

Numerische Lineare Algebra

Oliver Ernst

Professur Numerische Mathematik

Wintersemester 2017/18



Mathematik!
TU Chemnitz

① Einleitung

- 1.1 Lineare Gleichungssysteme
- 1.2 Matrixfunktionen
- 1.3 Modellreduktion
- 1.4 Eigenwertaufgaben

② Krylov-Unterraumverfahren

- 2.1 Projektionen
- 2.2 Orthogonale Projektionsverfahren
- 2.3 Krylov-Unterräume
- 2.4 Basen von Krylov-Räumen

③ Lineare Gleichungssysteme

- 3.1 Lösungsstrategien
- 3.2 Selbstadjungierte Probleme
- 3.3 Das Verfahren der konjugierten Gradienten
- 3.4 Krylov-Verfahren für lineare Ausgleichsprobleme
- 3.5 Vorkonditionierung

④ Matrixfunktionen

- 4.1 Erste Definition mithilfe der Jordanschen Normalform

- 4.2 Hermitesche Polynominterpolation
- 4.3 Alternative Darstellungen von Matrixfunktionen
- 4.4 Resolventenintegrale
- 4.5 Ein Beispiel

5 Krylov-Verfahren für Matrixfunktionen

- 5.1 Schranken für $\|f(\mathbf{A})\|$
- 5.2 Fehlerschranken für Krylov-Verfahren
- 5.3 Die Konvergenz des CG-Verfahrens

6 Das QR-Verfahren für Eigenwertaufgaben

- 6.1 Reduktion auf Hessenberg-Gestalt
- 6.2 Vektoriteration
- 6.3 QR-Iteration

- ① Einleitung
- ② Krylov-Unterraumverfahren
- ③ Lineare Gleichungssysteme
- ④ Matrixfunktionen
- ⑤ Krylov-Verfahren für Matrixfunktionen**
- ⑥ Das QR-Verfahren für Eigenwertaufgaben

Seien $\mathbf{A} \in \mathbb{C}^{N \times N}$, $\mathbf{b} \in \mathbb{C}^N$ und f eine Funktion, für die $f(\mathbf{A})$ definiert ist.

- In Definition 2.12 wurde die Rayleigh-Ritz-Approximation an $f(\mathbf{A})$ aus einem Unterraum eingeführt.
- Wählt man als Unterraum den Krylov-Raum $\mathcal{K}_m(\mathbf{A}, \mathbf{b})$ mit der Arnoldi-Zerlegung (Lanczos-Zerlegung)

$$\mathbf{A}\mathbf{V}_m = \mathbf{V}_m\mathbf{H}_m + h_{m+1,1m}\mathbf{v}_{m+1}\mathbf{e}_m^\top,$$

so ist diese Rayleigh-Ritz-Approximation gegeben durch

$$\mathbf{f}_m := \mathbf{V}_m f(\mathbf{V}_m^\mathbf{H}\mathbf{A}\mathbf{V}_m) \mathbf{V}_m^\mathbf{H}\mathbf{b} = \beta \mathbf{V}_m f(\mathbf{H}_m) \mathbf{e}_1, \quad (\beta = \|\mathbf{b}\|_2). \quad (5.1)$$

- Wir nennen \mathbf{f}_m die (m -te) **Arnoldi-Approximation** bzw. **Lanczos-Approximation** an $f(\mathbf{A})\mathbf{b}$.
- Im Spezialfall $f(\lambda) = 1/\lambda$ ist $\mathbf{f}_m = \beta \mathbf{V}_m \mathbf{H}_m^{-1} \mathbf{e}_1$ die m -te OR-Approximation aus $\mathcal{K}_m(\mathbf{A}, \mathbf{b})$ (die m -te CG Iterierte, falls \mathbf{A} Hermitesch und positiv definit ist) an die Lösung von $\mathbf{A}\mathbf{x} = \mathbf{b}$ (vgl. Satz 3.4).

Beachte: aus der Tatsache, dass $f(\mathbf{A})$ definiert ist, folgt keineswegs, dass auch

$$f(\mathbf{V}_m^H \mathbf{A} \mathbf{V}_m) = f(\mathbf{H}_m)$$

definiert ist.

Hinreichend hierfür ist z.B. dass f analytisch im Innern des Wertebereichs $W(\mathbf{A})$ von \mathbf{A} und auf $W(\mathbf{A})$ stetig ist; in diesem Fall ist $f(\mathbf{H}_m)$ definiert, denn

$$W(\mathbf{H}_m) = W(\mathbf{V}_m^H \mathbf{A} \mathbf{V}_m) \subseteq W(\mathbf{A}).$$

Wenn in der Folge $f(\mathbf{A})$ bzw. $f(\mathbf{H}_m)$ erwähnt wird, setzen wir immer voraus, dass diese Matrizen definiert sind.

Satz 5.1

Für die Arnoldi-Approximation \mathbf{f}_m aus (5.1) gilt

$$\mathbf{f}_m = p_{m-1}(\mathbf{A})\mathbf{b}. \quad (5.2)$$

Dabei bezeichnet p_{m-1} das Polynom vom Grad $m - 1$, welches f an den Nullstellen des charakteristischen Polynoms von \mathbf{H}_m im Hermiteschen Sinn interpoliert. Ist Γ eine einfach geschlossene, positiv orientierte Jordan-Kurve, die alle Eigenwerte sowohl von \mathbf{A} als auch von \mathbf{H}_m im Innern enthält, und ist f analytisch im Innengebiet $\text{int } \Gamma$ sowie stetig in $\overline{\text{int } \Gamma}$, so gilt

$$\mathbf{f}_m = \frac{1}{2\pi i} \int_{\Gamma} f(\lambda) \mathbf{x}_m(\lambda) d\lambda. \quad (5.3)$$

Dabei ist

$$\mathbf{x}_m(\lambda) = \beta \mathbf{V}_m(\lambda \mathbf{I}_m - \mathbf{H}_m)^{-1} \mathbf{e}_1$$

die OR-Approximation aus $\mathcal{K}_m(\mathbf{A}, \mathbf{b})$ an die Lösung von $(\lambda \mathbf{I} - \mathbf{A})\mathbf{x} = \mathbf{b}$.

- 1 Einleitung
- 2 Krylov-Unterraumverfahren
- 3 Lineare Gleichungssysteme
- 4 Matrixfunktionen
- 5 Krylov-Verfahren für Matrixfunktionen
 - 5.1 Schranken für $\|f(\mathbf{A})\|$
 - 5.2 Fehlerschranken für Krylov-Verfahren
 - 5.3 Die Konvergenz des CG-Verfahrens
- 6 Das QR-Verfahren für Eigenwertaufgaben

Krylov-Verfahren für Matrixfunktionen

Schranken für $\|f(\mathbf{A})\|$

Die erste Schranke, die wir für $\|f(\mathbf{A})\|$ beweisen werden, folgt unmittelbar aus der Definition einer Matrixfunktion mithilfe der Jordanschen Normalform (vgl. Definition 4.3).

Satz 5.2

Es sei $\mathbf{A} \in \mathbb{C}^{N \times N}$ mit Minimalpolynom

$$m_{\mathbf{A}}(\lambda) = \prod_{\mu=1}^k (\lambda - \lambda_{\mu})^{n_{\mu}}, \quad \lambda_i \neq \lambda_j \text{ für } i \neq j,$$

und Jordan-Normalform $\mathbf{J} = \mathbf{T}^{-1} \mathbf{A} \mathbf{T}$. Mit $\|\cdot\|_p$ bezeichnen wir die Hölder- p -Norm bzw. die davon induzierte Matrixnorm ($1 \leq p \leq \infty$). Dann gilt

$$\|f(\mathbf{A})\|_p \leq n_{\max} \operatorname{cond}_p(\mathbf{T}) \max_{\substack{\mu=1,2,\dots,k \\ j=0,1,\dots,n_{\mu}-1}} \frac{|f^{(j)}(\lambda_{\mu})|}{j!}$$

mit $n_{\max} := \max\{n_{\mu} : \mu = 1, 2, \dots, k\}$.

Krylov-Verfahren für Matrixfunktionen

Schranken für $\|f(\mathbf{A})\|$

Unmittelbare Folgerungen aus Satz 5.2 sind:

(1) Ist \mathbf{A} diagonalisierbar, d.h. ist $\mathbf{T}^{-1}\mathbf{A}\mathbf{T}$ eine Diagonalmatrix, dann ist

$$\|f(\mathbf{A})\|_p \leq \text{cond}_p(\mathbf{T}) \max_{\lambda \in \Lambda(\mathbf{A})} |f(\lambda)|.$$

(2) Ist \mathbf{A} normal, d.h. gibt es eine unitäre Matrix \mathbf{T} , sodass $\mathbf{T}^H\mathbf{A}\mathbf{T}$ Diagonalform besitzt, dann ist

$$\|f(\mathbf{A})\|_2 = \max_{\lambda \in \Lambda(\mathbf{A})} |f(\lambda)|.$$

(3) In der Regel wird das Spektrum von \mathbf{A} nicht bekannt sein; manchmal kennt man aber eine Einschließung des Spektrums, also eine kompakte Menge $K \subset \mathbb{C}$ mit $\Lambda(\mathbf{A}) \subseteq K$. Offenbar kann man dann $\max_{\lambda \in \Lambda(\mathbf{A})} |f(\lambda)|$ nach oben durch $\max_{\zeta \in K} |f(\zeta)|$ abschätzen. Für Hermitesche Matrizen ergibt sich etwa

$$\|f(\mathbf{A})\|_2 \leq \max_{\lambda_{\min}(\mathbf{A}) \leq \lambda \leq \lambda_{\max}(\mathbf{A})} |f(\lambda)|.$$

Krylov-Verfahren für Matrixfunktionen

Schranken für $\|f(A)\|$

Im Folgenden spielen rationale Transformationen der Form

$$r = r_\alpha : \zeta \mapsto \frac{\zeta + \alpha}{1 + \bar{\alpha}\zeta}, \quad \alpha \in \mathbb{C}, \quad |\alpha| < 1, \quad (5.4)$$

eine entscheidende Rolle. Dies sind Bijektionen der abgeschlossenen Einheitskreisscheibe $\bar{\mathbb{D}}$ mit

$$r(0) = \alpha \quad \text{und} \quad r(-\alpha) = 0.$$

Die Umkehrabbildung von r_α ist

$$r_\alpha^{-1} : \omega \mapsto \frac{\omega - \alpha}{1 - \bar{\alpha}\omega}.$$

Abbildung 6 zeigt die Bilder der

Keise	$ \zeta = 2j/10,$	$j = 1, 2, \dots, 5$ und der
Strahlen	$\zeta = \rho \exp(2\pi i j/16)$	$0 \leq \rho \leq 1, j = 0, 1, \dots, 15$

unter der Transformation r_α für $\alpha = 1/4$ (links) und $\alpha = -1/2(1 + i)$ (rechts).

Krylov-Verfahren für Matrixfunktionen

Schranken für $\|f(A)\|$

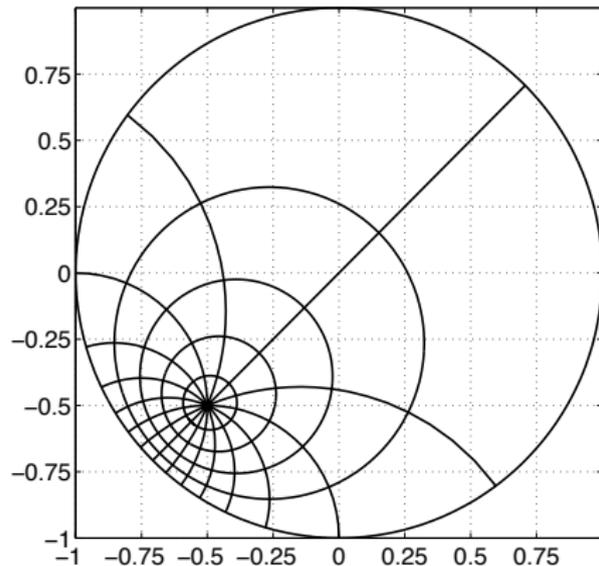
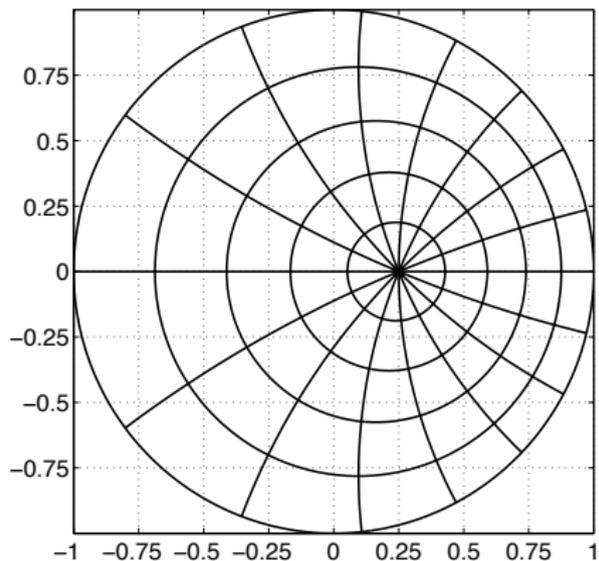


Abbildung 6: Die Transformation r_α .

Krylov-Verfahren für Matrixfunktionen

Schranken für $\|f(\mathbf{A})\|$

Die folgenden Abschätzungen entstammen dem Ideenkreis um den von John von Neumann stammenden Begriff der **Spektralmengen** linearer Operatoren in separablen Hilbert-Räumen.

Lemma 5.3

Sei r eine gebrochen lineare Transformation der Form (5.4). Für jede Matrix $\mathbf{A} \in \mathbb{C}^{N \times N}$ mit $\|\mathbf{A}\|_2 \leq 1$ gilt dann $\|r(\mathbf{A})\|_2 \leq 1$.

Satz 5.4 (von Neumann, 1951)

Seien $\mathbf{A} \in \mathbb{C}^{N \times N}$ und f analytisch in einer Umgebung von

$$\overline{\mathbb{D}}_{\mathbf{A}} := \{\lambda \in \mathbb{C} : |\lambda| \leq \|\mathbf{A}\|_2\}.$$

Dann ist $\|f(\mathbf{A})\|_2 \leq \max_{\lambda \in \overline{\mathbb{D}}_{\mathbf{A}}} |f(\lambda)|$.

Bemerkung: Ist \mathbf{A} normal, so folgt die Behauptung bereits aus der zweiten Bemerkung im Anschluss an Satz 5.2.

Korollar 5.5

Seien $\mathbf{A} \in \mathbb{C}^{N \times N}$, $\gamma \in \mathbb{C}$ und f analytisch in einer Umgebung von

$$\overline{\mathbb{D}}_{\mathbf{A}, \gamma} := \{\lambda \in \mathbb{C} : |\lambda - \gamma| \leq \|\mathbf{A} - \gamma \mathbf{I}\|_2\}.$$

Dann ist

$$\|f(\mathbf{A})\|_2 \leq \max_{\lambda \in \overline{\mathbb{D}}_{\mathbf{A}, \gamma}} |f(\lambda)|.$$

Dieses Korollar impliziert für ganze Funktionen f

$$\|f(\mathbf{A})\|_2 \leq \inf_{\gamma \in \mathbb{C}} \max_{\lambda \in \overline{\mathbb{D}}_{\mathbf{A}, \gamma}} |f(\lambda)|.$$

Ob diese Verschärfung praktisch relevant ist darf bezweifelt werden.

[Beispiel].

Krylov-Verfahren für Matrixfunktionen

Schranken für $\|f(\mathbf{A})\|$

Abschließend zitieren wir ein neueres Resultat von M. Crouzeix, dessen Beweis allerdings den Rahmen dieser Vorlesung sprengt.

Satz 5.6 (Crouzeix, 2017)

Es seien $\mathbf{A} \in \mathbb{C}^{N \times N}$ und f analytisch im Innern des Wertebereichs $W(\mathbf{A})$ (siehe (2.3)) von \mathbf{A} sowie stetig in $W(\mathbf{A})$. Dann gilt

$$\|f(\mathbf{A})\|_2 \leq C \max_{\lambda \in W(\mathbf{A})} |f(\lambda)|$$

mit einer universellen Konstanten C , $2 \leq C < 1 + \sqrt{2}$.

Crouzeix vermutet $C = 2$. Dies konnte er bereits 2004 für den Fall beweisen, dass $W(\mathbf{A})$ eine Kreisscheibe ist. 2007 wurde die obere Schranke auf 11.08 reduziert, die o.g. Schranke stammt von [Crouzeix & Palencia, 2017]. Darüberhinaus konnten [Greenbaum & Choi, 2013] sowie [Choi, 2014] (sehr spezielle) Klassen von Matrizen identifizieren, für die Satz 5.6 mit $C = 2$ richtig ist.

- 1 Einleitung
- 2 Krylov-Unterraumverfahren
- 3 Lineare Gleichungssysteme
- 4 Matrixfunktionen
- 5 Krylov-Verfahren für Matrixfunktionen
 - 5.1 Schranken für $\|f(A)\|$
 - 5.2 Fehlerschranken für Krylov-Verfahren
 - 5.3 Die Konvergenz des CG-Verfahrens
- 6 Das QR-Verfahren für Eigenwertaufgaben

Krylov-Verfahren für Matrixfunktionen

Fehlerschranken für Krylov-Verfahren

In diesem Abschnitt nehmen wir an, dass $\|\mathbf{b}\|_2 = 1$ gilt, was offensichtlich ohne Beschränkung der Allgemeinheit möglich ist. Die Arnoldi/Lanczos-Approximation an $f(\mathbf{A})\mathbf{b}$ ist dann $\mathbf{f}_m = \mathbf{V}_m f(\mathbf{H}_m) \mathbf{e}_1$.

Lemma 5.7

Für den Fehler $f(\mathbf{A})\mathbf{b} - \mathbf{f}_m$ der Arnoldi-Approximation gilt

$$\|f(\mathbf{A})\mathbf{b} - \mathbf{f}_m\|_2 \leq \min_{p \in \mathcal{P}_{m-1}} [\|(f - p)(\mathbf{A})\|_2 + \|(f - p)(\mathbf{H}_m)\|_2].$$

Die folgenden Aussagen sind unmittelbare Folgerungen aus Lemma 5.7 und den Abschätzungen aus dem letzten Abschnitt.

Satz 5.8

Seien $\mathbf{A} \in \mathbb{C}^{N \times N}$, $\mathbf{b} \in \mathbb{C}^N$ mit $\|\mathbf{b}\|_2 = 1$, f eine Funktion, für die $f(\mathbf{A})$ definiert ist, und \mathbf{f}_m die Arnoldi-Approximation an $f(\mathbf{A})\mathbf{b}$ aus $\mathcal{K}_m(\mathbf{A}, \mathbf{b})$.

(a) Ist f analytisch in einer Umgebung von $\overline{\mathbb{D}}_{\mathbf{A}} := \{\lambda \in \mathbb{C} : |\lambda| \leq \|\mathbf{A}\|_2\}$, so gilt

$$\|f(\mathbf{A})\mathbf{b} - \mathbf{f}_m\|_2 \leq 2 \min_{p \in \mathcal{P}_{m-1}} \max_{\lambda \in \overline{\mathbb{D}}_{\mathbf{A}}} |f(\lambda) - p(\lambda)|.$$

(b) Ist f analytisch in Innern des Wertebereich $W(\mathbf{A})$ von \mathbf{A} und stetig auf $W(\mathbf{A})$, so gilt

$$\|f(\mathbf{A})\mathbf{b} - \mathbf{f}_m\|_2 \leq 2C \min_{p \in \mathcal{P}_{m-1}} \max_{\lambda \in W(\mathbf{A})} |f(\lambda) - p(\lambda)|$$

mit der universellen Konstanten C aus Satz 5.6.

(c) [Stewart & Leyk, 1996] Ist $\mathbf{A} = \mathbf{A}^H$ und f stetig auf $[\lambda_{\min}(\mathbf{A}), \lambda_{\max}(\mathbf{A})]$, so gilt

$$\|f(\mathbf{A})\mathbf{b} - \mathbf{f}_m\|_2 \leq 2 \min_{p \in \mathcal{P}_{m-1}} \max_{\lambda_{\min}(\mathbf{A}) \leq \lambda \leq \lambda_{\max}(\mathbf{A})} |f(\lambda) - p(\lambda)|.$$

Krylov-Verfahren für Matrixfunktionen

Fehlerschranken für Krylov-Verfahren

Unter den Voraussetzungen von Satz 5.8 kann der Fehler der Krylov-Approximation also (modulo einer Konstanten) durch den Fehler der **polynomialen Bestapproximation** auf einer durch A definierten kompakten Menge abgeschätzt werden. Wir zitieren dazu folgenden Satz

Satz 5.9

Sei f stetig auf der kompakten Menge $K \subset \mathbb{C}$. Dann gibt es für jedes $m \in \mathbb{N}_0$ genau ein Polynom $p_m^* \in \mathcal{P}_m$ mit

$$\max_{\lambda \in K} |f(\lambda) - p_m^*(\lambda)| \leq \max_{\lambda \in K} |f(\lambda) - p(\lambda)| \quad \text{für alle } p \in \mathcal{P}_m.$$

Man nennt p_m^* die polynomiale Bestapproximation an f auf K und bezeichnet den zugehörigen Approximationsfehler mit

$$E_m(f, K) = \max_{\lambda \in K} |f(\lambda) - p_m^*(\lambda)| = \min_{p \in \mathcal{P}_m} \max_{\lambda \in K} |f(\lambda) - p(\lambda)|.$$

Die Bestapproximation und deren Fehler sind nur in Spezialfällen bekannt.

Ein solcher ist die (verschobene) Kehrwertfunktion

$$f(z) = \frac{1}{z - \gamma}, \quad z \in K = [-1, 1], \quad \gamma > 1.$$

Dann ist

$$\frac{1}{z - \gamma} - p_m^*(z) = \frac{(\gamma - \sqrt{\gamma^2 - 1})^m}{\gamma^2 - 1} \cos(m\varphi + \delta)$$

mit

$$z = \cos \varphi \quad \text{und} \quad \frac{\gamma z - 1}{z - \gamma} = \cos \delta$$

(vgl. [Meinardus, 1967, §4.3]). Das bedeutet

$$E_m((z - \gamma)^{-1}, [-1, 1]) = \frac{(\gamma - \sqrt{\gamma^2 - 1})^m}{\gamma^2 - 1}.$$

Krylov-Verfahren für Matrixfunktionen

Fehlerschranken für Krylov-Verfahren

Wir interessieren uns für die Bestapproximation $q_m^* \in \mathcal{P}_m$ an $g(\lambda) = 1/\lambda$ über dem Intervall $[\alpha, \beta] \subset (0, \infty)$, also $0 < \alpha < \beta$.

Die Transformation

$$\ell : [-1, 1] \rightarrow [\alpha, \beta], \quad z \mapsto \lambda = \ell(z) = \frac{\alpha - \beta}{2} z + \frac{\alpha + \beta}{2},$$

ist eine Bijektion. Ihre Umkehrabbildung ist

$$\ell^{-1} : [\alpha, \beta] \rightarrow [-1, 1], \quad \lambda \mapsto z = \frac{2}{\alpha - \beta} \lambda + \frac{\alpha + \beta}{\beta - \alpha}.$$

Daher ist

$$\max_{\lambda \in [\alpha, \beta]} |g(\lambda) - q(\lambda)| = \max_{z \in [-1, 1]} |g(\ell(z)) - q(\ell(z))|$$

für alle $q \in \mathcal{P}_m$.

Da mit q auch $q \circ \ell$ ein Polynom vom Grad m ist und weil

$$g(\ell(z)) = \frac{1}{\ell(z)} = \frac{2}{\alpha - \beta} \frac{1}{z - \gamma} = \frac{2}{\alpha - \beta} f(z)$$

mit $\gamma := \frac{\alpha + \beta}{\beta - \alpha} = 1 + 2\frac{\alpha}{\beta - \alpha} > 1$ gilt, folgt

$$\max_{\lambda \in [\alpha, \beta]} |g(\lambda) - q(\lambda)| = \frac{2}{\beta - \alpha} \max_{z \in [-1, 1]} \left| f(z) - \frac{\alpha - \beta}{2} q(\ell(z)) \right| \quad \forall q \in \mathcal{P}_m.$$

Deshalb sind (vgl. Abb. 7)

$$\frac{\alpha - \beta}{2} q_m^*(\ell(z)) = p_m^*(z) \quad \text{bzw.} \quad q_m^*(\lambda) = \frac{2}{\alpha - \beta} p_m^*(\ell^{-1}(\lambda))$$

und

$$E_m(g, [\alpha, \beta]) = \frac{2}{\beta - \alpha} E_m(f, [-1, 1]).$$

Krylov-Verfahren für Matrixfunktionen

Fehlerschranken für Krylov-Verfahren

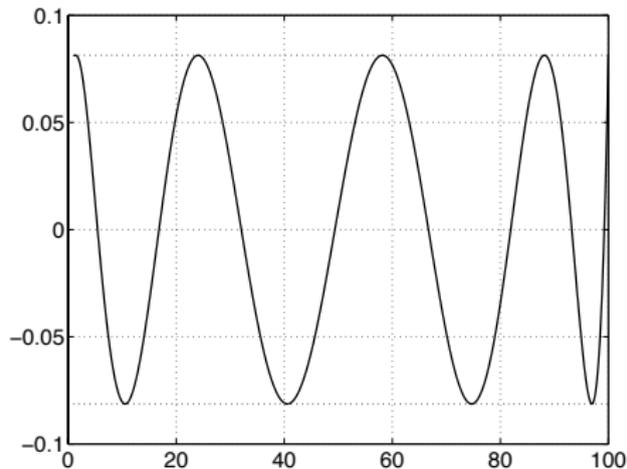
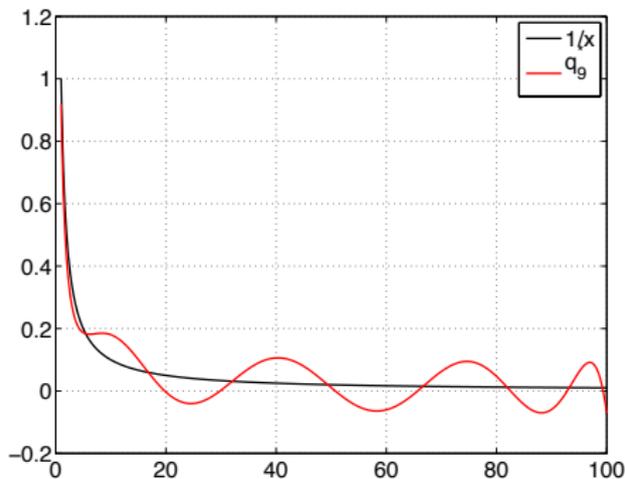


Abbildung 7: Die Funktion $g(\lambda) = 1/\lambda$ und ihre polynomiale Bestapproximation $q_9^* \in \mathcal{P}_9$ (links) sowie der zugehörige Fehler (rechts).

Satz 5.10

Sei $\mathbf{A} \in \mathbb{C}^{N \times N}$ Hermitesch und positiv definit. Dann genügen die Approximationen \mathbf{x}_m ($\mathbf{x}_0 = \mathbf{0}$) des CG-Verfahrens zur Lösung von $\mathbf{A}\mathbf{x} = \mathbf{b}$ der Fehlerabschätzung

$$\|\mathbf{A}^{-1}\mathbf{b} - \mathbf{x}_m\|_2 \leq \frac{\kappa - 1}{\lambda_{\max}(\mathbf{A})} \left[\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1} \right]^{m-1} \|\mathbf{b}\|_2$$

mit der Konditionszahl $\kappa := \text{cond}_2(\mathbf{A}) = \lambda_{\max}(\mathbf{A})/\lambda_{\min}(\mathbf{A})$ von \mathbf{A} .

Krylov-Verfahren für Matrixfunktionen

Fehlerschranken für Krylov-Verfahren

Ein weiteres Resultat aus der Approximationstheorie (vgl. [Meinardus, 1967,(6.29)]) sind die Schranken

$$\frac{1}{2^m (m+1)!} \left(1 + \frac{0.18}{m+2}\right) \leq E_m(e^z, [-1, 1]) \leq \frac{1}{2^m (m+1)!} \left(1 + \frac{1.6}{m+2}\right)$$

und allgemeiner (vgl. [Meinardus, 1967,(6.75)])

$$E_m(e^{tz}, [-1, 1]) = 2 I_{m+1}(|t|) (1 + o(1)) = \frac{|t|^{m+1}}{2^m (m+1)!} (1 + o(1)), \quad m \rightarrow \infty,$$

wobei $t \in \mathbb{R}$ und

$$I_m(t) = \sum_{j=0}^{\infty} \frac{(t/2)^{2j+m}}{j! (n+m)!}$$

die Bessel-Funktion m -ter Ordnung mit rein imaginärem Argument bezeichnet.

Hieraus erhalten wir sofort folgende asymptotische Fehlerabschätzung:

Satz 5.11

Seien $\mathbf{A} \in \mathbb{C}^{N \times N}$ Hermitesch mit $\Lambda(\mathbf{A}) \subset [\alpha, \beta] \subset \mathbb{R}$, $\mathbf{b} \in \mathbb{C}^N$ mit $\|\mathbf{b}\|_2 = 1$ und \mathbf{f}_m die Arnoldi-Approximation an $\exp(\mathbf{A})\mathbf{b}$ aus $\mathcal{K}_m(\mathbf{A}, \mathbf{b})$. Dann gilt für den Approximationsfehler

$$\|\exp(\mathbf{A})\mathbf{b} - \mathbf{f}_m\|_2 \leq \exp\left(\frac{\alpha + \beta}{2}\right) \frac{4}{m!} \left[\frac{\beta - \alpha}{4}\right]^m (1 + o(1)) \quad (m \rightarrow \infty).$$

- 1 Einleitung
- 2 Krylov-Unterraumverfahren
- 3 Lineare Gleichungssysteme
- 4 Matrixfunktionen
- 5 Krylov-Verfahren für Matrixfunktionen
 - 5.1 Schranken für $\|f(A)\|$
 - 5.2 Fehlerschranken für Krylov-Verfahren
 - 5.3 Die Konvergenz des CG-Verfahrens
- 6 Das QR-Verfahren für Eigenwertaufgaben

Krylov-Verfahren für Matrixfunktionen

Die Konvergenz des CG-Verfahrens

Die klassische Fehlerabschätzung für das CG-Verfahren basiert auf der Minimiereigenschaft (vgl. Satz 3.12)

$$\|\mathbf{x}_* - \mathbf{x}_m^{\text{CG}}\|_A = \min_{\mathbf{x} \in \mathbf{x}_0 + \mathcal{K}_m(\mathbf{A}, \mathbf{r}_0)} \|\mathbf{x}_* - \mathbf{x}\|_A. \quad (5.5)$$

Wegen $\mathbf{x}_* = \mathbf{x}_0 + \mathbf{A}^{-1} \mathbf{r}_0$, $\mathbf{r}_0 = \mathbf{A}(\mathbf{x}_* - \mathbf{x}_0)$ und

$$\mathbf{x}_0 + \mathcal{K}_m(\mathbf{A}, \mathbf{r}_0) = \mathbf{x}_0 + \{q_{m-1}(\mathbf{A})\mathbf{r}_0 : q_{m-1} \in \mathcal{P}_{m-1}\}$$

lässt sich der Fehlervektor $\mathbf{x}_* - \mathbf{x}$ für jedes $\mathbf{x} \in \mathbf{x}_0 + \mathcal{K}_m(\mathbf{A}, \mathbf{b})$ in der Form

$$\mathbf{x}_* - \mathbf{x} = [\mathbf{I} - \mathbf{A}q_{m-1}(\mathbf{A})](\mathbf{x}_* - \mathbf{x}_0) =: p_m(\mathbf{A})(\mathbf{x}_* - \mathbf{x}_0)$$

schreiben. Dabei ist

$$p_m(\lambda) = 1 - \lambda q_{m-1}(\lambda)$$

ein Polynom vom Grad m , welches durch $p_m(0) = 1$ normiert ist.

Aus (5.5) folgt

$$\begin{aligned}\|\mathbf{x}_* - \mathbf{x}_m^{\text{CG}}\|_A &= \min_{\substack{p \in \mathcal{P}_m, \\ p(0)=1}} \|p(\mathbf{A})(\mathbf{x}_* - \mathbf{x}_0)\|_A \\ &\leq \min_{\substack{p \in \mathcal{P}_m, \\ p(0)=1}} \|p(\mathbf{A})\|_A \|\mathbf{x}_* - \mathbf{x}_0\|_A\end{aligned}$$

bzw.

$$\begin{aligned}\frac{\|\mathbf{x}_* - \mathbf{x}_m^{\text{CG}}\|_A}{\|\mathbf{x}_* - \mathbf{x}_0\|_A} &\leq \min_{\substack{p \in \mathcal{P}_m, \\ p(0)=1}} \|p(\mathbf{A})\|_A = \min_{\substack{p \in \mathcal{P}_m, \\ p(0)=1}} \|p(\mathbf{A})\|_2 \\ &= \min_{\substack{p \in \mathcal{P}_m, \\ p(0)=1}} \max_{\lambda \in \Lambda(\mathbf{A})} |p(\lambda)|,\end{aligned}$$

denn $\|p(\mathbf{A})\|_A = \|\mathbf{A}^{1/2}p(\mathbf{A})\mathbf{A}^{-1/2}\|_2 = \|p(\mathbf{A})\|_2$ für alle Polynome p .

Damit ist folgendes Lemma bewiesen:

Lemma 5.12

Sei $K \subseteq (0, \infty)$ eine beliebige kompakte Menge, die sämtliche Eigenwerte von \mathbf{A} enthält, und sei $p_m \in \mathcal{P}_m$ ein beliebiges Polynom vom Grad m , das $p_m(0) = 1$ erfüllt. Dann gilt

$$\frac{\|\mathbf{x}_* - \mathbf{x}_m^{\text{CG}}\|_{\mathbf{A}}}{\|\mathbf{x}_* - \mathbf{x}_0\|_{\mathbf{A}}} \leq \max_{\zeta \in K} |p_m(\zeta)|.$$

- Auf dieser – fast trivialen – Beobachtung basieren alle Fehlerabschätzungen für das CG-Verfahren: Man geht von einer Einschließung K der Eigenwerte von \mathbf{A} aus und versucht, Polynome p_m zu konstruieren, die der Nebenbedingung $p_m(0) = 1$ genügen und auf K möglichst wenig von Null abweichen.
- Man wird so ganz natürlich mit Problemen der polynomialen Bestapproximation konfrontiert.

Bevor wir den Hauptsatz dieses Abschnitts formulieren, schließen wir noch zwei Bemerkungen zu Lemma 5.12 an:

Bemerkung 5.13

Die Abschätzung

$$\frac{\|\mathbf{x}_* - \mathbf{x}_m^{\text{CG}}\|_A}{\|\mathbf{x}_* - \mathbf{x}_0\|_A} \leq \min_{\substack{p \in \mathcal{P}_{m-1}, \\ p(0)=1}} \max_{\lambda \in \Lambda(\mathbf{A})} |p(\lambda)|$$

ist in gewissem Sinn scharf: Aus der Definition der induzierten Matrixnorm folgt nämlich, dass man zu jedem m einen Startvektor \mathbf{x}_0 finden kann, sodass hier Gleichheit herrscht. Dies zeigt, dass die Konvergenzgeschwindigkeit des CG-Verfahrens ausschließlich durch die Eigenwerte von \mathbf{A} bestimmt wird.

Bemerkung 5.14

Besitzt die Matrix \mathbf{A} nur d verschiedene Eigenwerte $\lambda_1, \dots, \lambda_d$, dann berechnet das CG-Verfahren bekanntlich im d -ten Schritt die exakte Lösung.

Diese Aussage folgt auch unmittelbar aus Lemma 5.12, wenn man für p_d das Interpolationspolynom vom Grad d wählt, das durch die $(d + 1)$ Bedingungen

$$p_d(0) = 1, \quad p_d(\lambda_j) = 0, \quad (j = 1, \dots, d)$$

bestimmt ist.

Satz 5.15

Für die Iterierten des CG-Verfahrens gilt die Fehlerabschätzung

$$\frac{\|\mathbf{x}_* - \mathbf{x}_m^{\text{CG}}\|_A}{\|\mathbf{x}_* - \mathbf{x}_0\|_A} \leq \frac{1}{T_m\left(\frac{\kappa+1}{\kappa-1}\right)} \leq 2 \left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^m. \quad (5.6)$$

Dabei ist $\kappa = \text{cond}_2(\mathbf{A}) = \lambda_{\max}/\lambda_{\min}$ die Konditionszahl von \mathbf{A} bez. der Euklid-Norm.

Die obere Schranke aus Satz 5.15 ist eine monoton wachsende Funktion von κ , d.h. wächst die Konditionszahl, so verschlechtert sich zumindest die Abschätzung für das Konvergenzverhalten des CG-Verfahrens.

Es ist klar, dass die Abschätzung in vielen Fällen zu pessimistisch ist:

- Für eine Matrix \mathbf{A} mit nur zwei Eigenwerten, λ_{\min} und λ_{\max} , konvergiert das CG-Verfahren in zwei Schritten, während die obere Schranke hier beliebig schlecht werden kann.
- Ähnlich liegt der Fall, wenn die Eigenwerte von \mathbf{A} sich an wenigen Punkten in $[\lambda_{\min}, \lambda_{\max}]$ konzentrieren, also sogenannte „Cluster“ bilden: Auch dann wird das CG-Verfahren wesentlich schneller konvergieren als es Satz 5.15 vorhersagt, weil hier $K = [\lambda_{\min}, \lambda_{\max}]$ keine gute Einschließung für $\Lambda(\mathbf{A})$ ist.

Beispiel 5.16

Wir betrachten vier Diagonalmatrizen $\mathbf{A} \in \mathbb{R}^{100 \times 100}$ mit gleicher Konditionszahl $\kappa = 100$. Satz 5.15 liefert daher für jede dieser Matrizen dieselbe Abschätzung.

Die Diagonaleinträge von \mathbf{A}_1 sind die Extremalstellen des auf $[1, 100]$ transformierten Tschebyschew-Polynoms T_{99} ,

$$\mathbf{A}_2 = \text{diag}([1 : 100]),$$

$$\mathbf{A}_3 = \text{diag}(51 + [0 : 99](\rho^{(99 - [0:99])})) \quad \text{mit } \rho = 0.8.$$

Neben 1 und 100 besitzt \mathbf{A}_4 als Diagonaleinträge 98 $N(50.5, 10)$ -verteilte Zufallszahlen.

Die Konvergenzkurven des CG-Verfahrens für die rechte Seite $\mathbf{b} = [1, 1, \dots, 1]^T$ und den Startvektor $\mathbf{x}_0 = \mathbf{0}$ sind in Abb. 8 dargestellt.

Krylov-Verfahren für Matrixfunktionen

Die Konvergenz des CG-Verfahrens

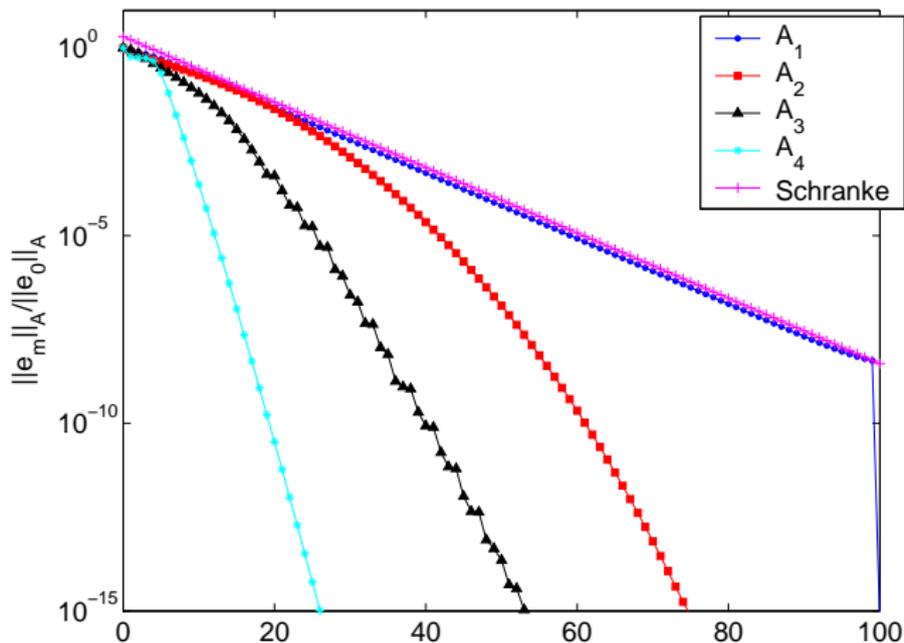


Abbildung 8: Konvergenzkurven des CG-Verfahrens.

Krylov-Verfahren für Matrixfunktionen

Die Konvergenz des CG-Verfahrens

„As we have seen the conjugate gradients algorithm is just an efficient implementation of the Lanczos algorithm. The eigenvalues of the implicitly generated tridiagonal matrices T_i are the Ritz values of A with respect to the current Krylov subspace. It is known from Lanczos theory that these Ritz values converge to the eigenvalues of A and that in general the extremal eigenvalues of A are first well approximated [...]. Furthermore, the speed of convergence depends on how well these eigenvalues are separated from the others (gap ratio) [...]. This helps us to understand the so-called superlinear convergence behavior of the conjugate gradient method (as well as other Krylov subspace methods). It can be shown that as soon as one of the extremal eigenvalues is modestly well approximated by a Ritz value, the procedure converges from then on as a process in which this eigenvalue is absent, i.e., a process with a reduced condition number.“

Henk A. Van der Vorst
Lecture Notes on Iterative Methods
University of Utrecht 1994.

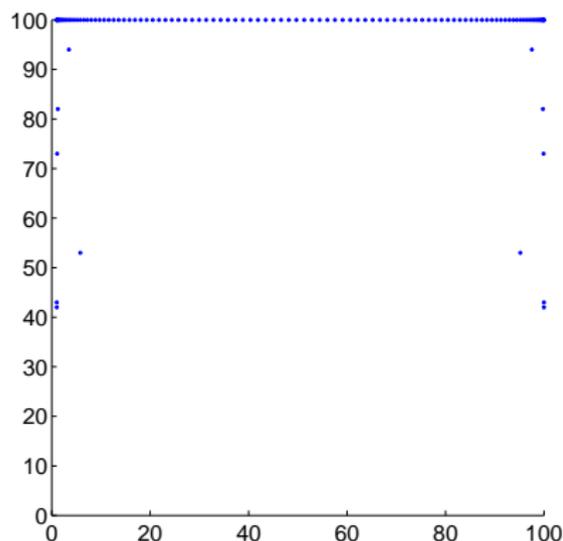
Krylov-Verfahren für Matrixfunktionen

Die Konvergenz des CG-Verfahrens

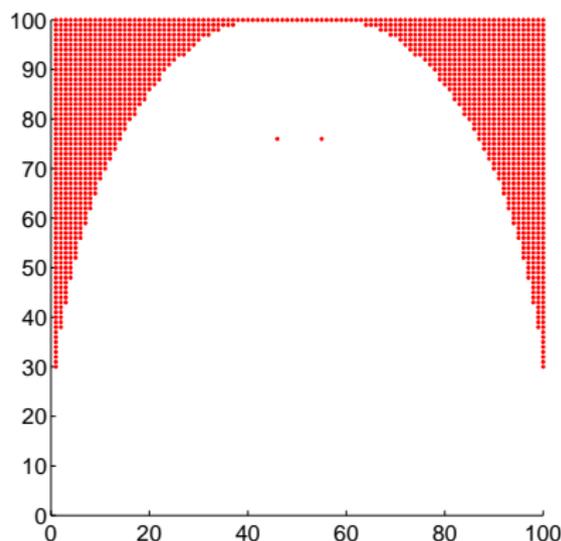
Hinreichend gut approximierte Eigenwerte von A im Lauf der CG/Lanczos-Iteration:

Abszisse: Index j der Eigenwerte $\{\lambda_j\}_{j=1}^{100}$ von A

Ordinate: Iterationsindex des CG/Lanczos-Verfahrens



A_1



A_2

Krylov-Verfahren für Matrixfunktionen

Die Konvergenz des CG-Verfahrens

Hinreichend gut approximierte Eigenwerte von A im Lauf der CG/Lanczos-Iteration:

Abszisse: Index j der Eigenwerte $\{\lambda_j\}_{j=1}^{100}$ von A

Ordinate: Iterationsindex des CG/Lanczos-Verfahrens

