}_k r_{k-2} \in S_{k-3}^{\perp} . \tag{3.10}$$

Man muß also $\tilde{Z}_k$ so bestimmen, daß die drei Vektoren $\tilde{Z}_k^T r_{k-l_k}$, $\tilde{Z}_k^T A\tilde{Z}_k r_{k-1}$, $\tilde{Z}_k^T A\tilde{Z}_k r_{k-2}$ im Orthogonalraum von S_{k-3} liegen. Wir zerteilen diese Aufgabe in zwei Teile. Zuerst bestimmt man drei Vektoren $h_{1,k}, h_{2,k}, h_{3,k} \in S_{k-3}^{\perp}$. Danach berechnet mit Hilfe des folgenden Satzes $\tilde{Z}_k$.

Satz 3
Für $k \geq 3$ seien die $(n \times 3)$-Matrizen

$$G_k = (r_{k-l_k}, A\tilde{Z}_k r_{k-1}, A\tilde{Z}_k r_{k-2}) , \tag{3.11}$$
$$H_k = (h_{1,k}, h_{2,k}, h_{3,k}) \tag{3.12}$$

mit $h_{1,k}, h_{2,k}, h_{3,k} \in S_{k-3}^{\perp}$ definiert.

Falls für jedes $k \geq 3$ ein beliebiges $W_k \in \mathrm{GL}_3(\mathbb{R})$ existiert mit

$$\tilde{Z}_k^T G_k = H_k W_k , \tag{3.13}$$

so hat das Verfahren 2 eine zweistufige Rekursion.*

Beweis. Für ein festes $j \in \{3, \ldots, k\}$ sind die Bedingungen (3.3) für eine zweistufige Rekursion äquivalent zu $G_k^T \tilde{Z}_k r_{k-j} = 0$. Das folgt wegen $H_k^T r_{k-j} = 0$ sofort aus (3.13). □

Damit ergibt sich nun das Problem, im k-ten Schritt die drei Vektoren $h_{1,k}, h_{2,k}, h_{3,k}$ zu bestimmen. Leider kann man sie nicht beliebig wählen, sie müssen nämlich linear unabhängig sein. Das liegt daran, daß auch die drei Spaltenvektoren von G_k i. a. linear unabhängig sind. Aus $\mathrm{Rang}(H_k) < 3$ würde aus (3.13) folgen, daß $\tilde{Z}_k^T$ und damit auch $\tilde{Z}_k$ singulär sind. Das darf aber nach Definition der CKS-Verfahren nicht sein.

Für die ersten zwei Vektoren bieten sich

$$h_{1,k} = \tilde{Z}_{k-1}^T r_{k-1} ,$$
$$h_{2,k} = \tilde{Z}_{k-2}^T r_{k-2}$$

an, denn aus der Orthogonalitätsbedingung (1.12) auf Seite 6 folgt, daß $\tilde{Z}_{k-1}^T r_{k-1}$ und $\tilde{Z}_{k-2}^T r_{k-2}$ in $S_{k-3}^{\perp}$ liegen. Nun fehlt nur noch ein weiterer Vektor in $S_{k-3}^{\perp}$. Um diesen zu konstruieren, greifen wir auf die Darstellung (1.11) von

r_k zurück. Aus ihr folgt mit der zweistufigen Rekursion

$$\sum_{i=1}^{\sigma_k} \mu_{i,k} A\tilde{Z}_k r_{k-i} = r_k - r_{k-l_k} \tag{3.14}$$

für $k \geq 1$. Daraus folgt für $2 \leq j \leq k$

$$\begin{aligned} r_k^T \tilde{Z}_k A\tilde{Z}_{k-j+1} \sum_{i=1}^{\sigma_{k-j+1}} \mu_{i,k-j+1} r_{k-j+1-i} &\overset{(3.14)}{=} r_k^T \tilde{Z}_k (r_{k-j+1} - r_{k-j+1-l_{k-j+1}}) \\ &\overset{(1.12)}{=} 0\,. \end{aligned} \tag{3.15}$$

Wegen (3.8) und (3.9) folgt aus (3.15) mit $\tilde{Z}_1 = \tilde{Z}_2 = I$ für alle $k \geq 1$

$$r_k^T \tilde{Z}_k A r_0 = r_k^T \tilde{Z}_k A r_1 = 0\,. \tag{3.16}$$

Das Ziel ist es nun, diese Gleichung außer für r_0 bzw. r_1 auch für $r_2, \ldots, r_{k-2}$ zu erhalten. Dann hätte man mit $A^T \tilde{Z}_{k-1}^T r_{k-1}$ einen dritten Vektor in $S_{k-3}^{\perp}$ konstruiert. Dazu ist es notwendig, an die $\tilde{Z}_k$ gewisse Voraussetzungen zu machen.

Hierzu gibt es verschiedene Möglichkeiten. Im folgenden werden drei Ansätze vorgestellt, mit denen dieses Ziel erreicht wird.

3.1 Ansatz 2

Wie man im letzten Kapitel gesehen hat, führt die Wahl von $b_{i,k} = r_{k-i}$ nicht zum Erfolg. Man legt deshalb die $b_{i,k}$ vorerst nicht fest. Dafür setzt man wieder

$$Z = I \tag{3.17}$$

und fordert außerdem

$$C_k^T r_{k-1} = 0\,, \tag{3.18a}$$
$$C_k^T r_{k-2} = 0 \tag{3.18b}$$

für alle $k \geq 3$, woraus mit (3.7) und $\tilde{Z}_1 = \tilde{Z}_2 = I$

$$\tilde{Z}_k r_{k-1} = r_{k-1}\,, \tag{3.19a}$$
$$\tilde{Z}_k r_{k-2} = r_{k-2} \tag{3.19b}$$

für alle $k \geq 2$ folgt. Damit verschwindet zum einen die Abhängigkeit von $\tilde{Z}_k$ bei G_k, denn nun gilt nach (3.11)

$$G_k = (r_{k-l_k}, Ar_{k-1}, Ar_{k-2})\,, \tag{3.20}$$

andererseits kann man eine Aussage analog zum Lemma **??** treffen.

Lemma 2
Haben im Verfahren 2 die Orthogonalitätsmatrizen $\tilde{Z}_k$ für alle $k \geq 3$ die Form $\tilde{Z}_k = I + B_k C_k^T$ mit (3.18), und gilt für die Iterierten $x_1, \ldots, x_{k-1}$ eine zweistufige Rekursion mit $\mu_{1,1}, \ldots, \mu_{1,k-1} \neq 0$, so gilt im k-ten Schritt*

$$r_k^T \tilde{Z}_k A r_{k-j} = 0 \tag{3.21}$$

für $j = 2, \ldots, k$.

Beweis. Der Beweis erfolgt durch vollständige Induktion über absteigendes j. Für $j = k$ bzw. $j = k-1$ folgt die Behauptung aus (3.16). Sei nun die Behauptung richtig für $j = \tilde{\jmath}+1, \ldots, k$ für ein festes $\tilde{\jmath} \in \{2, \ldots, k-2\}$. Zu zeigen ist, daß die Behauptung dann auch für $j = \tilde{\jmath}$ gilt. Wegen der zweistufigen Rekursion und $k - \tilde{\jmath} + 1 \geq 3$ gilt (3.15) mit $\sigma_{k-\tilde{\jmath}+1} = 2$.

$$\begin{aligned}
0 &= r_k^T \tilde{Z}_k A \tilde{Z}_{k-\tilde{\jmath}+1} \sum_{i=1}^{\sigma_{k-\tilde{\jmath}+1}} \mu_{i,k-\tilde{\jmath}+1} r_{k-\tilde{\jmath}+1-i} \\
&= \mu_{1,k-\tilde{\jmath}+1} r_k^T \tilde{Z}_k A \tilde{Z}_{k-\tilde{\jmath}+1} r_{k-\tilde{\jmath}} + \mu_{2,k-\tilde{\jmath}+1} r_k^T \tilde{Z}_k A \tilde{Z}_{k-\tilde{\jmath}+1} r_{k-\tilde{\jmath}-1} \\
&\overset{(3.19)}{=} \mu_{1,k-\tilde{\jmath}+1} r_k^T \tilde{Z}_k A r_{k-\tilde{\jmath}} + \mu_{2,k-\tilde{\jmath}+1} r_k^T \tilde{Z}_k A r_{k-(\tilde{\jmath}+1)} \\
&\overset{\text{I.A.}}{=} \mu_{1,k-\tilde{\jmath}+1} r_k^T \tilde{Z}_k A r_{k-\tilde{\jmath}}
\end{aligned}$$

Aus $\mu_{1,k-\tilde{\jmath}+1} \neq 0$ folgt die Behauptung. □

Man kann also wie im ersten Ansatz

$$H_k = (\tilde{Z}_{k-1}^T r_{k-1}, \tilde{Z}_{k-2}^T r_{k-2}, A^T \tilde{Z}_{k-1}^T r_{k-1}) \tag{3.22}$$

verwenden, denn alle drei Spaltenvektoren liegen in $S_{k-3}^{\perp}$. Ebenso berechnet man nun C_k aus

$$C_k B_k^T G_k = H_k W_k - G_k\,, \tag{3.23}$$

wobei die nichtsinguläre Matrix $W_k \in \mathbb{R}^{3\times 3}$ noch frei gewählt werden kann. Falls $B_k^T G_k$ nichtsingulär ist, wovon wir im folgenden ausgehen, so ergibt sich C_k analog durch

$$C_k = (H_k W_k - G_k)(B_k^T G_k)^{-1}\,. \tag{3.24}$$

Wenn man (3.24) in die Forderung (3.18) einsetzt, so erhält man

$$\begin{aligned}
& C_k^T r_{k-j} = 0 \\
\iff\ & (G_k^T B_k)^{-1}(W_k^T H_k^T - G_k^T) r_{k-j} = 0 \\
\iff\ & W_k^T H_k^T r_{k-j} = G_k^T r_{k-j}
\end{aligned} \tag{3.25}$$

für $j = 1, 2$. Definiert man

$$y_j^k = H_k^T r_{k-j}\,,$$

$$z_j^k = G_k^T r_{k-j}$$

für $j = 1, 2$, so wird (3.25) äquivalent zu

$$W_k^T y_j^k = z_j^k \tag{3.26}$$

für $j = 1, 2$. Falls y_1^k und y_2^k bzw. z_1^k und z_2^k jeweils linear unabhängig sind, so existiert ein reguläres W_k mit (3.26). Eine Möglichkeit, ein solches W_k zu konstruieren, ist z. B.

$$W_k = (y_1^k, y_2^k, y_1^k \times y_2^k)^{-T} (z_1^k, z_2^k, z_1^k \times z_2^k)^T \,. \tag{3.27}$$

Hierbei ist $\times$ das Kreuzprodukt zweier Vektoren im $\mathbb{R}^3$.

An die Matrix B_k wurden bisher nur zwei Bedingungen gestellt:

- Der von den Spaltenvektoren von B_k aufgespannte Raum sollte nicht in $\operatorname{span}\{r_{k-1}, r_{k-2}, r_{k-3}\}$ liegen und
- $B_k^T G_k$ darf nicht singulär werden.

Das folgende Lemma zeigt, daß es bei der Wahl von B_k nur auf den von den Spaltenvektoren $b_{i,k}$ aufgespannten Raum ankommt.

Lemma 3
Seien $\tilde{Z}_k = Z + \tilde{B}_k \tilde{C}_k$ und $Z_k = Z + B_k C_k^T$ zwei Matrizen, die beide (3.13) erfüllen. Weiterhin gelte

$$\tilde{B}_k = B_k U$$

mit einem $U \in \mathrm{GL}_3(\mathbb{R})$, und sowohl $\tilde{B}_k^T G_k$ als auch $B_k^T G_k$ seien invertierbar. Dann gilt

$$\tilde{Z}_k = Z_k \,.$$

Beweis. Aus (3.13) folgt, daß sowohl $\tilde{C}_k$ bzw. $\tilde{B}_k$ als auch C_k bzw. B_k die Gleichung (**??**) erfüllen. Damit gilt

$$\begin{aligned}
\tilde{C}_k &= (H_k W_k - G_k)(\tilde{B}_k^T G_k)^{-1} \\
&= (H_k W_k - G_k)(U^T B_k^T G_k)^{-1} \\
&= (H_k W_k - G_k)(B_k^T G_k)^{-1} U^{-T} \\
&= C_k U^{-T}
\end{aligned}$$

Daraus folgt $\tilde{C}_k^T = U^{-1} C_k^T$ und damit

$$\begin{aligned}
\tilde{Z}_k &= Z + \tilde{B}_k \tilde{C}_k^T \\
&= Z + B_k U U^{-1} C_k^T
\end{aligned}$$

$$
\begin{aligned}
&= Z + B_k C_k \\
&= Z_k \,. \qquad \square
\end{aligned}
$$

Als mögliche Wahl für B_k bietet sich

$$B_k = G_k$$

an. Dann wird $B_k^T G_k$ nur dann singulär, wenn schon $\mathrm{Rang}(G_k) < 3$ gilt. Eine andere Möglichkeit für B_k ist

$$B_k = (Ar_{k-1}, r_{k-1}, r_{k-2})\,.$$

Mit diesem B_k wird die Darstellung der Konvergenzaussagen später etwas einfacher.

3.1.1 Algorithmus 1

Wir fassen die Vorgehensweise zusammen, indem wir für die in diesem Kapitel konstruierte spezielle Variante von Verfahren 2^* einen Algorithmus angeben.

Algorithmus 1 *Man wählt eine Startnäherung x_0 für das Gleichungssystem $Ax = b$ und setzt $r_0 = Ax_0 - b$.*

Für $k \geq 1$ setzt man

$$l_k = \sigma_k = \min(2, k)\,.$$

Für $k \geq 1$ berechnet man x_k aus dem Ansatz

$$x_k = x_{k-l_k} + \sum_{i=1}^{\sigma_k} \mu_{i,k} r_{k-i}\,. \tag{3.28}$$

Für die zugehörigen Residuen r_k gilt

$$r_k = r_{k-l_k} + \sum_{i=1}^{\sigma_k} \mu_{i,k} A r_{k-i}\,. \tag{3.29}$$

Die Koeffizienten $\mu_{i,k}$ berechnen sich aus

$$r_k^T r_{k-j} = 0 \tag{3.30}$$

für $j = 1, \ldots, \sigma_k$, d. h. als Lösung des linearen Gleichungssystems

$$\sum_{i=1}^{\sigma_k} r_{k-i}^T A^T r_{k-j} \mu_{i,k} = -r_{k-l_k}^T r_{k-j}$$

für $j = 1, \ldots, \sigma_k$.

3.1.2 Eigenschaften von Algorithmus 1

Wegen der Forderung (3.19) verschwindet $\tilde{Z}_k$ im Ansatz (3.29) von r_k und auch aus der Orthogonalitätsbedingung (3.30). Damit wird der Algorithmus 1 ganz unabhängig von der Matrix $\tilde{Z}_k$ und entpuppt sich als eine spezielle abgeschnittene Version eines verallgemeinerten cg-Verfahrens mit $P = I$ und $Z = I$. Gleichzeitig ist der Algorithmus 1 auch ein nichtstationäres $(2,2)$-stufiges Verfahren nach der Definition von Gutknecht [7]. Dort werden als (k,l)-stufige Verfahren lineare Gleichungslöser klassifiziert, bei denen jeder neue Iterationsvektor sich zusammensetzt aus einer Linearkombination der letzten k zurückliegenden Iterierten und der letzten l zurückliegenden Residuen, d. h. man macht den Ansatz

$$x_m = \sum_{j=1}^{\min(m,k)} \gamma_{j,m} x_{m-j} + \sum_{i=1}^{\min(m,l)} \alpha_{i,m} r_{m-i}\,.$$

Falls man im Algorithmus 1 immer $l_k = 1$ für alle $k \geq 1$ setzt, so erhält man ein $(1,2)$-stufiges Verfahren.

3.2 Ansatz 3

Der dritte Ansatz unterscheidet sich nur geringfügig vom vorhergehenden. Man setzt

$$Z = I \tag{3.31}$$

und fordert im Gegensatz zum Ansatz 2 hier

$$\tilde{Z}_k r_{k-1} = r_{k-1}\,, \tag{3.32a}$$

$$\tilde{Z}_k r_{k-2} = x_{k-1} - x_{k-2}\,. \tag{3.32b}$$

Damit wird aus dem Ansatz (1.10) der Iterierten x_k mit $k \geq 2$

$$x_k = x_{k-l_k} + \mu_{2,k}(x_{k-1} - x_{k-2}) + \mu_{1,k} r_{k-1}\,. \tag{3.33}$$

Man kann also in diesem Ansatz ohne Einschränkung

$$l_k = 1$$

setzen. Durch die Forderungen (3.32) wird

$$G_k = \Big(r_{k-1}, A r_{k-1}, A(x_{k-1} - x_{k-2})\Big) \tag{3.34}$$

wie beim Ansatz 2 unabhängig von $\tilde{Z}_k$ und damit werden aus den quadratischen Gleichungen (3.3b) und (3.3c) lineare Gleichungen für $\tilde{Z}_k$.

Um für diesen Ansatz eine Aussage analog zu Lemma **??** und Lemma 2 machen zu können, beweisen wir zuerst ein vorbereitendes Lemma.

Lemma 4
Haben die Orthogonalitätsmatrizen $\tilde{Z}_k$ im Verfahren 2 die Form $\tilde{Z}_k = I + B_k C_k^T$ mit (3.32), und hat das Verfahren in den ersten $k-1$ Schritten eine zweistufige Rekursion, so gilt*

$$x_{\tilde{k}} - x_{\tilde{k}-j} \in S_{\tilde{k}-1} \tag{3.35}$$

für alle $\tilde{k} = 1, \ldots, k-1$ und alle $j \leq \tilde{k}$.

Beweis. Wegen der zweistufigen Rekursion ergibt sich für die $x_{\tilde{k}}$ für $\tilde{k} \in \{1, \ldots, k-1\}$ die Darstellung

$$x_{\tilde{k}} = x_{\tilde{k}-1} + \sum_{i=1}^{\sigma_{\tilde{k}}} \mu_{i,\tilde{k}} \tilde{Z}_{\tilde{k}} r_{\tilde{k}-i} \tag{3.36}$$

Der Beweis erfolgt durch vollständige Induktion über $\tilde{k}$. Für $\tilde{k} = 1$ gilt wegen (3.36)

$$x_1 - x_0 = \mu_{1,1} \tilde{Z}_1 r_0 = \mu_{1,1} r_0 \in S_0 \,, \tag{3.37}$$

d. h. die Behauptung gilt für $\tilde{k} = 1$ und $j = 1$.

Sei nun die Behauptung richtig für $1, \ldots, \tilde{k} - 1$. Zu zeigen ist dann, daß sie auch für $\tilde{k}$ gilt. Da hier $\tilde{k} \geq 2$ vorausgesetzt werden kann, gilt $\sigma_{\tilde{k}} = 2$. Für $j = 1$ folgt die Behauptung sofort aus (3.36), aus $l_k = 1$ und der Induktionsannahme:

$$\begin{aligned} x_{\tilde{k}} - x_{\tilde{k}-1} &\overset{(3.36)}{=} \mu_{1,\tilde{k}} \tilde{Z}_k r_{k-1} + \mu_{2,k} \tilde{Z}_k r_{\tilde{k}-2} \\ &\overset{(3.32)}{=} \mu_{1,\tilde{k}} r_{k-1} + \mu_{2,k} (x_{\tilde{k}-1} - x_{\tilde{k}-2}) \\ &\in \; S_{k-1} \,. \end{aligned}$$

Daraus folgt zusammen mit der Induktionsannahme für $\tilde{k} - 1$ die Aussage für $j \in \{2, \ldots, \tilde{k}\}$

$$x_{\tilde{k}} - x_{\tilde{k}-j} = (x_{\tilde{k}} - x_{\tilde{k}-1}) - (x_{\tilde{k}-1} - x_{\tilde{k}-j}) \in S_{k-1} \,. \qquad \square$$

Lemma 5
Haben die Orthogonalitätsmatrizen $\tilde{Z}_k$ im Verfahren 2 die Form $\tilde{Z}_k = I + B_k C_k^T$ mit (3.32), und hat das Verfahren in den ersten $k-1$ eine zweistufige Rekursion, so gilt*

$$r_k^T \tilde{Z}_k A r_{k-j} = 0 \tag{3.38}$$

für $j = 2, \ldots, k$.

Beweis. Das vorangegangene Lemma 4 besagt, daß für alle $\tilde{k} < k$ gilt $x_{\tilde{k}} - x_{\tilde{k}-1} \in S_{\tilde{k}-1}$. Wegen der Voraussetzung (3.9) an die allgemeine Lage der Differenzen aufeinanderfolgender Iterierten folgt

$$x_{\tilde{k}} - x_{\tilde{k}-1} \in S_{\tilde{k}-1} \setminus S_{\tilde{k}-2}\,,$$

d. h. für $\tilde{k} = 1, \ldots, k-1$ gilt $\mu_{1,\tilde{k}} \neq 0$.

Der eigentliche Beweis erfolgt durch vollständige Induktion über absteigendes j. Für $j = k$ bzw. $j = k-1$ folgt die Behauptung aus (3.16). Sei nun die Behauptung richtig für $j = \tilde{\jmath}+1, \ldots, k$ für ein festes $\tilde{\jmath} \in \{2, \ldots, k-2\}$. Zu zeigen ist, daß die Behauptung dann auch für $j = \tilde{\jmath}$ gilt. Wegen der zweistufigen Rekursion und $k - \tilde{\jmath} + 1 \geq 3$ gilt (3.15) mit $\sigma_{k-\tilde{\jmath}+1} = 2$.

$$\begin{aligned}
0 &= r_k^T \tilde{Z}_k A \tilde{Z}_{k-\tilde{\jmath}+1} \sum_{i=1}^{\sigma_{k-\tilde{\jmath}+1}} \mu_{i,k-\tilde{\jmath}+1} r_{k-\tilde{\jmath}+1-i} \\
&= \mu_{1,k-\tilde{\jmath}+1} r_k^T \tilde{Z}_k A \tilde{Z}_{k-\tilde{\jmath}+1} r_{k-\tilde{\jmath}} + \mu_{2,k-\tilde{\jmath}+1} r_k^T \tilde{Z}_k A \tilde{Z}_{k-\tilde{\jmath}+1} r_{k-\tilde{\jmath}-1} \\
&\overset{(3.32)}{=} \mu_{1,k-\tilde{\jmath}+1} r_k^T \tilde{Z}_k A r_{k-\tilde{\jmath}} + \mu_{2,k-\tilde{\jmath}+1} r_k^T \tilde{Z}_k A (\underbrace{x_{k-\tilde{\jmath}} - x_{k-(\tilde{\jmath}+1)}}_{\in S_{k-(\tilde{\jmath}+1)}}) \\
&\overset{\text{I.A.}}{=} \mu_{1,k-\tilde{\jmath}+1} r_k^T \tilde{Z}_k A r_{k-\tilde{\jmath}}\,.
\end{aligned}$$

Aus $\mu_{1,k-\tilde{\jmath}+1} \neq 0$ folgt die Behauptung. □

Damit kann man die gleiche Matrix H_k nach (3.22) wie im Ansatz 2 verwenden und erhält analog für C_k die Formeln (3.23) und (3.24). Nun muß man W_k noch so bestimmen, daß $\tilde{Z}_k$ die Forderungen (3.32) erfüllt. Aus (3.32a) ergibt sich, falls $G_k^T B_k$ nichtsingulär ist, an W_k die folgende Forderung

$$\begin{aligned}
& \tilde{Z}_k r_{k-1} = r_{k-1} \\
\iff\ & B_k C_k^T r_{k-1} = 0 \\
\iff\ & C_k^T r_{k-1} = 0 \\
\iff\ & (G_k^T B_k)^{-1} (W_k^T H_k^T - G_k^T) r_{k-1} = 0 \\
\iff\ & W_k^T H_k^T r_{k-1} = G_k^T r_{k-1}
\end{aligned} \tag{3.39}$$

Die Gleichung (3.32b) läßt sich wegen $\tilde{Z}_k = I + B_k C_k^T$ umformen zu

$$\begin{aligned}
& \tilde{Z}_k r_{k-2} = x_{k-1} - x_{k-2} \\
\iff\ & B_k C_k^T r_{k-2} = x_{k-1} - x_{k-2} - r_{k-2}\,.
\end{aligned} \tag{3.40}$$

Daraus folgt, daß

$$x_{k-1} - x_{k-2} - r_{k-2} \in \operatorname{span}\{b_{1,k}, b_{2,k}, b_{3,k}\}$$

sein muß. O. B. d. A. setzt man

$$b_{3,k} = x_{k-1} - x_{k-2} - r_{k-2} \tag{3.41}$$

und erhält aus (3.40) mit $e_3^3 = (0, 0, 1)^T$

$$\begin{aligned}
(3.40) \iff & C_k^T r_{k-2} = e_3^3 \\
\iff & (G_k^T B_k)^{-1}(W_k^T H_k^T - G_k^T) r_{k-2} = e_3^3 \\
\iff & W_k^T H_k^T r_{k-2} = G_k^T r_{k-2} + G_k^T B_k e_3^3 \\
\iff & W_k^T H_k^T r_{k-2} = G_k^T (r_{k-2} + b_{3,k}) \\
\iff & W_k^T H_k^T r_{k-2} = G_k^T (x_{k-1} - x_{k-2}) \qquad (3.42)
\end{aligned}$$

Definiert man

$$\begin{aligned}
y_1^k &= H_k^T r_{k-1}\,, \\
y_2^k &= H_k^T r_{k-2}\,, \\
z_1^k &= G_k^T r_{k-1}\,, \\
z_2^k &= G_k^T (x_{k-1} - x_{k-2})\,,
\end{aligned}$$

so sind (3.39) und (3.42) äquivalent zu

$$W_k^T y_j^k = z_j^k \tag{3.43}$$

für $j = 1, 2$. Auch hier gilt, daß, falls y_1^k und y_2^k bzw. z_1^k und z_2^k jeweils linear unabhängig sind, ein reguläres W_k mit (3.43) existiert.

Wenn man nun noch $b_{1,k}$ bzw. $b_{2,k}$ geeignet festlegt, etwa als

$$\begin{aligned}
b_{1,k} &= A r_{k-1}\,, \\
b_{2,k} &= r_{k-1}\,,
\end{aligned}$$

so hat man eine zweite spezielle Variante von Verfahren 2* mit zweistufiger Rekursion konstruiert.

3.2.1 Algorithmus 2

Auch für die mit dem Ansatz 3 konstruierte Variante von Verfahren 2* wollen wir wieder einen Algorithmus angeben.

Algorithmus 2 *Man wählt eine Startnäherung x_0 für das Gleichungssystem $Ax = b$ und setzt $r_0 = Ax_0 - b$.*

Man berechnet x_k für $k = 1$ aus dem Ansatz

$$x_1 = x_0 + \beta_{1,k} r_0 \,,$$

für $k \geq 2$ aus

$$x_k = x_{k-1} + \beta_{1,k} r_{k-1} + \beta_{2,k}(x_{k-1} - x_{k-2}) \,.$$

Für die zugehörigen Residuen r_k gilt

$$r_1 = r_0 + \beta_{1,k} A r_0$$

und für $k \geq 2$

$$r_k = r_{k-1} + \beta_{1,k} A r_{k-1} + \beta_{2,k}(r_{k-1} - r_{k-2}) \,.$$

Die Koeffizienten $\beta_{i,k}$ berechnen sich aus

$$r_k^T r_{k-1} = 0$$

und für $k \geq 2$ zusätzlich aus

$$r_k^T (x_{k-1} - x_{k-2}) = 0 \,. \tag{3.44}$$

Für $k \geq 2$ ist das äquivalent zu dem linearen Gleichungssystem

$$\begin{pmatrix} r_{k-1}^T A^T r_{k-1} & (r_{k-1} - r_{k-2})^T r_{k-1} \\ r_{k-1}^T A^T (x_{k-1} - x_{k-2}) & (r_{k-1} - r_{k-2})^T (x_{k-1} - x_{k-2}) \end{pmatrix} \begin{pmatrix} \beta_{1,k} \\ \beta_{2,k} \end{pmatrix} = - \begin{pmatrix} r_{k-1}^T r_{k-1} \\ r_{k-1}^T (x_{k-1} - x_{k-2}) \end{pmatrix} .$$

3.2.2 Bemerkungen zu Algorithmus 2

Der Algorithmus 2 unterscheidet sich vom klassischen cg-Verfahren nur durch die zweite Orthogonalitätsbedingung (3.44), die beim klassischen cg-Verfahren

$$r_k^T r_{k-2} = 0 \tag{3.45}$$

lautet. Es ist interessant, in den numerischen Testrechnungen in Kapitel 4 herauszufinden, wie stark sich die beiden Verfahren unterscheiden, d. h. wie stark sich die Änderung *einer* Orthogonalitätsbedingung auf die Iteration auswirkt.

3.3 Exakt und kurze Rekursion?

Die in Algorithmus 1 bzw. Algorithmus 2 angegebenen Verfahren sind exakte pCKS-Verfahren mit fester Orthogonalitätsmatrix $Z_k = I$, aber schrittabhängigen Suchrichtungen. Gleichzeitig sind sie aber auch spezielle abgeschnittene (also *nichtexakte*) verallgemeinerte cg-Verfahren.

Es wird damit in zwei Fällen die Äquivalenz eines exakten pCKS-Verfahrens mit einem abgeschnittenen verallgemeinerten cg-Verfahren gezeigt. Die ersteren erfüllen im k-ten Schritt eine Minimierungseigenschaft über einem k-dimensionalen affinen Unterrraum des $\mathbb{R}^n$, die letzeren nicht. Das bedeutet, daß die pCKS-Verfahren im n-ten Schritt mit der Lösung terminieren müßten. Diese Eigenschaft haben die verallgemeinerten cg-Verfahren im allgemeinen nicht. Dieser scheinbare Widerspruch löst sich durch den nachfolgenden Satz auf.

Satz 4
Die Algorithmen 1 bzw. 2 brechen spätestens im n-ten Schritt zusammen.

Beweis. Nach der Voraussetzung über die allgemeine Lage der Residuen r_k gilt $\dim(S_k) = k + 1$ und damit

$$\dim(S_{n-3}^{\perp}) = n - \dim(S_{n-3}) = 2\,.$$

Damit ist es nicht mehr möglich, drei linear unabhängige Vektoren in $S_{n-3}^{\perp}$ zu finden. $\square$

Die drei von Algorithmus 1 bzw. 2 berechneten Vektoren h_{1n}, h_{2n}, h_{3n} sind also linear abhängig, womit beide Verfahren im letzten Schritt zusammenbrechen. Genaugenommen brechen die Verfahren auch nur insofern zusammen, als es um die Berechnung von $\tilde{Z}_k$ geht. Da aber $\tilde{Z}_k$ nicht in die Berechnung von x_k bzw. r_k eingeht, kann man die Iteration auch über $k = n$ hinaus laufen lassen, wie man es von einem abgeschnittenen verallgemeinerten cg-Verfahren ja erwartet. Nur die Orthogonalitätsbedingungen und Minimierungseigenschaften des pCKS-Verfahrens gelten dann nicht mehr.

In der Praxis ist dies aber keine Einschränkung, weil n in der Regel so groß ist, daß man selten bis $k = n$ iteriert.

3.3.1 Kann man eine Dimension dazugewinnen?

Eine andere Idee, die Dualität von präkonditioniertem CKS-Verfahren und verallgemeinertem cg-Verfahren auszunutzen, wäre es, im Algorithmus 1 außer (3.19) für $k \geq 3$ auch

$$\tilde{Z}_k r_{k-3} = r_{k-3} \tag{3.46}$$

zu fordern, um eine zusätzliche Orthogonalität zu erzwingen. So ein $\tilde{Z}_k$ existiert aber im allgemeinen nicht.

3.4 Spezielle Konvergenzaussagen

Nun ist es natürlich interessant, was für Konvergenzaussagen man über den Umweg über präkonditionierte CKS-Verfahren gewonnen hat. Man kann die Aussage des Korollars 4 von Seite 6, das für den Algorithmus 1 gilt, noch verfeinern, indem man die spezielle Form von $\tilde{Z}_k$ ausnutzt.

Satz 5
Falls A positiv reell ist, d. h. der symmetrische Anteil von A positiv definit ist, so gilt für das im Algorithmus 1 angegebene Verfahren

$$\begin{aligned}
\|r_k\|_{A^{-1}} &\leq \sqrt{\frac{1+\varrho^2(R)}{\mu_m^2}} \min_{\eta \in \mathbb{R}^k} \left\| r_{k-l_k} + A(Ar_{k-1}, r_{k-1}, \ldots, r_0) K_k \eta \right\|_{A^{-1}} \\
&= \sqrt{\frac{1+\varrho^2(R)}{\mu_m^2}} \min_{\eta \in \mathbb{R}^k} \left\| r_{k-l_k} + A(r_k, r_{k-1}, \ldots, r_0) L_k \eta \right\|_{A^{-1}} \\
&= \sqrt{\frac{1+\varrho^2(R)}{\mu_m^2}} \min_{\eta \in \mathbb{R}^k} \left\| r_{k-l_k} + A^2 r_{k-1} \zeta^T \eta + \sum_{i=1}^{k} \eta_i A r_{k-i} \right\|_{A^{-1}}, \\
\|e_k\|_{A^T} &\leq \sqrt{\frac{1+\varrho^2(R)}{\mu_m^2}} \min_{\eta \in \mathbb{R}^k} \left\| e_{k-l_k} + A(Ae_{k-1}, e_{k-1}, \ldots, e_0) K_k \eta \right\|_{A^T} \\
&= \sqrt{\frac{1+\varrho^2(R)}{\mu_m^2}} \min_{\eta \in \mathbb{R}^k} \left\| e_{k-l_k} + A(e_k, e_{k-1}, \ldots, e_0) L_k \eta \right\|_{A^T} \\
&= \sqrt{\frac{1+\varrho^2(R)}{\mu_m^2}} \min_{\eta \in \mathbb{R}^k} \left\| e_{k-l_k} + A^2 e_{k-1} \zeta^T \eta + \sum_{i=1}^{k} \eta_i A e_{k-i} \right\|_{A^T}.
\end{aligned}$$

Dabei ist $\varrho(R)$ der Spektralradius des schiefsymmetrischen Anteils R von A^{-1}. μ_m ist der kleinste Eigenwert von M, dem symmetrischen Anteil von A^{-1}. Falls insbesondere A symmetrisch ist, so gilt

$$\begin{aligned}
\|r_k\|_{A^{-1}} &= \min_{\eta \in \mathbb{R}^k} \left\| r_{k-l_k} + A(Ar_{k-1}, r_{k-1}, \ldots, r_0) K_k \eta \right\|_{A^{-1}} \\
&= \min_{\eta \in \mathbb{R}^k} \left\| r_{k-l_k} + A(r_k, r_{k-1}, \ldots, r_0) L_k \eta \right\|_{A^{-1}}
\end{aligned}$$

$$
\begin{aligned}
&= \min_{\eta\in\mathbb{R}^k} \left\| r_{k-l_k} + A^2 r_{k-1}\zeta^T\eta + \sum_{i=1}^{k} \eta_i A r_{k-i} \right\|_{A^{-1}}, \\
\|e_k\|_{A^T} &= \min_{\eta\in\mathbb{R}^k} \left\| e_{k-l_k} + A(Ae_{k-1}, e_{k-1}, \ldots, e_0)K_k\eta \right\|_{A^T} \\
&= \min_{\eta\in\mathbb{R}^k} \left\| e_{k-l_k} + A(e_k, e_{k-1}, \ldots, e_0)K_k\eta \right\|_{A^T} \\
&= \min_{\eta\in\mathbb{R}^k} \left\| e_{k-l_k} + A^2 e_{k-1}\zeta^T\eta + \sum_{i=1}^{k} \eta_i A e_{k-i} \right\|_{A^T}.
\end{aligned}
$$

Hierbei sind $K_k, L_k \in \mathbb{R}^{(k+1)\times k}$ die folgenden Matrizen

$$
K_k = \begin{pmatrix} 0 & 0 & \zeta_3 & \cdots & \zeta_k \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 1 \end{pmatrix} \quad \textit{und} \quad L_k = \begin{pmatrix} 0 & 0 & \zeta_3 & \cdots & \zeta_k \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & * & \cdots & * \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & * & \cdots & * \end{pmatrix} \tag{3.47}
$$

mit $\zeta_i = c_{1,k}^T r_{k-i}$ für $i = 3, \ldots, k$ und $\zeta = (0, 0, \zeta_3, \ldots, \zeta_n)^T \in \mathbb{R}^n$.

Beweis. Um diesen Satz zu beweisen, wird der Ausdruck $\sum_{i=1}^{k} \eta_i A\tilde{Z}_k r_{k-i}$, der in den Aussagen von Korollar 4 auftritt, geeignet umgeformt. Dabei wird die spezielle Form von $\tilde{Z}_k$ ausgenutzt.

Zur Vereinfachung definiert man die beiden folgenden $n\times(k+1)$-Matrizen

$$
\begin{aligned}
R_k &= (r_k, r_{k-1}, \ldots, r_0), \\
Q_k &= (Ar_{k-1}, r_{k-1}, \ldots, r_0).
\end{aligned}
$$

Mit

$$
\tilde{I}_k = \begin{pmatrix} 0 & & \\ 1 & & \\ & \ddots & \\ & & 1 \end{pmatrix} \in \mathbb{R}^{(k+1)\times k}
$$

gilt dann

$$
R_{k-1} = R_k\tilde{I}_k = Q_k\tilde{I}_k. \tag{3.48}
$$

Weiterhin sei

$$
E_k = \begin{pmatrix} 0 & 0 & \varepsilon_{3,k}^1 & \cdots & \varepsilon_{k,k}^1 \\ 0 & 0 & \varepsilon_{3,k}^2 & \cdots & \varepsilon_{k,k}^2 \\ 0 & 0 & \varepsilon_{3,k}^3 & \cdots & \varepsilon_{k,k}^3 \end{pmatrix} := C_k^T R_{k-1}. \tag{3.49}
$$

Die ersten beiden Spalten von E_k sind Null wegen $C_k^T r_{k-1} = C_k^T r_{k-2} = 0$.

Nach Korollar **??** und wegen der Darstellung (3.29) von r_k kann man Ar_{k-1} darstellen als

$$Ar_{k-1} = \sum_{j=0}^{k} \xi_j r_{k-j}$$

mit gewissen $\xi_j \in \mathbb{R}$. Dabei gilt wegen (3.29) $\xi_0 \neq 0$. Daraus folgt

$$Q_k = R_k J_k \tag{3.50}$$

mit

$$J_k = \begin{pmatrix} \xi_0 & & & \\ \xi_1 & 1 & & \\ \vdots & & \ddots & \\ \xi_k & & & 1 \end{pmatrix} \in \mathrm{GL}_{k+1}(\mathbb{R})\,.$$

Das Ziel ist nun, die Vektoren $\tilde{Z}_k r_{k-i}$ für $i = 1, \ldots, k$ mit Hilfe von R_k bzw. Q_k darzustellen. Wegen $B_k = (Ar_{k-1}, r_{k-1}, r_{k-2})$ gilt

$$\begin{aligned} B_k &= Q_k I_{3,k} && (3.51)\\ &\overset{(3.50)}{=} R_k J_k I_{3,k} && (3.52) \end{aligned}$$

mit

$$I_{3,k} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 & 0 \end{pmatrix} \in \mathbb{R}^{(k+1)\times 3}\,. \tag{3.53}$$

Um später die spezielle Form von K_k bzw. L_k erreichen zu können, muß man noch erwähnen, daß man bei beiden beliebige Spaltenumformungen machen darf. Das folgt aus der Tatsache, daß

$$\{\, \eta \mid \eta \in \mathbb{R}^k \,\} = \{\, W\eta \mid \eta \in \mathbb{R}^k \,\} \tag{3.54}$$

für alle $W \in \mathrm{GL}_k(\mathbb{R})$ gilt. Damit kann man nun den Ausdruck aus Korollar 4 umformen.

$$\begin{aligned} \left\{ \sum_{i=1}^{k} \eta_i A\tilde{Z}_k r_{k-i} \,\middle|\, \eta_1, \ldots, \eta_k \in \mathbb{R} \right\} &= \{\, A\tilde{Z}_k R_{k-1}\eta \mid \eta \in \mathbb{R}^k \,\} \\ &\overset{(\mathbf{??})}{=} \{\, A(I + B_k C_k^T) R_{k-1}\eta \mid \eta \in \mathbb{R}^k \,\} \\ &\overset{(3.49)}{=} \{\, A(R_{k-1} + B_k E_k)\eta \mid \eta \in \mathbb{R}^k \,\} \end{aligned}$$

$$
\begin{aligned}
&\overset{(3.48)}{=} \{\, A(Q_k\tilde I_k + Q_k I_{3,k}E_k)\eta \mid \eta \in \mathbb{R}^k \,\} \\
&= \{\, AQ_k(\tilde I_k + I_{3,k}E_k)\eta \mid \eta \in \mathbb{R}^k \,\} \\
&= \{\, AQ_k \underbrace{(\tilde I_k + I_{3,k}E_k)W}_{=:K_k} \eta \mid \eta \in \mathbb{R}^k \,\}
\end{aligned}
$$

für beliebiges $W \in \mathrm{GL}_k(\mathbb{R})$. Da der Ausdruck in der Klammer die Form

$$
\tilde I_k + I_{3,k}E_k = \begin{pmatrix} 0 & 0 & \varepsilon^1_{3,k} & \cdots & \varepsilon^1_{k,k} \\ 1 & 0 & \varepsilon^2_{3,k} & \cdots & \varepsilon^2_{k,k} \\ & 1 & \varepsilon^3_{3,k} & \cdots & \varepsilon^3_{k,k} \\ & & 1 & & \\ & & & \ddots & \\ & & & & 1 \end{pmatrix}
$$

hat, erreicht man mit der Wahl

$$
W = \begin{pmatrix} 1 & 0 & -\varepsilon^2_{3,k} & \cdots & -\varepsilon^2_{k,k} \\ & 1 & -\varepsilon^3_{3,k} & \cdots & -\varepsilon^3_{k,k} \\ & & 1 & & \\ & & & \ddots & \\ & & & & 1 \end{pmatrix} \in \mathrm{GL}_k(\mathbb{R})\,,
$$

daß K_k die geforderte Form (3.47) hat. Dabei beachte man, daß nach Definition $\zeta_i = \varepsilon^1_{i,k}$ für $i = 1, \ldots, k$ gilt.

Die Behauptung mit L_k folgt aus der analogen Rechnung wegen

$$
\begin{aligned}
A\tilde Z_k R_{k-1} &= A(I + B_k C_k^T)R_{k-1} \\
&= A(R_{k-1} + B_k E_k) \\
&= A(R_k \tilde I_k + R_k J_k I_{3,k} E_k) \\
&= AR_k(\tilde I_k + J_k I_{3,k} E_k)
\end{aligned}
$$

Wählt man dann das gleiche W wie oben, so erhält man die entsprechende Aussage mit einer Matrix von der Form von L_k.

Die jeweils dritten Aussagen folgen sofort aus den Aussagen mit K_k. Wegen dessen spezieller Form gilt

$$
K_k = \tilde I_k + e^1_{k+1}\zeta^T\,,
$$

wobei $e^1_{k+1} = (1, 0, \ldots, 0)^T \in \mathbb{R}^{k+1}$ der erste Einheitsvektor ist. Daraus folgt dann wegen $Q_k e^1_{k+1} = Ar_{k-1}$

$$
Q_k K_k \eta \;=\; Q(\tilde I_k + e^1_{k+1}\zeta^T)\eta
$$

$$\stackrel{(3.48)}{=} R_{k-1}\eta + \zeta^T \eta A r_{k-1}$$

und damit die Behauptung. □

Entsprechend kann man die Aussagen des Korollars 4 auch für das Verfahren aus Algorithmus 2 spezialisieren.

Satz 6
Falls A positiv reell ist, d. h. der symmetrische Anteil von A positiv definit ist, so gilt für das im Algorithmus 2 angegebene Verfahren

$$\begin{aligned}
\|r_k\|_{A^{-1}} \leq & \sqrt{\frac{1+\varrho^2(R)}{\mu_m^2}} \\
& \cdot \min_{\eta \in \mathbb{R}^k} \left\| r_{k-1} + A^2 r_{k-1} \zeta^T \eta + A(r_{k-1}, x_{k-1} - x_{k-2}, r_{k-3}, \ldots, r_0)\eta \right\|_{A^{-1}} \\
\|e_k\|_{A^T} \leq & \sqrt{\frac{1+\varrho^2(R)}{\mu_m^2}} \\
& \cdot \min_{\eta \in \mathbb{R}^k} \left\| e_{k-l_k} + A^2 e_{k-1} \zeta^T \eta + (Ae_{k-1}, e_{k-1} - e_{k-2}, Ae_{k-3}, \ldots, Ae_0)\eta \right\|_{A^T} .
\end{aligned}$$

Dabei ist $\varrho(R)$ der Spektralradius des schiefsymmetrischen Anteils R von A^{-1}. μ_m ist der kleinste Eigenwert von M, dem symmetrischen Anteil von A^{-1}. Weiterhin ist $\zeta = (0, 0, \zeta_3, \ldots, \zeta_n)^T \in \mathbb{R}^n$ mit $\zeta_i = c_{1,k}^T r_{k-i}$ für $i = 3, \ldots, k$. Falls insbesondere A symmetrisch ist, so gilt

$$\begin{aligned}
\|r_k\|_{A^{-1}} &= \min_{\eta \in \mathbb{R}^k} \left\| r_{k-1} + A^2 r_{k-1} \zeta^T \eta + A(r_{k-1}, x_{k-1} - x_{k-2}, r_{k-3}, \ldots, r_0)\eta \right\|_{A^{-1}} \\
\|e_k\|_{A^T} &= \min_{\eta \in \mathbb{R}^k} \left\| e_{k-l_k} + A^2 e_{k-1} \zeta^T \eta + (Ae_{k-1}, e_{k-1} - e_{k-2}, Ae_{k-3}, \ldots, Ae_0)\eta \right\|_{A^T} .
\end{aligned}$$

Beweis. Analog zum Beweis von Satz 5. □

Kapitel 4

Numerische Testrechnungen

Für die Testrechnungen wird die dreidimensionale partielle Differentialgleichung

$$\Delta v + v + \varrho(v^T \nabla)v = h \tag{4.1}$$

mit $v : \mathbb{R}^3 \to \mathbb{R}^3$ betrachtet (vergleiche [11, Kapitel 6]). Hierbei ist $\nabla = (\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z})^T$ und $\Delta = \nabla^T \nabla$. Die rechte Seite h wird so gewählt, daß sich als Lösung

$$v(x,y,z) = \begin{pmatrix} \sin(\pi x)\cos(\pi y)\cos(\pi z) \\ \cos(\pi x)\sin(\pi y)\cos(\pi z) \\ \cos(\pi x)\cos(\pi y)\sin(\pi z) \end{pmatrix} \tag{4.2}$$

ergibt. Diese partielle Differentialgleichung ist eine grobe Näherung für die Navier-Stokes-Gleichungen. Sie wird auf dem Einheitswürfel mit Finite-Differenzen-Methoden mit Konsistenzordnung 2 auf einem $20 \times 20 \times 20$-Gitter diskretisiert und im ersten Schritt des Newton-Verfahrens linearisiert. Die Matrix ist so normalisiert, daß die Betragssummen der Zeilenvektoren Eins sind und die Hauptdiagonaleinträge nichtnegativ.

Das so entstehende lineare Gleichungssystem hat die Dimension $n = 24\,000$. Der Parameter ϱ simuliert eine Reynolds-Zahl. Mit wachsendem ϱ nimmt der schiefsymmetrische Anteil der Matrix zu. Dadurch wird die Konvergenzrate von verallgemeinerten cg-Verfahren i. a. vermindert, was die Konvergenzaussagen von Satz **??** ja nahelegen.

Gegeneinander getestet werden die folgenden Verfahren:

- ORTHORES als exaktes Verfahren.
- Das klassische cg-Verfahren, also ein abgeschnittenes ORTHORES mit $\sigma_{\max} = 2$.
- Ein kombiniertes ORTHORES-Verfahren mit $\sigma_{\max} = 5$ und $\sigma_{\text{per}} = 20$. Dieses Verfahren wird PRES20 [10] genannt.

- Das in Algorithmus 1 vorgestellte Verfahren.
- Das in Algorithmus 2 vorgestellte Verfahren.
- Das in Algorithmus 1 vorgestellte Verfahren als neugestartete Version mit $\sigma_{\text{per}} = 20$ (in Analogie zu PRES20).
- Das in Algorithmus 2 vorgestellte Verfahren als neugestartete Version mit $\sigma_{\text{per}} = 20$.

Bei allen neugestarteten Verfahren wird das geglättete x_k^s nach dem Neustart als Startvektor verwendet. Das exakte ORTHORES-Verfahren wurde als Referenzverfahren verwendet, weil es das Optimum darstellt, was mit einem verallgemeinerten cg-Verfahren zu erreichen ist, d. h. es ist mindestens ebenso gut wie alle anderen verallgemeinerten cg-Verfahren.

Das klassische cg-Verfahren ist wie die beiden Algorithmen 1 und 2 auch ein zweistufiges abgeschnittenes Verfahren, hat also den gleichen Rechenaufwand wie diese.

Das PRES20-Verfahren ist ein Verfahren aus der Praxis, das sich als besonders schnell herausgestellt hat. Die Parameter $\sigma_{\max} = 5$ und $\sigma_{\text{per}} = 20$ wurden durch aufwendige Tests so bestimmt, daß eine optimale Abwägung zwischen Aufwand und Leistungsfähigkeit des Verfahrens gewährleistet wird. PRES20 ist also im Sinne der Praxistauglichkeit ein „optimales" abgeschnittenes verallgemeinertes cg-Verfahren.

Aufgetragen wird die relative Euklidnorm der Residuen $\|r_k\|_2 \,/\, \|r_0\|_2$ über der Anzahl der Iterationsschritte, was gleichbedeutend mit der Anzahl der Matrix-Vektor-Multiplikationen (MVM) ist, da alle hier behandelten Verfahren mit einer MVM pro Iterationsschritt auskommen.

Teilweise werden auch die relativen Euklidnormen der geglätteten Residuen s_k aufgetragen. Die geglätteten Iterierten x_k^s und die zugehörigen Residuen $s_k = Ax_k^s - b$ werden durch den Glättungsalgorithmus von Schönauer [10, 14] berechnet, durch den die Folge der Euklidnormen der s_k monoton fallend wird und außerdem immer unterhalb der Euklidnorm der echten Residuen r_k liegt, d. h. es gilt für alle $k \geq 1$

$$\|s_k\|_2 \leq \min(\|r_k\|_2\,, \|s_{k-1}\|_2)\,. \tag{4.3}$$

Da das PRES20-Verfahren nach Definition schon ein geglättetes Verfahren ist, ergibt sich durch nochmaliges nachträgliches Glätten kein Unterschied mehr.

4.1 $\varrho = 1$

Im Falle $\varrho = 1$ ist der schiefsymmetrische Anteil der Matrix A sehr klein. Es ist also zu erwarten, daß alle Verfahren sehr gut konvergieren. In der Abbildung 4.1 sind die relativen Residuumsnormen für die beiden Algorithmen 1 und 2, jeweils mit und ohne Neustart, aufgetragen.

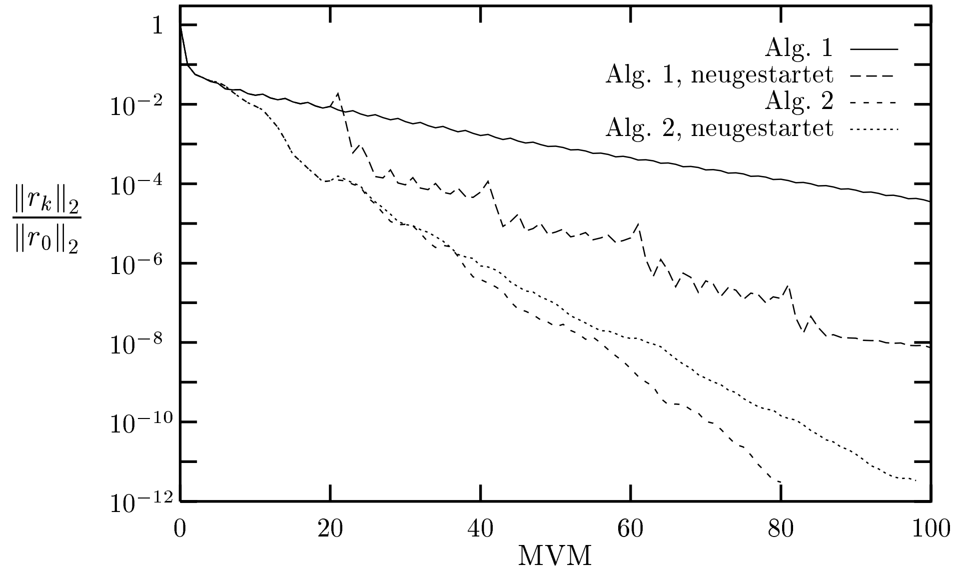

Abbildung 4.1: $\varrho = 1$

Man erkennt, daß bei diesem linearen Gleichungssystem der Algorithmus 1 *mit* Neustart besser läuft. Beim zweiten Algorithmus hingegen ist die Variante *ohne* Neustart besser. Deshalb werden in der Abbildung 4.2 die jeweils besseren Varianten von Algorithmus 1 bzw. 2 mit den bekannten Verfahren klassisches cg, exaktes ORTHORES und PRES20 verglichen.

In Abbildung 4.2 sind die geglätteten Residuen aufgetragen, um einen besseren Vergleich mit dem ebenfalls geglätteten PRES20-Verfahren zu ermöglichen.

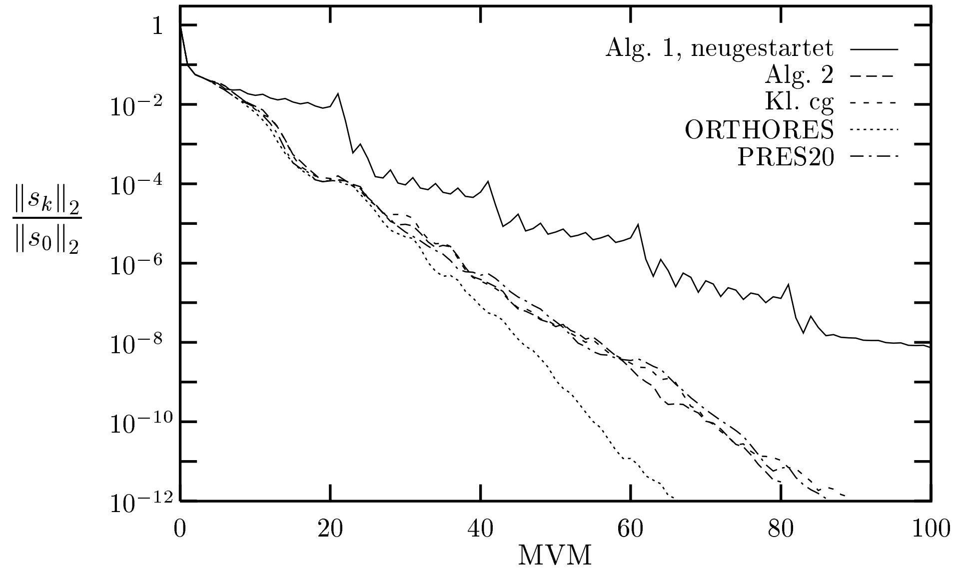

Abbildung 4.2: $\varrho = 1$, geglättete Residuen

Wie zu erwarten ist, läuft das exakte ORTHORES-Verfahren am besten. Die anderen Verfahren laufen fast alle gleich gut, wobei nur die neugestartete Version von Algorithmus 1 negativ aus dem Rahmen fällt.

4.2 $\varrho = 100$

Auch für $\varrho = 100$ werden zunächst die neuen Verfahren verglichen. Man

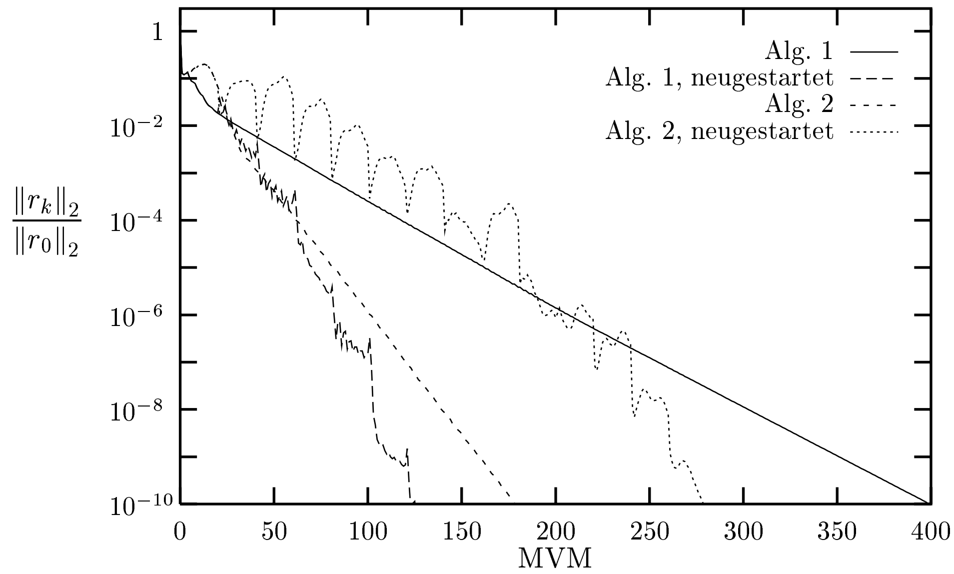

Abbildung 4.3: $\varrho = 100$

erkennt hier ebenfalls, daß sich nur der erste Algorithmus durch einen Neustart nach 20 Schritten verbessert. In Abbildung 4.4 zeigt sich im Vergleich mit den anderen Verfahren, daß beide Algorithmen mit PRES20 mithalten können. Das klassische cg-Verfahren ist schon deutlich langsamer.

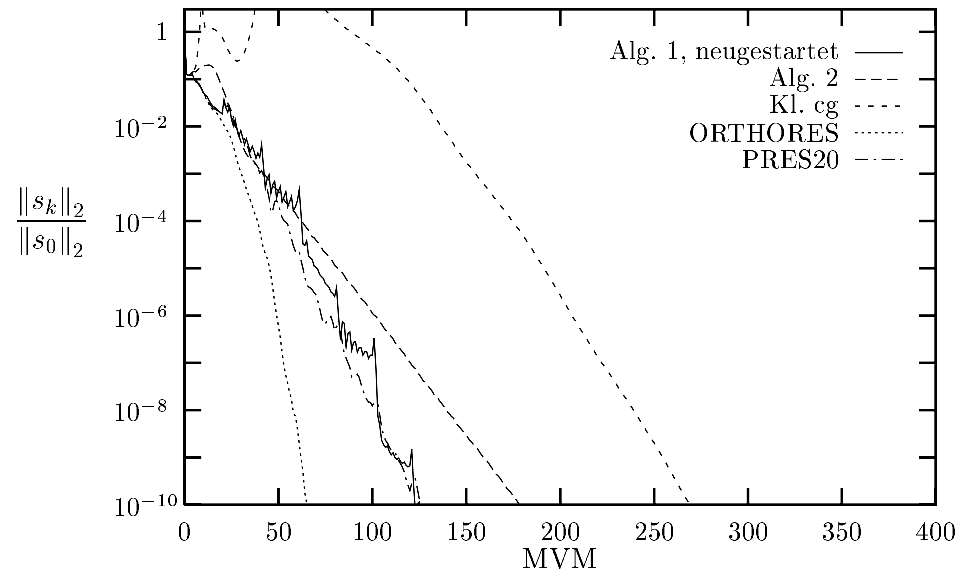

Abbildung 4.4: $\varrho = 100$, geglättete Residuen

4.3 $\varrho = 1000$

Bei $\varrho = 1000$ erreichen die abgeschnittenen Verfahren bei weitem nicht das Konvergenzverhalten des exakten ORTHORES-Verfahrens. In Abbildung 4.5 sind für alle getesteten Verfahren die relativen Normen der geglätteten Residuen für die ersten 400 Iterationsschritte aufgetragen.

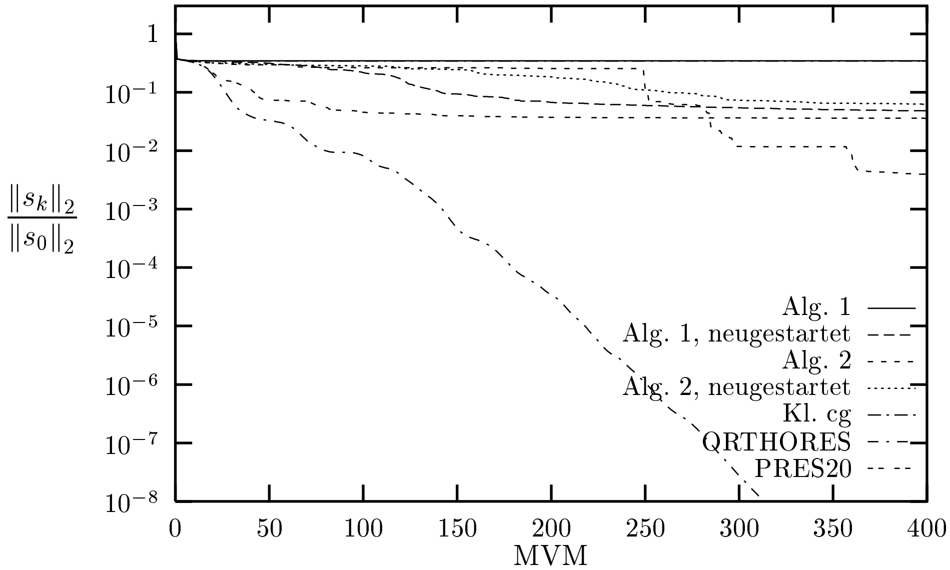

Abbildung 4.5: $\varrho = 1000$, geglättete Residuen

Die Kurven vom klassischen cg-Verfahren und vom Algorithmus 1 ohne Neustart liegen aufeinander. Beide Verfahren „schlafen“ nach den ersten 10 Iterationen ein. Dies liegt daran, daß die ungeglätteten Residuen dieser Verfahren divergieren.

Um die Unterschiede zwischen den abgeschnittenen Verfahren genauer erkennen zu können, sind in Abbildung 4.6 die ersten 2000 Iterationen eingezeichnet. Es zeigt sich, daß außer dem Algorithmus 2 ohne Neustart alle anderen Verfahren „einschlafen“.

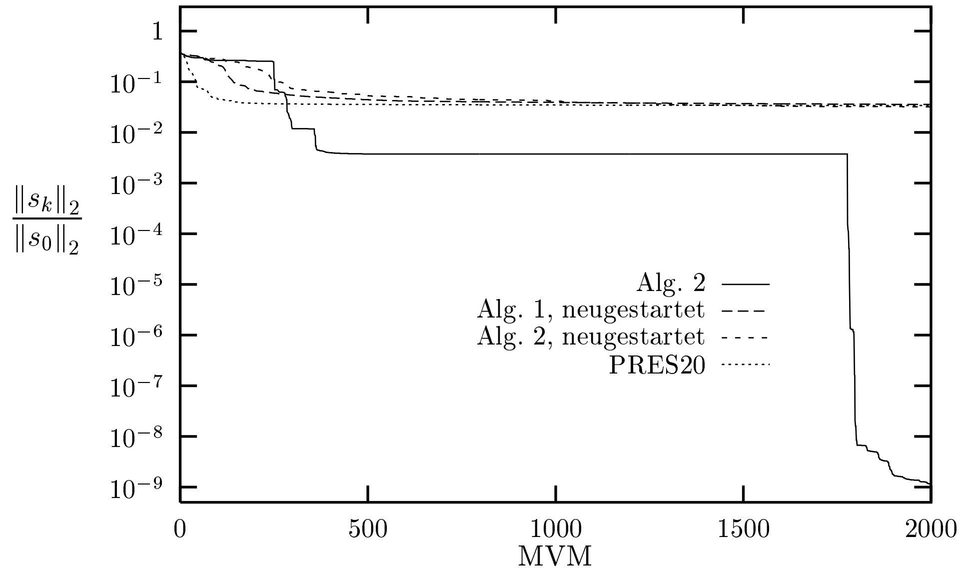

Abbildung 4.6: $\varrho = 1000$, geglättete Residuen

Das Verhalten von Algorithmus 2 wird deutlicher, wenn man statt den geglätteten die ungeglätteten Residuen aufträgt.

Die Kurve von Algorithmus 2 ähnelt einer Sägezahnkurve. Warum das Verfahren nach ungefähr 1700 Schritten so plötzlich „einrastet“, ist nicht zu erklären.

Man erkennt ganz links, daß das Residuum vom klassischen cg-Verfahren schon nach wenigen Schriten eine Euklidnorm von 10^8 erreicht, d. h. das Verfahren divergiert.

Betrachtet man sogar 4000 Iterationen, so sieht man, daß auch PRES20 konvergiert. Hier sind wieder die geglätteten Residuen aufgetragen.

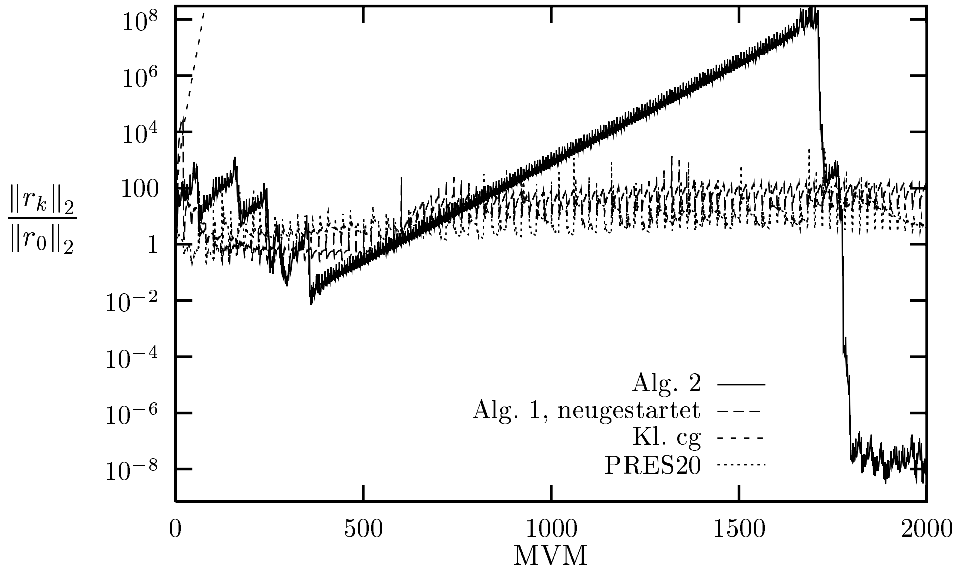

Abbildung 4.7: $\varrho = 1000$

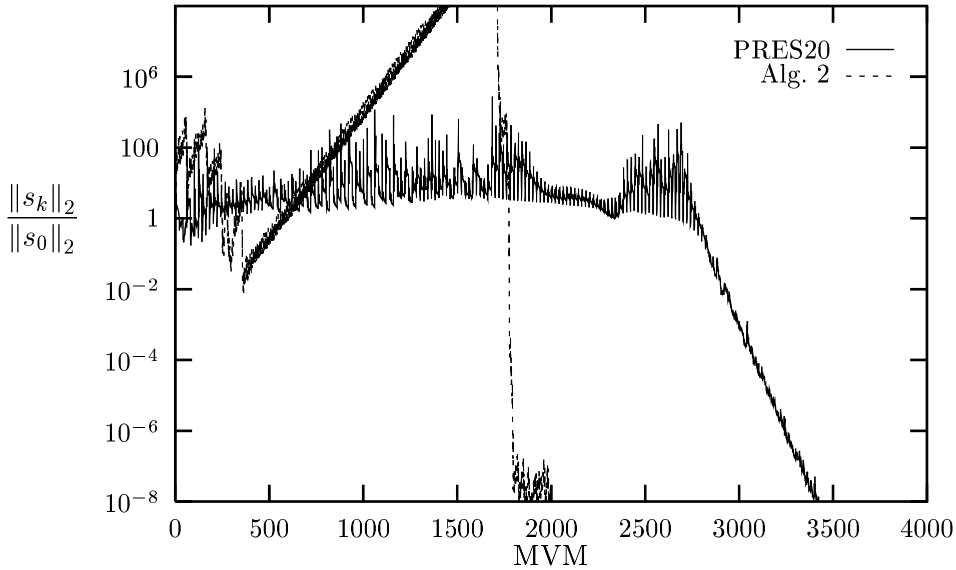

Abbildung 4.8: $\varrho = 1000$, geglättete Residuen

4.4 $\varrho = 10000$

Für $\varrho = 10000$ ist die Iterationsmatrix A nicht mehr positiv reell, d. h. die Konvergenzsätze gelten nicht mehr. In Abbildung 4.9 erkennt man, daß auch das exakte ORTHORES-Verfahren in den ersten 400 Schritten keine merkliche Verbesserung erzielen kann. Damit ist für die abgeschnittenen Verfahren auch keine Konvergenz zu erwarten.

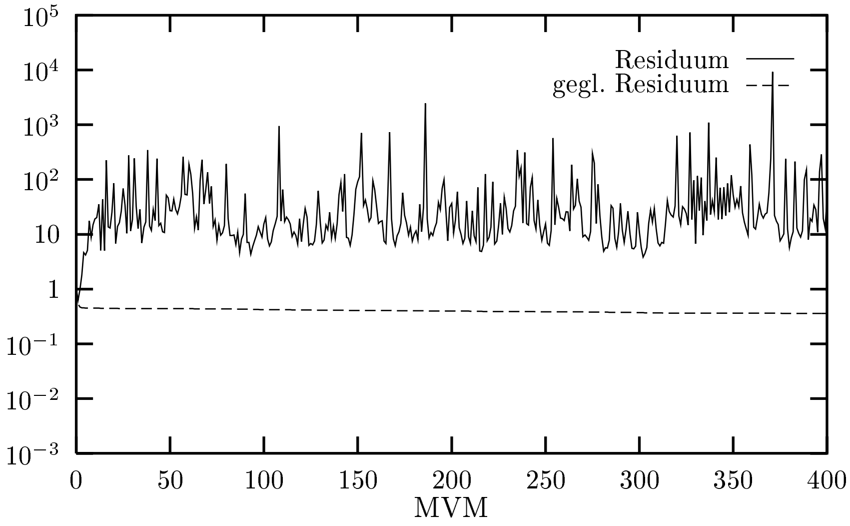

Abbildung 4.9: $\varrho = 10^4$, ORTHORES-Verfahren

Kapitel 5

Conclusion

Im ersten Kapitel werden die verallgemeinerten cg-Verfahren und als Spezialfall davon das klassische cg-Verfahren von Hestenes und Stiefel vorgestellt. Für exakte verallgemeinerte cg-Verfahren wird eine Minimierungseigenschaft über dem Krylowraum $K_{k-1}(A, r_0)$ in der A^{-1}-Norm bewiesen.

Durch eine Verallgemeinerung der verallgemeinerten cg-Verfahren erhält man im zweiten Kapitel die CKS-Verfahren. Es wird gezeigt, daß diese ebenfalls eine Minimierungseigenschaft über dem Krylowraum $K_{k-1}(A, r_0)$ erfüllen, jedoch im allgemeinen nicht in der festen A^{-1}-Norm, sondern in einer schrittabhängigen $Z_k A^{-1}$-Norm.

Um dieses Manko zu umgehen, wird in den folgenden Kapiteln auf eine Theorie für schrittabhängig präkonditionierte CKS-Verfahren hingearbeitet. Dazu werden im vierten Kapitel nach einer kurzen Einführung in die Präkonditionierung die Orthogonalisierungsverfahren vorgestellt. Sie sind wiederum eine Verallgemeinerung der CKS-Verfahren. Sie werden benötigt, da man bei der Rückformulierung von schrittabhängig präkonditionierten CKS-Verfahren auf das ursprüngliche System keine CKS-Verfahren mehr erhält, sondern spezielle Orthogonalisierungsverfahren.

Die im Kapitel 5 erzielten Ergebnisse über die Rückformulierung von präkonditionierten Orthogonalisierungsverfahren werden in Kapitel 6 auf den Spezialfall der CKS-Verfahren angewendet. Es wird gezeigt, daß man verschiedene Verfahren erhalten kann, wenn man auf dem präkonditionierten System verschiedene Ansätze verwendet. Zwei spezielle Verfahren werden exemplarisch ausgewählt und durch geeignete Wahl der Präkonditionierungsmatrizen gelingt es, für beide Minimierungseigenschaften in der A^{-1}-Norm zu erhalten.

In Kapitel 7 werden die Bedingungen für eine kurze Rekursion bei den beiden Verfahren aufgestellt. Im darauffolgenden Kapitel wird gezeigt, daß sich diese Bedingungen für das erste Verfahren nicht mit sinnvollem Aufwand erfüllen lassen.

Für das zweite Verfahren hingegen werden in Kapitel 9 zwei Varianten konstruiert, die beide die Bedingungen für eine kurze (zweistufige) Rekursion erfüllen. Für diese Varianten werden Algorithmen angegeben. Es zeigt sich, daß beide Verfahren eine Minimierungseigenschaft in der A^{-1}-Norm über einem k-dimensionalen Unterraum besitzen. Außerdem stellt sich heraus, daß die so konstruierten Verfahren äquivalent zu speziellen abgeschnittenen verallgemeinerten cg-Verfahren sind. Damit können für diese abgeschnittenen Verfahren erstmals Minimierungseigenschaften in einem k-dimensionalen Unterraum angegeben werden.

In Kapitel 10 werden die zwei konstruierten Algorithmen mit anderen verallgemeinerten cg-Verfahren verglichen. Es zeigt sich, daß einer davon sogar mit dem abgeschnittenen Verfahren PRES20 konkurrieren kann, das statt einer zweistufigen Rekursion eine fünfstufige besitzt, also in jedem Schritt mehr Informationen verwendet.

5.1 Ausblick

Außer den in Kapitel 1 definierten Verfahren 1 und 2 gibt es noch viele andere präkonditionierte CKS-Verfahren. Diese erhält man, indem man auf dem präkonditionierten System andere Ansätze für das dort verwendete CKS-Verfahren benutzt. Prinzipiell besteht auch für diese pCKS-Verfahren die Möglichkeit, sie mit einer kurzen Rekursion zu formulieren.

Statt der hier erreichten Minimierungseigenschaft in der A^{-1}-Norm könnte man auch versuchen, eine solche in anderen, festen Normen wie z. B. der Euklidnorm zu bekommen. Auch fehlerminimierende Verfahren sind denkbar, bei denen die Euklidnorm des Fehlers e_k minimiert wird.

Aber auch für Verfahren 2 sind in dieser Arbeit die Möglichkeiten noch nicht ausgeschöpft worden. Es gibt außer den drei verwendeten Ansätzen zur Erfüllung der Bedingungen für die kurze Rekursion eventuell noch weitere, die dann zu anderen Verfahren führen.

Es existieren also noch viele Möglichkeiten, aus der Klasse der präkonditionierten CKS-Verfahren neue Erkenntnisse für die Theorie und neue Verfahren für die Praxis zu gewinnen.

Literaturverzeichnis

[1] T. Barth and T. Manteuffel. Estimating the spectrum of A using the roots of the polynomials associated with the QMR iteration. *SIAM J. Matrix. Anal.*, March 1993. submitted to.

[2] T. Barth and T. Manteuffel. Variable metric conjugate gradient methods. In N. Natori and T. Nodera, editors, *Advances in Numerical Methods for Large Sparse Sets of Linear Equations, No 10, Matrix Analysis and Parallel Computing, PCG '94*, pages 165–188. Keio University, Yokohama, Japan, 1994.

[3] V. Faber and T. Manteuffel. Necessary and sufficient conditions for the existence of a conjugate gradient method. *SIAM J. Numer. Anal.*, 21:352–362, 1984.

[4] V. Faber and T. A. Manteuffel. Orthogonal error methods. *SIAM J. Numer. Anal.*, 24(1):170–187, 1987.

[5] S. Fujino and S. Doi. Optimizing multicolor ICCG methods on some vectorcomputers. In R. Beauwens and P. de Groen, editors, *Iterative Methods in linear Algebra*, pages 349–358. Elservier Science Publishers B.V. (North-Holland), 1992.

[6] S. Fujino, M. Mori, and T. Takeuchi. Performance of hyperplane ordering on vector computers. *Journal of Computational and Applied Mathematics*, 38:125–136, 1991.

[7] M.H. Gutknecht. Stationary and almost stationary (k,l)-step methods for linear and nonlinear systems of equations. *Numer. Math.*, 56:179–213, 1989.

[8] Günther Hämmerlein and Karl-Heinz Hoffmann. *Numerische Mathematik*. Grundwissen Mathematik 7. Springer-Verlag, Berlin, Heidelberg, New York, Tokyo, second edition, 1989.

[9] Max Koecher. *Lineare Algebra und analytische Geometrie*. Springer-Verlag, Berlin, Heidelberg, New York, Tokyo, second edition, 1985.

[10] W. Schönauer. *Scientific Computing on Vector Computers*. North-Holland, Amsterdam, New York, Oxford, Tokyo, 1987.

[11] R. Weiss. Minimization properties and short recurrences for Krylov subspace methods. *Electronic Transactions on Numerical Analysis (ET-NA)*, 2:57–75, 1994.

[12] R. Weiss. Orthogonalization methods. Internal report 52/94, University of Karlsruhe, Computing Center, 1994.

[13] R. Weiss, H. Häfner, and W. Schönauer. Tuning the matrix-vector multiplication in diagonal form. In G.R. Joubert, D.J. Evans, and F.J. Peters, editors, *Proceedings of Parallel Computing 89, Leiden*, pages 93–98. Elsevier Science Publishers B. V., North-Holland, 1990.

[14] R. Weiss and W. Schönauer. Accelerating generalized conjugate gradient methods by smoothing. In R. Beauwens and P. de Groen, editors, *Iterative Methods in Linear Algebra*, pages 283–292. North-Holland, 1992.