Numerische Mathematik für Elektrotechniker

Zusammenfassung

Benjamin Berkels

Karl-Heinz Brakhage, Thomas Jankuhn, Christian Löbbert

Institut für Geometrie und Praktische Mathematik
RWTH Aachen

Wintersemester 2019/2020

Heute in der Vorlesung

Zusammenfassung des Stoffes Dahmen & Reusken Kap. 1-10

1. Einleitung
2. Fehleranalyse: Kondition, Rundungsfehler, Stabilität
3. Lineare Gleichungssysteme, direkte Lösungsverfahren
4. Lineare Ausgleichsrechnung
5. Nichtlineare Gleichungssysteme, iterative Lösungsverfahren
6. Nichtlineare Ausgleichsrechnung
7. Berechnung von Eigenwerten
8. Interpolation
9. Splinefunktionen
10. Numerische Integration

Klausuraufgaben: Verständnisfragen

Zusammenfassung

Zusammenfassung

Kondition

Kondition: Definition

- ist unabhängig von einem speziellen Lösungweg (Algorithmus)
- gibt an, welche Genauigkeit man bestenfalls (bei exakter Rechnung) bei gestörten Eingangsdaten erwarten kann.

Frage:

- Was ist die (relative) Kondition eines Problems?

$\Rightarrow$ Die relative Kondition eines Problems bezeichnet das Verhältnis des relativen Ausgabefehlers zum relativen Eingabefehler, d.h. die Sensitivität des Problems unter Störungen der Eingabedaten.

Kondition: Berechnung

Wie wird die Kondition analysiert?

- Wir haben folgende Fälle betrachtet
  - $f : \mathbb{R} \to \mathbb{R}$
  - $f : \mathbb{R}^n \to \mathbb{R}$
  - $f : \mathbb{R}^n \to \mathbb{R}^n$, f linear

Wie sind die elementaren Rechenoperationen konditioniert?

- Multiplikation und Division sind für alle Eingangsdaten gut konditioniert.
- Addition ist
  - gut konditioniert, wenn beide Zahlen das gleiche Vorzeichen haben;
  - sehr schlecht konditioniert, wenn $x_1 \approx -x_2$.

Kondition: Berechnung

- $f : \mathbb{R}^n \to \mathbb{R}$ (Eingabe: Vektor, Ausgabe: Skalar)

$$\left|\frac{f(\tilde{x}) - f(x)}{f(x)}\right| \dot{\leq} \kappa_{\text{rel}}(x) \sum_{j=1}^{n} \left|\frac{\tilde{x}_j - x_j}{x_j}\right|$$

mit

$$\kappa_{\text{rel}}(x) = \kappa_{\text{rel}}^{\infty}(x) := \max_j \left|\frac{\partial f(x)}{\partial x_j} \cdot \frac{x_j}{f(x)}\right|$$

- $f : \mathbb{R}^n \to \mathbb{R}^n$ (Eingabe: Vektor, Ausgabe: Vektor)
  Für $f(x) = Ax$ linear, invertierbar gilt

$$\frac{\|f(\tilde{x}) - f(x)\|}{\|f(x)\|} \leq \underbrace{\|A\| \cdot \|A^{-1}\|}_{\kappa(A)} \frac{\|\tilde{x} - x\|}{\|x\|}$$

Maschinenzahlen

Wie werden Zahlen im Computer dargestellt?

- Maschinenzahlen $\mathbb{M}(\boldsymbol{b}, \boldsymbol{m}, \boldsymbol{r}, \boldsymbol{R})$
  $\Rightarrow \ \boldsymbol{x}_{\mathrm{MIN}}, \ \boldsymbol{x}_{\mathrm{MAX}}, \ \textbf{eps}, \ |\varepsilon| \leq \textbf{eps}$

Welche Probleme können dabei/deswegen auftreten?

- Assoziativ- und Distributivgesetz nicht mehr gültig
- Gefahr der Auslöschung bei $\boldsymbol{\nabla} \in \{+, -\}$

Maschinenzahlen: Darstellung

Seien $b \in \mathbb{N}$, $b > 1$, fest gewählt. Jedes $x \in \mathbb{R}$, $x \neq 0$, lässt sich in der Form

$$x = \pm \left(\sum_{j=1}^{\infty} d_j \, b^{-j} \right) \cdot b^e$$

darstellen, mit $d_j \in \{0, 1, \ldots, b-1\}$, $d_1 \neq 0$,
und e eine ganze Zahl.

Nur endliche Anzahl darstellbar $\Rightarrow$ Maschinenzahlen $\mathbb{M}(b, m, r, R)$

- Betragsmäßig kleinste ($\neq 0$) und größte Zahl:

$$x_{\text{MIN}} = b^{r-1} \quad \text{und} \quad x_{\text{MAX}} = (1 - b^{-m}) \times b^R$$

- Bildbereich $\mathbb{D} := [-x_{\text{MAX}}, -x_{\text{MIN}}] \cup [x_{\text{MIN}}, x_{\text{MAX}}]$
- Unterlauf, wenn $0 \neq |x| < |x_{\text{MIN}}|$;
- Überlauf, wenn $|x| > |x_{\text{MAX}}|$.

Maschinengenauigkeit

- Für den relativen Rundungsfehler erhält man für $x \in \mathbb{D}$

$$\left|\frac{\mathrm{fl}(x) - x}{x}\right| \leq \frac{\frac{b^{-m}}{2} \cdot b^e}{b^{-1} \cdot b^e} = \frac{b^{1-m}}{2}.$$

- Die (relative) Maschinengenauigkeit

$$\mathsf{eps} := \frac{b^{1-m}}{2}$$

  charakterisiert das Auflösungsvermögen des Rechners, d.h.

$$\mathsf{eps} = \inf\{\delta > 0 \mid \mathrm{fl}(1+\delta) > 1\}$$

- Der Rundungsfehler ε erfüllt $|\varepsilon| \leq \mathsf{eps}$ und es gilt

$$\mathrm{fl}(x) = x\,(1+\varepsilon).$$

Rückwärtsstabilität

Definition (Rückwärtsstabilität)

Das Verfahren heißt rückwärts stabil, wenn es für alle $x \in X$ ein $\tilde{x} \in X$ gibt, so dass

$$\tilde{f}(x) = f(\tilde{x}) \quad \text{und} \quad \frac{\|x - \tilde{x}\|}{\|x\|} = \mathcal{O}(\text{eps}).$$

Satz

Wird ein rückwärts stabiler Algorithmus zur Lösung des Problems f mit Kondition $\kappa(x)$ angewendet, so gilt

$$\frac{\|\tilde{f}(x) - f(x)\|}{\|f(x)\|} = \mathcal{O}(\kappa(x)\,\text{eps}).$$

Stabilität vs. Kondition

- Bei einem stabilen Lösungsverfahren bleiben die im Laufe der Rechnung erzeugten Rundungsfehler in der Größenordnung der durch die Kondition des Problems bedingten unvermeidbaren Fehler.
- Kenntnisse über die Kondition eines Problems sind oft für die Interpretation oder Bewertung der Ergebnisse von entscheidender Bedeutung.
  - "Schlechtes Ergebnis" bedeutet nicht unbedingt gleich "instabiler Algorithmus", sondern deutet evtl. auf eine schlechte Kondition des Problems hin.
- In einem Algorithmus sollen (wegen Stabilität) Auslöschungseffekte vermieden werden.

Operationen $\cdot$ und $/$ sind für alle Eingangsdaten gut konditioniert. Operationen $+$ und $-$ können schlecht konditioniert sein. Dadurch können Rundungsfehler enorm verstärkt werden: Auslöschung.

Lineare Gleichungssysteme: Kondition

Satz 3.9

Sie $x + \Delta x$ die Lösung von $(A + \Delta A)(x + \Delta x) = b + \Delta b$. Dann gilt falls $\kappa(A)\frac{\|\Delta A\|}{\|A\|} < 1$:

$$\frac{\|\Delta x\|}{\|x\|} \le \frac{\kappa(A)}{1 - \kappa(A)\frac{\|\Delta A\|}{\|A\|}} \left(\frac{\|\Delta A\|}{\|A\|} + \frac{\|\Delta b\|}{\|b\|} \right).$$

wobei

$$\kappa_{\|\cdot\|}(A) \equiv \|A^{-1}\| \, \|A\|$$

die (relative) Konditionszahl der Matrix A (bzgl. $\|\cdot\|$) ist.

Der Fehler ist proportional zum Residuum $\tilde{r} = b - A\,\tilde{x}$

$$\kappa(A)^{-1}\frac{\|\tilde{r}\|}{\|b\|} \le \frac{\|x - \tilde{x}\|}{\|x\|} \le \kappa(A)\frac{\|\tilde{r}\|}{\|b\|}$$

Matrixzerlegungen

- LR-Zerlegung/Gauß-Elimination mit Spaltenpivotisierung
  $\boldsymbol{P A = L R}$
  - (praktisch) stabile und effiziente Methode
  - Aufwand $\sim \frac{1}{3} n^3$
- Cholesky-Zerlegung: $\boldsymbol{A = L D L^T}$
  - Für symmetrisch positiv definite Matrizen ("symmetrische" LR-Zerlegung)
  - Berechnung durch elementweise Auswertung von $\boldsymbol{A = L D L^T}$: Cholesky-Verfahren
- QR-Zerlegung: $\boldsymbol{A = Q R}$
  - Existiert auch für rechteckige Matrizen
  - Berechnung über Givens-Rotation, Housholder-Transformation (oder Gram-Schmidt)

Gauß-Elimination/ LR-Zerlegung

- Dreiecksmatrizen ergeben leicht lösbare Systeme:
  Aufwand $\approx \frac{1}{2} n^2$ Operationen.
- $\kappa(A)$ spielt eine zentrale Rolle bei der Kondition des Problems $(A, b) \to x = A^{-1}b$, siehe Satz 3.9.
- Residuum als Maß für die Genauigkeit:
  Nur aussagekräftig für $\kappa(A) \approx 1$.
- Zeilenskalierung vs. Pivotisierung
  - Skalierung/Äquilibrierung verbessert die Kondition des Gleichungssystems. (sogenannte Vorkonditionierung)
  - Pivotisierung verbessert die Stabilität der Gauß-Elimination/ LR-Zerlegung.

Cholesky-Zerlegung

Definition 3.31. (Symmetrisch postiv definite Matrix)

$\boldsymbol{A} \in \mathbb{R}^{\boldsymbol{n}\times\boldsymbol{n}}$ heißt symmetrisch positiv definit (s.p.d.), falls

$$\boldsymbol{A}^{\boldsymbol{T}} = \boldsymbol{A} \qquad \text{(Symmetrie)}$$

und für alle $\boldsymbol{x} \in \mathbb{R}^{\boldsymbol{n}}$, $\boldsymbol{x} \neq \mathbf{0}$ gilt

$$\boldsymbol{x}^{\boldsymbol{T}}\boldsymbol{A}\,\boldsymbol{x} > \mathbf{0} \qquad \text{(positiv definit).}$$

- $\boldsymbol{A}$ ist invertierbar, und $\boldsymbol{A}^{-\mathbf{1}}$ ist s.p.d.
- $\boldsymbol{A}$ hat nur strikt positive (insbesondere reelle) Eigenwerte.
- $\boldsymbol{A}$ hat nur strikt positive Diagonaleinträge und der betragsgrößte Eintrag von $\boldsymbol{A}$ liegt auf der Diagonalen.
- Bei Gauß-Elimination ohne Pivotisierung sind alle Pivotelemente strikt positiv.

QR-Zerlegung

Eine Matrix $Q \in \mathbb{R}^{m\times m}$ heißt orthogonal, falls

$$Q^T Q = I.$$

Eigenschaften:

- $Q^{-1} = Q^T$.
- $\|Q\,x\|_2 = \|x\|_2$, $\kappa_2(Q) = 1$.
- $\|Q\,A\|_2 = \|A\|_2$, $\kappa_2(Q\,A) = \kappa_2(A)$.

Satz QR-Zerlegung

Zu $A \in \mathbb{R}^{m\times n}$ existiert eine Orthogonale Matrix Q, so dass

$$A = Q\,R,$$

mit R obere Dreiecksmatrix.

Givens-Rotationen

Grundaufgabe

Gegeben sei $(a, b)^T \in \mathbb{R}^2 \setminus \{0\}$. Finde $c, s, r \in \mathbb{R}$ mit

$$\begin{pmatrix} c & s \\ -s & c \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} r \\ 0 \end{pmatrix}$$

und

$$c^2 + s^2 = 1.$$

Die Lösung ist: $r = \pm\sqrt{a^2 + b^2}, \quad c := \frac{a}{r}, \quad s := \frac{b}{r}$

Beachte

Drehung verändert nicht die Euklidische Länge eines Vektors, d.h.

$$\|(r, 0)^T\|_2 = |r| = \sqrt{a^2 + b^2} = \|(a, b)^T\|_2.$$

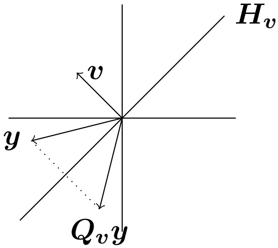
- Die obige (Rotations-)Matrix ist orthogonal.

Householder-Transformation

Definition

Für $v = (v_1, \ldots, v_n)^T \in \mathbb{R}^n \setminus \{0\}$, ist die Householder-Transformation definiert als

$$Q_v = I - 2\,\frac{v\,v^T}{v^T\,v}.$$



$$H_v = \{x \in \mathbb{R}^n \,|\, x^T v = 0\}.$$

Lösung von $Q_v y = \pm\|y\|_2 e^1$

$$v = y + \alpha\, e^1$$
$$\alpha = \operatorname{sign}(y_1)\,\|y\|_2$$
$$Q_v y = -\alpha\, e^1$$

Linearer Ausgleich

- Aufgabe: $b \in \mathbb{R}^m$, $A \in \mathbb{R}^{m \times n}$, $m \geq n$,

$$\min_{x \in \mathbb{R}^n} \|A\,x - b\|_2 \qquad \Leftrightarrow \qquad A^T A\,x = A^T b$$

- Eindeutige Lösung $\Leftrightarrow$ $\mathrm{Rang}(A) = n$

- Kondition (nur Störung in b):

$$\frac{\|\tilde{x} - x^*\|_2}{\|x^*\|_2} \leq \frac{\kappa_2(A)}{\cos\Theta} \cdot \frac{\|\tilde{b} - b\|_2}{\|b\|_2}$$

wobei $\cos\Theta = \dfrac{\langle A\,x^*, b\rangle}{\|A\,x^*\|\,\|b\|} \overset{b - Ax^* \perp Ax^*}{=} \dfrac{\|A\,x^*\|}{\|b\|}$

- Lösungsverfahren:
  - über Normalgleichungen $A^T A\,x = A^T b$ (Cholesky-Verfahren)
  - über QR-Zerlegung (Householder, Givens)

Linearer Ausgleich: Stabilität

	Normalgleichungen	QR-Zerlegung
Rechenaufwand $(m \gg n)$	$\approx \frac{1}{2}\, m\, n^2$	$\approx m\, n^2$ (Householder)
Stabilität	instabil, wenn $\kappa_2(A) \gg 1$ stabil, wenn $\kappa_2(A)$ moderat	stabil

Singulärwertzerlegung

Satz 4.27

Zu jeder Matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ existieren orthogonale Matrizen $\boldsymbol{U} \in \mathbb{R}^{m\times m}$, $\boldsymbol{V} \in \mathbb{R}^{n\times n}$ und eine Diagonalmatrix

$$\boldsymbol{\Sigma} := \mathbf{diag}_{mn}(\boldsymbol{\sigma}_1, \ldots, \boldsymbol{\sigma}_p) \in \mathbb{R}^{m\times n}, \quad p = \min(m, n),$$

mit

$$\boldsymbol{\sigma}_1 \geq \boldsymbol{\sigma}_2 \geq \cdots \geq \boldsymbol{\sigma}_p \geq 0,$$

so dass

$$\boldsymbol{A} = \boldsymbol{U}\,\boldsymbol{\Sigma}\,\boldsymbol{V}^T.$$

Hierbei heißen:

- Singulärwerte von $\boldsymbol{A}$: $\boldsymbol{\sigma}_i,\ i = 1, \ldots, p$
- Linkssingulärvektoren : Spalten von $\boldsymbol{U} = [\boldsymbol{u}_1\, \boldsymbol{u}_2 \cdots \boldsymbol{u}_m]$
- Rechtssingulärvektoren: Spalten von $\boldsymbol{V} = [\boldsymbol{v}_1\, \boldsymbol{v}_2 \cdots \boldsymbol{v}_n]$

Singulärwertzerlegung: Eigenschaften

Sei $A = U\,\Sigma\,V^T$ eine Singulärwertzerlegung von $A \in \mathbb{R}^{m\times n}$ mit

$$\sigma_1 \geq \cdots \geq \sigma_r > \sigma_{r+1} = \cdots = \sigma_p = 0.$$

- $\|A\|_2 = \sigma_1 = \sigma_{\max}$
- $\|A^{-1}\|_2 = \frac{1}{\sigma_n} = \frac{1}{\sigma_{\min}}$, falls $A \in \mathbb{R}^{n\times n}$ regulär
- $\kappa_2(A) = \|A\|_2\,\|A^{-1}\|_2 = \frac{\sigma_1}{\sigma_n}$, falls $A \in \mathbb{R}^{n\times n}$ regulär
- Die Pseudoinverse $A^+ \in \mathbb{R}^{n\times m}$ ist definiert durch

  $$A^+ = V\,\Sigma^+\,U^T \text{ mit } \Sigma^+ = \mathrm{diag}_{n\times m}(\sigma_1^{-1}, \ldots, \sigma_r^{-1}, 0, \ldots).$$

- $\kappa_2(A) = \|A\|_2\,\|A^+\|_2 = \frac{\sigma_1}{\sigma_r} = \frac{\sigma_{\max}}{\sigma_{\min}}$,
  falls $\mathrm{Rang}(A) = r$ mit A^+ die Pseudoinverse.
- Linearer Ausgleich: $x^* = A^+ b = V\,\Sigma^+\,U^T b$ ist Minimierer von $\min_{x\in\mathbb{R}^n} \|A\,x - b\|_2$ mit $\|x^*\|_2$ minimal.

Nichtlineare Gleichungssysteme

- Nullstellenproblem $f(x) = 0 \Leftrightarrow$ Fixpunktproblem $\Phi(x) = x$.
  Es gibt viele Möglichkeiten für Φ,
  zum Beispiel: für invertierbare Matrix M_x:
  $\Phi(x) = x - M_x\, f(x)$.
- Fixpunktiteration:
  $$x_{k+1} = \Phi(x_k)$$
- Banachscher Fixpunktsatz:

  $\Phi : E \to E$, E abgeschloßen, konvex (Selbstabbildung)

  Φ Kontraktion auf E

  hinreichende Bedingung für Konvergenz der Fixpunktiteration.
- $n = 1$ (skalares Problem): geometrische Darstellung der Fixpunktiteration.

Nichtlineare Gleichungssysteme: Verfahren

- Konvergenzordnung: Maß für die Konvergenzgeschwindigkeit
- Fehlerschätzung hängt von p (Konvergenzordnung) ab.
  Skalare Folgen : einfache Formeln $p = 1$, $p > 1$
  Vektorfolgen : einfache Formel nur für $p > 1$.
- Methoden für skalare Probleme $f(x) = 0$:
  Bisektion, Newton-Verfahren, Sekanten-Verfahren.
- Newton-Verfahren:
  - $x^{k+1} = x^k + s^k$ mit $f'(x^k)\, s^k = -f(x^k)$
  - Konvergenzordnung: $p = 2$, falls $f'(x^*)$ regulär
  - lokale Konvergenz
  - Stopp-Kriterium: $\|x^* - x^k\| \approx \|x^{k+1} - x^k\|$ "klein"
  - Vereinfachtes Newton-Verfahren: feste Jacobi-Matrix
  - Wahl des Startvektors: Homotopieverfahren, problemabhängig
  - über Dämpfung kann man den Konvergenzbereich vergrößern

Das Gauß-Newton-Verfahren

Nichtlineares Ausgleichsproblem

Gegeben $F : \mathbb{R}^n \to \mathbb{R}^m$, bestimme $x^* \subset \mathbb{R}^n$, so dass

$$x^* \in \arg\min_{x \in \mathbb{R}^n} \|F(x)\|_2.$$

Ansatz:

1. Ersetze $F(x)$ durch lineare Approximation (Taylor-Entwicklung)
2. Schrittweise Annäherung an x^* durch Lösung linearer Probleme in jedem Schritt

Zur Erinnerung:

- Wir bezeichnen die Lösung am Iterationsschritt k mit
$$x^k = (x_1^k, \ldots, x_n^k)^T \in \mathbb{R}^n.$$
- Taylor-Entwicklung
$$F(x) = F(x^k) + F'(x^k)\,(x - x^k) + \mathcal{O}(\|x - x^k\|_2^2).$$

Das Gauß-Newton-Verfahren

Algorithmus 6.3. (Gauß-Newton)

Wähle Startwert x^0.
Für $k = 0, 1, 2, \ldots$:

1. Berechne $F(x^k), F'(x^k)$.
2. Finde s^k, so dass
$$s^k \in \arg\min_{s \in \mathbb{R}^n} \|F'(x^k)\, s + F(x^k)\|_2$$
3. Setze $x^{k+1} = x^k + s^k$.

Beachte

- Schritt 2 erfordert die Lösung eines linearen Ausgleichsproblems (Normalgleichung, QR-Zerlegung, SVD)
- Falls $F'(x)$ nicht vollen Rang hat, hat das Ausgleichsproblem keine eindeutige Lösung. $\to$ Wähle s^k mit minimaler 2-Norm.

Gauß-Newton als Fixpunktiteration

Gauss-Newton mit $\mathbf{Rang}\, F'(x) = n$ ist Fixpunktiteration mit

$$\Phi(x) = x - [F'(x)^T\, F'(x)]^{-1}\, F'(x)^T F(x).$$

Lokale Konvergenz um $x^* = \Phi(x^*)$ hängt von $\rho(K)\, \|F(x^*)\|_2$, wobei

$$K := A^{-1} \left(\sum_{i=1}^{m} \frac{F_i(x^*)}{\|F(x^*)\|_2}\, F_i''(x^*) \right) A^{-1}$$

und $A^2 = F'(x^*)^T F'(x^*)$.

- Falls x^* lokales Maximum oder Sattelpunkt von $\phi := \frac{1}{2}\|F\|_2^2$, gilt immer $\rho(K)\, \|F(x^*)\|_2 \geq 1$, also abstoßend.
- Falls x^* lokales Minima mit $\rho(K)\, \|F(x^*)\|_2 < 1$, dann lokale Konvergenz
- Falls x^* lokales Minima mit $\rho(K)\, \|F(x^*)\|_2 > 1$, dann abstoßend

Nichtlinearer Ausgleich: Zusammenfassung

- Beim Gauß-Newton-Verfahren wird das nichtlineare Ausgleichsproblem über eine Folge linearer Ausgleichsprobleme gelöst.
- Lokale Konvergenz im allgemeinen nur 1. Ordnung, falls $\boldsymbol{F}(\boldsymbol{x}^*) = \boldsymbol{0}$, sogar von 2. Ordnung.
- Es kann lokale Divergenz auftreten.
- Matrix $\boldsymbol{F}'(\boldsymbol{x}^k)$ kann Rang $< \boldsymbol{n}$ haben.
- Bei Levenberg-Marquardt:
  - Parameter $\boldsymbol{\mu}$ zur Vergrößerung des Einzugbereichs.
  - Matrix $\begin{pmatrix} \boldsymbol{F}'(\boldsymbol{x}^k) \\ \boldsymbol{\mu I} \end{pmatrix}$ hat stets Rang $= \boldsymbol{n}$.

Polynominterpolation

Lagrange-Interpolationsaufgabe: Es gibt eindeutiges $P_n \in \Pi_n$, mit

$$P_n(x_j) = f(x_j), \quad j = 0, 1, \dots, n$$

Effiziente Auswertung an einer Stelle mit Neville-Aitken-Schema

Werten Sie $P(f|x_0, x_1, x_2)$ an der Stelle $x = 0.5$ aus.

	$P_{i,0}$	$P_{i,1}$	$P_{i,2}$
$x_0 = 0$	$P_{0,0} = f(x_0) = 1$		
$x_1 = 1$	$P_{1,0} = f(x_1) = 4$	2.5	
$x_2 = 2$	$P_{2,0} = f(x_2) = 2$	5.0	$3\frac{1}{8}$

1. Einsetzen der gegebenen Werte
2. Auswerten der Rekursion

Effiziente Methode zur Auswertung der Potenzform

Sei $p \in \Pi_n$ ein Polynom, das in der Potenzform vorliegt, d.h.,

$$p(x) = a_0 + a_1 x + \cdots + a_n x^n$$

mit bekannten Koeffizienten $a_0, \ldots, a_n$. Es gilt

$$p(x) = a_0 + x \Big(a_1 + x \left(a_2 + \cdots + x \left(a_{n-1} + x\, a_n\right)\right)\Big).$$

Algorithmus 8.12. (Horner-Schema)

Setze $b_n = a_n$,

für $k = n-1, n-2, \ldots, 0$ berechne

$$b_k = a_k + x\, b_{k+1}$$

Dann ist $p(x) = b_0$.

Der Rechenaufwand wird etwa halbiert.

Lösungsdarstellung

Die eindeutige Lösung $P_n = P(f \mid x_0, \ldots, x_n)$ hat verschiedene Darstellungen abhängig von der Wahl der Basis in Π_n:

1. Lagrange-Basis

$$P_n(x) = \sum_{j=0}^{n} f(x_j)\, \ell_{jn}(x), \qquad \ell_{jn}(x) = \prod_{\substack{k=0\\ k\neq j}}^{n} \frac{x - x_k}{x_j - x_k}.$$

2. monomiale Basis

$$P_n(x) = \sum_{j=0}^{n} a_j\, x^j, \qquad V_n \begin{pmatrix} a_0 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} f(x_0) \\ \vdots \\ f(x_n) \end{pmatrix}.$$

3. Newtonsche Basis

$P_n(x) = \sum_{j=0}^{n} [x_0, \ldots, x_j] f\, \omega_j(x)$, Knotenpolynome ω_j und

$$[x_0, \ldots, x_n] f = \frac{[x_1, \ldots, x_n] f - [x_0, \ldots, x_{n-1}] f}{x_n - x_0}.$$

Zusammenfassung: Knotenwahl

Definiere

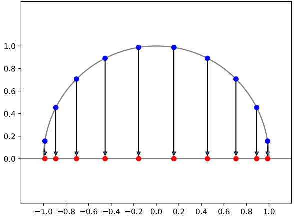
$$\omega_{n+1}(x) := \prod_{j=0}^{n} (x - x_j) \text{ und } M_{n+1}(f) := \max_{x\in[a,b]} \frac{\left|f^{(n+1)}(x)\right|}{(n+1)!}$$

Dann erhält man für die Fehlerschranke

$$|f(x) - P(f|x_0,\ldots,x_n)(x)| \leq \left|\omega_{n+1}(x)\right| M_{n+1}(f) .$$

Tschebyscheff-Knoten minimieren den Fehler der Knotenpolynome. Auf $[-1,1]$ durch gegeben:

$$x_j = \cos\left(\frac{2j+1}{2n+2}\pi\right), \; j = 0,\ldots,n$$



Interpolationsfehler

- Sei $a := \min\{x_0, \ldots, x_n\}$ und $b := \max\{x_0, \ldots, x_n\}$ mit n fest.
- Die Länge des Intervalls $h := b - a$ sei veränderbar.
- Falls $x \in I := [a, b]$ liegt, erhält man sofort die grobe Abschätzung
$$|\omega_{n+1}(x)| \leq h^{n+1},$$
und somit
$$\|f - P(f|x_0, \ldots, x_n)\|_{L_\infty(I)} \leq \frac{h^{n+1}}{(n+1)!} \|f^{n+1}\|_{L_\infty(I)}.$$
- Der Fehler wird kleiner, wenn die Stützstellen gemäß h zusammenrücken.

Dieser Effekt wird für die numerische Differentiation und Spline-Interpolation genutzt.

Interpolation: Bemerkungen

- Für die Numerik günstige Darstellung:
  Newtonsche Interpolationsformel
- Sehr hoher Polynomgrad: oft nicht günstig (Runge-Beispiel).
- Besser: fester Grad $+$ Wiederholung
  $+$ Verkleinerung des Interpolationsintervalls
- Bei der Numerischen Differentiation ist bei der Wahl von h ein Kompromiss zwischen Diskretisierungsfehler und Rundungsfehler (Auslöschung) zu finden

Allgemeine Quadraturformel

Satz 10.3 und Lemma 10.4

Für gegebene Knoten $c \le x_0 < \cdots < x_m \le d$ mit $h = d - c$ gibt es Gewichte $c_0, \ldots, c_m$ so dass

$$I_m(f) = \int_c^d P(f|x_0, \ldots, x_m)(x)\, dx = h \sum_{j=0}^{m} c_j\, f(x_j)\,.$$

$$c_j = \frac{1}{h} \int_c^d \prod_{\substack{k=0\\k\neq j}}^{m} \frac{x - x_k}{x_j - x_k} dx = \frac{1}{h} \int_c^d \ell_{jm}(x)\, dx\,.$$

Die Quadraturformel ist exakt vom Grade m, d.h.

für jedes Polynom $Q \in \Pi_m$ gilt $I_m(Q) = \int_c^d Q(x)\, dx$.

Newton-Cotes-Formeln

Newton-Cotes-Formeln für äquidistante Stützstellen

$$I_m(f) = h \sum_{j=0}^{m} c_j\, f(c+\xi_j\, h), \quad \text{wobei } \xi_j = \begin{cases} \frac{1}{2}, & j = m = 0 \\ \frac{j}{m}, & j = 0, \ldots, m > 0\,. \end{cases}$$

m		ξ_j	c_j	$I_m(f) - \int_c^d f(x)dx$
1	Trapezregel	$0, 1$	$\frac{1}{2}, \frac{1}{2}$	$\frac{1}{12}\, h^3\, f^{(2)}(\xi)$
2	Simpson-Regel	$0, \frac{1}{2}, 1$	$\frac{1}{6}, \frac{4}{6}, \frac{1}{6}$	$\frac{1}{90}\, \left(\frac{1}{2}h\right)^5\, f^{(4)}(\xi)$

Summierte Formel: $I_{m,n}(f) = \sum_{k=1}^{n} h \sum_{j=0}^{m} c_j\, f(t_{k-1} + \xi_j\, h)$

Fehlerschranke (Simpson): $E(h) = I_{m,n}(f) - \int_a^b f(x)\, dx$

$$E(h) = \frac{h^4}{2880} \sum_{k=1}^{n} h\, f^{(4)}(\xi_k) \begin{cases} \leq \frac{h^4}{2880}\, (b-a)\, \|f^{(4)}\|_\infty, \\ \approx \frac{h^4}{2880}\, \left(f^{(3)}(b) - f^{(3)}(a)\right)\,. \end{cases}$$

Gauß-Quadratur

Newton-Cotes c_j ändern ab $m \geq 8$ Vorzeichen $\Rightarrow$ Auslöschung.

Ziel: c_j positiv und maximaler Exaktheitsgrad $\Rightarrow$ Gauß-Quadratur

Satz 10.6

Sei $m \geq 0$. Es existieren Stützstellen $x_0, \ldots, x_m \in (c, d)$ und positive Gewichte $w_0, \ldots, w_m$, so dass mit $h = d - c$

$$h \sum_{i=0}^{m} w_i \, f(x_i) = \int_c^d f(x) \, dx + E_f(h)$$

und

$$E_Q = 0 \quad \text{für alle } Q \in \Pi_{2m+1}.$$

Ferner gilt für passendes $\xi \in [c, d]$

$$|E_f(h)| = \frac{\big((m+1)!\big)^4}{\big((2m+2)!\big)^3 \; (2m+3)} \, h^{2m+3} \, \left| f^{(2m+2)}(\xi) \right| .$$

Zusammenfassung: Newton-Cotes / Gauß-Quadratur

- Newton-Cotes: äquidistante Stützstellen $x_0, \dots, x_m$
  Idee: Integriere $P(f \mid x_0, \dots, x_m)(x)$.
- Gauß-Quadratur: spezifisch gewählte Stützstellen
  Idee: Finde $x_0, \dots, x_m$, $c_0, \dots, c_m$, so dass I_m exakt für alle $Q \in \Pi_{2m+1}$.

Unterschiede

- Newton-Cotes: einfacher, Exaktheitsgrad m oder $m+1$
- Gauß-Quadratur: Gewichte $c_j > 0$, Exaktheitsgrad $2\,m+1$

Verständnisfragen VF-1

Es seien x_{MIN} bzw. x_{MAX} die kleinste bzw. größte (strikt) positive Zahl sowie **eps** die relative Maschinengenauigkeit in der Menge $\mathbb{M}(b, m, r, R)$ der Maschinenzahlen gemäß Vorlesung/Buch und $\mathbb{D} := [-x_{\text{MAX}}, -x_{\text{MIN}}] \cup [x_{\text{MIN}}, x_{\text{MAX}}]$.
Ferner beschreibe $fl : \mathbb{D} \to \mathbb{M}(b, m, r, R)$ die Standardrundung, und es sei $\ominus$ (gem. Vorlesung/Buch) der Minusoperator für $\mathbb{M}$, d.h.: $x \ominus y := fl(fl(x) - fl(y))$ wobei wir hier annehmen, dass alle Zwischenergebnisse in $\mathbb{D}$ liegen.
Alle Zahlen sind im Dezimalsystem angegeben.

$\boxed{\text{f}}$ Für jedes $x \in \mathbb{D}$ existiert eine Zahl ϵ mit $|\epsilon| \leq$ **eps** und $fl(x) = x + \epsilon$.

$\boxed{\text{w}}$ Die Zahl 31 ist in $\mathbb{M}(2, 6, -8, 8)$ exakt darstellbar.

$\boxed{\text{w}}$ Es gilt $\frac{|(x \ominus y) - (x - y)|}{|x - y|} \leq$ **eps** für alle $x, y \in \mathbb{M}(b, m, r, R)$

Verständnisfragen VF-1

Es seien x_{MIN} bzw. x_{MAX} die kleinste bzw. größte (strikt) positive Zahl sowie **eps** die relative Maschinengenauigkeit in der Menge $\mathbb{M}(b, m, r, R)$ der Maschinenzahlen gemäß Vorlesung/Buch und $\mathbb{D} := [-x_{\mathrm{MAX}}, -x_{\mathrm{MIN}}] \cup [x_{\mathrm{MIN}}, x_{\mathrm{MAX}}]$.
Ferner beschreibe $\mathbf{fl} : \mathbb{D} \to \mathbb{M}(b, m, r, R)$ die Standardrundung, und es sei $\ominus$ (gem. Vorlesung/Buch) der Minusoperator für $\mathbb{M}$, d.h.: $x \ominus y := \mathbf{fl}(\mathbf{fl}(x) - \mathbf{fl}(y))$ wobei wir hier annehmen, dass alle Zwischenergebnisse in $\mathbb{D}$ liegen.
Alle Zahlen sind im Dezimalsystem angegeben.

- $\boxed{\text{f}}$ Für jedes $x \in \mathbb{D}$ existiert eine Zahl ϵ mit $|\epsilon| \leq$ **eps** und $\mathbf{fl}(x) = x + \epsilon$.
- $\boxed{\text{w}}$ Die Zahl $\mathbf{31}$ ist in $\mathbb{M}(2, 6, -8, 8)$ exakt darstellbar.
- $\boxed{\text{w}}$ Es gilt $\frac{|(x \ominus y) - (x - y)|}{|x - y|} \leq$ **eps** für alle $x, y \in \mathbb{M}(b, m, r, R)$ mit $x \neq y$.

Frage 1: Kann für (betragsmäßig) sehr große Zahlen ($|x| >> 1$) nicht gelten:

eps $\leftrightarrow$ relativer Fehler.

Frage 2: $\mathbf{31 = 11111_2 = 0.11111_2 \cdot 2^5}$.

Frage 3: Da x und y Maschinenzahlen sind und nach Voraussetzung alle Zwischenergebnisse in $\mathbb{D}$ liegen, steht hier ($z = x - y$) im Prinzip:

$$\frac{|\mathbf{fl}(z) - z|}{|z|} \leq \textbf{eps}$$

.

Verständnisfragen VF-1

Berechnen Sie x_{MAX} für $\mathbb{M}(3, 2, -1, 3)$. $\boxed{24}$

$\boxed{f}$ Es gilt $\frac{|(x \ominus y)-(x-y)|}{|x-y|} \leq$ **eps** für alle $x, y \in \mathbb{D}$ mit $x \neq y$.

$\boxed{f}$ Bei einem stabilen Algorithmus ist der Ausgabefehler nicht viel größer als der Eingabefehler.

$\boxed{f}$ Die Subtraktion zweier Zahlen mit demselben Vorzeichen ist immer schlecht konditioniert.

$\boxed{w}$ Die Funktion $f(x_1, x_2) = x_2\, e^{x_1}$ ist für alle (x_1, x_2) mit $|x_1| \leq 1$ gut konditioniert.

Es sei $f(x) = 1/(1 + x)$ und $\tilde{x}$ ein Näherungswert für $x = 3$, der mit einem relativen Fehler von maximal 2% behaftet ist. Bestimmen Sie in erster Näherung eine (scharfe) Schranke für den relativen Fehler in $f(\tilde{x})$ als Annäherung für $f(x)$. $\boxed{0.015}$

Verständnisfragen VF-1

Berechnen Sie x_{MAX} für $\mathbb{M}(3, 2, -1, 3)$. $\boxed{24}$

$\boxed{\text{f}}$ Es gilt $\frac{|(x \ominus y)-(x-y)|}{|x-y|} \leq \mathbf{eps}$ für alle $x, y \in \mathbb{D}$ mit $x \neq y$.

$\boxed{\text{f}}$ Bei einem stabilen Algorithmus ist der Ausgabefehler nicht viel größer als der Eingabefehler.

$\boxed{\text{f}}$ Die Subtraktion zweier Zahlen mit demselben Vorzeichen ist immer schlecht konditioniert.

$\boxed{\text{w}}$ Die Funktion $f(x_1, x_2) = x_2\, e^{x_1}$ ist für alle (x_1, x_2) mit $|x_1| \leq 1$ gut konditioniert.

Es sei $f(x) = 1/(1+x)$ und $\tilde{x}$ ein Näherungswert für $x = 3$, der mit einem relativen Fehler von maximal 2% behaftet ist. Bestimmen Sie in erster Näherung eine (scharfe) Schranke für den relativen Fehler in $f(\tilde{x})$ als Annäherung für $f(x)$. $\boxed{0.015}$

Frage 1: $x_{\mathrm{MAX}} = 0.22_3 \cdot 3^3 = \left(\frac{2}{3} + \frac{2}{9}\right) \cdot 27 = 18 + 6 = 24$.

Frage 2: Jetzt sind x und y **keine** Maschinenzahlen. Daher ist bedingt durch die Eingangsrundung die schlechte Kondition der Subtraktion möglich.

Frage 3: Man kann die Kondition nicht umgehen.
Beispiel:
$f(x) = x^{1000},\ x = 1 \ \rightarrow\ \tilde{x} = 1.0001$. Dieser Eingangsfehler von 0.01% wird (bei exakter Rechnung) zu einm Ausgabefehler von mehr als 10.5%

Frage 4: Gegenbeispiel: $x = 101,\ y = 1\ :\rightarrow\ \kappa_{rel} = 1.01$, also gut konditioniert.

Frage 5: Man rechnet leicht nach: $\kappa_{rel,x_1} = |x_1|$ und $\kappa_{rel,x_2} = 1$.

Frage 6: Hier ist $\kappa_{rel}(x) = \left|-\dfrac{1}{(1+x)^2}\, x\right| \Big/ \left|\dfrac{1}{1+x}\right| = \left|\dfrac{x}{1+x}\right|$.
Also $\kappa_{rel}(3) = 0.75$ und somit werden aus den 2% in der Eingabe in erster Näherung $1.5\% = 0.015$ in der Ausgabe.

Verständnisfragen VF-2

Es seien $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, beliebig aber regulär, und $\boldsymbol{b} \in \mathbb{R}^{n}$ und gesucht sei die Lösung $\boldsymbol{x} \in \mathbb{R}^{n}$ von $\boldsymbol{A}\,\boldsymbol{x} = \boldsymbol{b}$.

$\boxed{\text{f}}$ Es seien $\tilde{\boldsymbol{x}}$ eine Annäherung der Lösung $\boldsymbol{x}$ und $\boldsymbol{r} := \boldsymbol{b} - \boldsymbol{A}\,\tilde{\boldsymbol{x}}$ das zugehörige Residuum.
Es gilt $\|\boldsymbol{r}\| \leq \boldsymbol{\kappa}(\boldsymbol{A})\,\|\tilde{\boldsymbol{x}} - \boldsymbol{x}\|$, mit $\boldsymbol{\kappa}(\boldsymbol{A}) := \|\boldsymbol{A}\|\,\|\boldsymbol{A}^{-1}\|$.

$\boxed{\text{f}}$ Es existieren stets eine untere Dreiecksmatrix $\boldsymbol{L}$ und eine obere Dreiecksmatrix $\boldsymbol{R}$, so dass $\boldsymbol{A} = \boldsymbol{L}\,\boldsymbol{R}$ gilt.

$\boxed{\text{w}}$ Falls $\boldsymbol{A}$ orthogonal ist, gilt $\boldsymbol{A}^{T}\,\boldsymbol{A} = \boldsymbol{I}$.

$\boxed{\text{w}}$ Es seien $\boldsymbol{B} \in \mathbb{R}^{n \times n}$ beliebig aber regulär und $\boldsymbol{\kappa}(\cdot)$ die Konditionszahl bzgl. $\|\cdot\|$. Es gilt $\boldsymbol{\kappa}(\boldsymbol{A}\,\boldsymbol{B}) \leq \boldsymbol{\kappa}(\boldsymbol{A})\,\boldsymbol{\kappa}(\boldsymbol{B})$.

Es sei $\boldsymbol{B} := \boldsymbol{D}\,\boldsymbol{A}$ die zeilenäquilibrierte Matrix zu $\boldsymbol{A}$.
Geben Sie $\|\boldsymbol{B}\|_{\infty}$ an. $\boxed{1}$

$\boxed{\text{f}}$ Für die Matrix $\boldsymbol{A} = \begin{pmatrix} \mathbf{2} & \mathbf{0} \\ \text{-}\mathbf{1} & \mathbf{3} \end{pmatrix}$ existiert eine Cholesky-Zerlegung

Verständnisfragen VF-2

Es seien $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, beliebig aber regulär, und $\boldsymbol{b} \in \mathbb{R}^n$ und gesucht sei die Lösung $\boldsymbol{x} \in \mathbb{R}^n$ von $\boldsymbol{A}\,\boldsymbol{x} = \boldsymbol{b}$.

- [f] Es seien $\tilde{\boldsymbol{x}}$ eine Annäherung der Lösung $\boldsymbol{x}$ und $\boldsymbol{r} := \boldsymbol{b} - \boldsymbol{A}\,\tilde{\boldsymbol{x}}$ das zugehörige Residuum.
  Es gilt $\|\boldsymbol{r}\| \leq \kappa(\boldsymbol{A})\,\|\tilde{\boldsymbol{x}} - \boldsymbol{x}\|$, mit $\kappa(\boldsymbol{A}) := \|\boldsymbol{A}\|\,\|\boldsymbol{A}^{-1}\|$.
- [f] Es existieren stets eine untere Dreiecksmatrix $\boldsymbol{L}$ und eine obere Dreiecksmatrix $\boldsymbol{R}$, so dass $\boldsymbol{A} = \boldsymbol{L}\,\boldsymbol{R}$ gilt.
- [w] Falls $\boldsymbol{A}$ orthogonal ist, gilt $\boldsymbol{A}^T\,\boldsymbol{A} = \boldsymbol{I}$.
- [w] Es seien $\boldsymbol{B} \in \mathbb{R}^{n\times n}$ beliebig aber regulär und $\kappa(\cdot)$ die Konditionszahl bzgl. $\|\cdot\|$. Es gilt $\kappa(\boldsymbol{A}\,\boldsymbol{B}) \leq \kappa(\boldsymbol{A})\,\kappa(\boldsymbol{B})$.

Es sei $\boldsymbol{B} := \boldsymbol{D}\,\boldsymbol{A}$ die zeilenäquilibrierte Matrix zu $\boldsymbol{A}$.
Geben Sie $\|\boldsymbol{B}\|_\infty$ an. [1]

- [f] Für die Matrix $\boldsymbol{A} = \begin{pmatrix} 2 & 0 \\ -1 & 3 \end{pmatrix}$ existiert eine Cholesky-Zerlegung

Frage 1: $\|\boldsymbol{r}\| = \|\boldsymbol{b} - \boldsymbol{A}\,\tilde{\boldsymbol{x}}\| = \|\boldsymbol{A}\,\boldsymbol{x} - \boldsymbol{A}\,\tilde{\boldsymbol{x}}\| = \|\boldsymbol{A}\,(\boldsymbol{x} - \tilde{\boldsymbol{x}})\| \leq \|\boldsymbol{A}\| \cdot \|\boldsymbol{x} - \tilde{\boldsymbol{x}}\|$
$\|\boldsymbol{A}^{-1}\|$ kann sehr klein sein. Daher ist diese Aussage falsch.

Frage 2: Gegenbeispiel: $\boldsymbol{A} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$. Die Gauß-Elimination zur Berechnung von $\boldsymbol{L}\,\boldsymbol{R}$ ist nicht immer ohne Pivotisierung möglich.

Frage 3: Definition von orthogonal.

Frage 4: $\kappa(\boldsymbol{A}\,\boldsymbol{B}) = \|\boldsymbol{A}\,\boldsymbol{B}\|\,\|(\boldsymbol{A}\,\boldsymbol{B})^{-1}\|$

$$\left.\begin{array}{r} \|\boldsymbol{A}\,\boldsymbol{B}\| \leq \|\boldsymbol{A}\|\,\|\boldsymbol{B}\| \\ \|(\boldsymbol{A}\,\boldsymbol{B})^{-1}\| = \|\boldsymbol{B}^{-1}\,\boldsymbol{A}^{-1}\| \leq \|\boldsymbol{B}^{-1}\|\;\|\boldsymbol{A}^{-1}\| \end{array}\right\} \text{Behauptung}$$

Frage 5: Zeilenäquilibrieren $d_{ii} = 1/\sum_j |a_{ij}|$ ("in jeder Zeile")
also so konzipiert, dass $\|\boldsymbol{D}\boldsymbol{A}\|_\infty = 1 = \|\boldsymbol{B}\|_\infty$

Frage 6: $\boldsymbol{A}$ ist nicht symmetrisch.

Verständnisfragen VF-2

Es seien $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, beliebig aber regulär, $\boldsymbol{b} \in \mathbb{R}^{n}$ und gesucht sei die Lösung $\boldsymbol{x} \in \mathbb{R}^{n}$ von $\boldsymbol{A}\,\boldsymbol{x} = \boldsymbol{b}$.

f Der Rechenaufwand zur Bestimmung der Lösung $\boldsymbol{x}$ über die Gauß-Elimination mit Spaltenpivotisierung beträgt etwa $\frac{1}{6}\,n^3$ Operationen (gem. Vorlesung/Buch).

f Pivotisierung verbessert die Kondition der Gauß-Elimination.

w Es sei $\boldsymbol{P}\,\boldsymbol{A} = \boldsymbol{L}\,\boldsymbol{R}$ die über den Gauß-Algorithmus mit Spaltenpivotisierung berechnete Faktorisierung. Dann gilt:

$$\left|\det\left(\boldsymbol{A}^{-1}\right)\right| = \frac{1}{|\det(\boldsymbol{R})|}.$$

Es sei $\boldsymbol{A} = \begin{pmatrix} 3 & 2 & 5 \\ -1 & 0 & 1 \\ 2 & -2 & -2 \end{pmatrix}$. Berechnen Sie $\|\boldsymbol{A}\|_1$. $\boxed{8}$

Verständnisfragen VF-2

Es seien $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, beliebig aber regulär, $\boldsymbol{b} \in \mathbb{R}^{n}$ und gesucht sei die Lösung $\boldsymbol{x} \in \mathbb{R}^{n}$ von $\boldsymbol{A}\,\boldsymbol{x} = \boldsymbol{b}$.

$\boxed{\text{f}}$ Der Rechenaufwand zur Bestimmung der Lösung $\boldsymbol{x}$ über die Gauß-Elimination mit Spaltenpivotisierung beträgt etwa $\frac{1}{6}\,n^3$ Operationen (gem. Vorlesung/Buch).

$\boxed{\text{f}}$ Pivotisierung verbessert die Kondition der Gauß-Elimination.

$\boxed{\text{w}}$ Es sei $\boldsymbol{P}\,\boldsymbol{A} = \boldsymbol{L}\,\boldsymbol{R}$ die über den Gauß-Algorithmus mit Spaltenpivotisierung berechnete Faktorisierung. Dann gilt:

$$\left|\det\left(\boldsymbol{A}^{-1}\right)\right| = \frac{1}{|\det(\boldsymbol{R})|}.$$

Es sei $\boldsymbol{A} = \begin{pmatrix} 3 & 2 & 5 \\ -1 & 0 & 1 \\ 2 & -2 & -2 \end{pmatrix}$. Berechnen Sie $\|\boldsymbol{A}\|_1$. $\boxed{8}$

Frage 1: Gauß/$\boldsymbol{LR}$-Zerlegung $\frac{1}{3}\,\boldsymbol{n}^{3}$, $\frac{1}{6}\,\boldsymbol{n}^{3}$ ist Cholesky/$\boldsymbol{L}\,\boldsymbol{D}\,\boldsymbol{L}^{T}$-Zerlegung

Frage 2: Pivotisierung stabilisiert (den Algorithmus)
Zusatz: äquilibrierung liefert eine Matrix mit i. A. besserer ∞-Kondition.

Frage 3: Da $\boldsymbol{P}\,\boldsymbol{A} = \boldsymbol{LR}$ gilt $\det\boldsymbol{P} \cdot \det\boldsymbol{A} = \det\boldsymbol{L} \cdot \det\boldsymbol{R}$ also

$$\pm\det\boldsymbol{A} = \det\boldsymbol{R}$$

Da aus $\boldsymbol{A}\,\boldsymbol{A}^{-1} = \boldsymbol{I}$ zudem $\det\boldsymbol{A} \cdot \det\left(\boldsymbol{A}^{-1}\right) = 1$ folgt, gilt

$$\left|\det\left(\boldsymbol{A}^{-1}\right)\right| = \frac{1}{|\det\boldsymbol{A}|} = \frac{1}{|\det\boldsymbol{R}|}$$

Frage 4: $\|\cdot\|_1$ Spaltensummennorm $\max\{6, 4, 8\} = 8$

Verständnisfragen VF-3

Es seien $\boldsymbol{A}$ eine symmetrisch positiv definite $\boldsymbol{n} \times \boldsymbol{n}$-Matrix, $\boldsymbol{b} \in \mathbb{R}^{\boldsymbol{n}}$ und $\boldsymbol{A} = \boldsymbol{L}\,\boldsymbol{D}\,\boldsymbol{L}^{\boldsymbol{T}}$ die Cholesky-Zerlegung von $\boldsymbol{A}$.

- **f** Es gilt $\|\boldsymbol{A}\|_2 = \|\boldsymbol{D}\|_2$.
- **f** Das Problem $\boldsymbol{A}\,\boldsymbol{x} = \boldsymbol{b}$ ist immer gut konditioniert.
- **w** Das Cholesky-Vefahren zur Bestimmung der Cholesky-Zerlegung ist ein stabiles Verfahren.
- **f** Für die stabile Berechnung einer $\boldsymbol{L}\,\boldsymbol{R}$-Zerlegung $\boldsymbol{A} = \boldsymbol{L}\,\boldsymbol{R}$ von $\boldsymbol{A}$ ist Pivotisierung notwendig.

Es seien $\boldsymbol{Q}$ eine orthogonale Matrix und $\boldsymbol{Q}\begin{pmatrix} \mathbf{5} \\ \mathbf{0} \\ \mathbf{12} \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \boldsymbol{c} \\ \mathbf{0} \end{pmatrix}$.
Geben Sie $|\boldsymbol{c}|$ an. **13**

- **f** Es gilt $\boldsymbol{A}^{-1} = \boldsymbol{L}\,\boldsymbol{D}^{-1}\,\boldsymbol{L}^{\boldsymbol{T}}$.

Verständnisfragen VF-3

Es seien $\boldsymbol{A}$ eine symmetrisch positiv definite $n \times n$-Matrix, $\boldsymbol{b} \in \mathbb{R}^n$ und $\boldsymbol{A} = \boldsymbol{L}\,\boldsymbol{D}\,\boldsymbol{L}^T$ die Cholesky-Zerlegung von $\boldsymbol{A}$.

$\boxed{\text{f}}$ Es gilt $\|\boldsymbol{A}\|_2 = \|\boldsymbol{D}\|_2$.

$\boxed{\text{f}}$ Das Problem $\boldsymbol{A}\,\boldsymbol{x} = \boldsymbol{b}$ ist immer gut konditioniert.

$\boxed{\text{w}}$ Das Cholesky-Vefahren zur Bestimmung der Cholesky-Zerlegung ist ein stabiles Verfahren.

$\boxed{\text{f}}$ Für die stabile Berechnung einer $\boldsymbol{L}\,\boldsymbol{R}$-Zerlegung $\boldsymbol{A} = \boldsymbol{L}\,\boldsymbol{R}$ von $\boldsymbol{A}$ ist Pivotisierung notwendig.

Es seien $\boldsymbol{Q}$ eine orthogonale Matrix und $\boldsymbol{Q}\begin{pmatrix}5\\0\\12\end{pmatrix} = \begin{pmatrix}0\\c\\0\end{pmatrix}$.
Geben Sie $|c|$ an. $\boxed{13}$

$\boxed{\text{f}}$ Es gilt $\boldsymbol{A}^{-1} = \boldsymbol{L}\,\boldsymbol{D}^{-1}\,\boldsymbol{L}^T$.

Frage 1: Dazu müsste z. B. $\boldsymbol{L}$ orthogonal sein. (Also $\|\boldsymbol{L}\| = \|\boldsymbol{I}\|$). Gegenbeispiel:

$$\boldsymbol{L} = \begin{pmatrix}1 & 0\\1 & 1\end{pmatrix}, \quad \boldsymbol{D} = \boldsymbol{I}, \quad \boldsymbol{L}\,\boldsymbol{D}\,\boldsymbol{L}^T = \begin{pmatrix}1 & 1\\1 & 2\end{pmatrix} \text{ offenbar } \|\boldsymbol{I}\|_1 \neq \left\|\begin{pmatrix}1 & 1\\1 & 2\end{pmatrix}\right\|_1$$

Frage 2: Es gibt auch schlecht konditionierte “s.p.d.” Probleme.
Beispielsweise Hilbert-Matrizen. (hier mit Shift)

$$\boldsymbol{H} = \begin{pmatrix}\frac{1}{1000} & \frac{1}{1001}\\ \frac{1}{1001} & \frac{1}{1002}\end{pmatrix}, \quad \kappa_2(\boldsymbol{H}) = 42.1513$$

Frage 3: Cholesky ist von der Stabilität mit Skalierung und Pivotisierung bei nicht s.p.d. Matrizen vergleichbar.(Pivotisierung weder notwendig noch sinvoll.)

Frage 4: Da $\boldsymbol{A}$ s.p.d., hier nicht erforderlich.

Frage 5: Anwendung von $\boldsymbol{Q}$ ändert Länge nicht. $\rightarrow |c| = \sqrt{5^2 + 12^2} = \sqrt{169} = 13$

Frage 6: $\boldsymbol{A} = \boldsymbol{L}\,\boldsymbol{D}\,\boldsymbol{L}^T \Rightarrow \boldsymbol{A}^{-1} = \boldsymbol{L}^{-T}\,\boldsymbol{D}^{-1}\,\boldsymbol{L}^{-1}$

Verständnisfragen VF-3

Es seien $\boldsymbol{A}$ eine symmetrisch positiv definite $\boldsymbol{n} \times \boldsymbol{n}$-Matrix, $\boldsymbol{b} \in \mathbb{R}^{\boldsymbol{n}}$ und $\boldsymbol{A} = \boldsymbol{L}\,\boldsymbol{D}\,\boldsymbol{L}^{\boldsymbol{T}}$ die Cholesky-Zerlegung von $\boldsymbol{A}$.

w Das Produkt zweier Givens-Rotations-Matrizen ist eine orthogonale Matrix.

f Es seien $\boldsymbol{Q_v} \in \mathbb{R}^{\boldsymbol{m}\times\boldsymbol{m}}$ eine Householder-Transformations-Matrix und $\boldsymbol{x} \in \mathbb{R}^{\boldsymbol{m}}$ beliebig. Es gilt $\|\boldsymbol{Q_v}\,\boldsymbol{x}\|_\infty = \|\boldsymbol{x}\|_\infty$.

f Die Berechnung einer $\boldsymbol{QR}$-Zerlegung $\boldsymbol{B} = \boldsymbol{Q}\,\boldsymbol{R}$ von $\boldsymbol{B} \in \mathbb{R}^{\boldsymbol{m}\times\boldsymbol{n}}$ über Householder-Transformationen ist nur dann stabil, wenn die Matrix $\boldsymbol{B}$ vollen Spaltenrang hat.

Es sei $\boldsymbol{Q_v}$ eine Householder-Transformation. Geben Sie den Wert des größten Eigenwertes der Matrix $\boldsymbol{Q_v}$ an. **1**

Verständnisfragen VF-3

Es seien $\boldsymbol{A}$ eine symmetrisch positiv definite $\boldsymbol{n} \times \boldsymbol{n}$-Matrix, $\boldsymbol{b} \in \mathbb{R}^{\boldsymbol{n}}$ und $\boldsymbol{A} = \boldsymbol{L}\,\boldsymbol{D}\,\boldsymbol{L}^{\boldsymbol{T}}$ die Cholesky-Zerlegung von $\boldsymbol{A}$.

$\boxed{\text{w}}$ Das Produkt zweier Givens-Rotations-Matrizen ist eine orthogonale Matrix.

$\boxed{\text{f}}$ Es seien $\boldsymbol{Q_v} \in \mathbb{R}^{\boldsymbol{m}\times\boldsymbol{m}}$ eine Householder-Transformations-Matrix und $\boldsymbol{x} \in \mathbb{R}^{\boldsymbol{m}}$ beliebig. Es gilt $\|\boldsymbol{Q_v}\,\boldsymbol{x}\|_\infty = \|\boldsymbol{x}\|_\infty$.

$\boxed{\text{f}}$ Die Berechnung einer $\boldsymbol{QR}$-Zerlegung $\boldsymbol{B} = \boldsymbol{Q}\,\boldsymbol{R}$ von $\boldsymbol{B} \in \mathbb{R}^{\boldsymbol{m}\times\boldsymbol{n}}$ über Householder-Transformationen ist nur dann stabil, wenn die Matrix $\boldsymbol{B}$ vollen Spaltenrang hat.

Es sei $\boldsymbol{Q_v}$ eine Householder-Transformation. Geben Sie den Wert des größten Eigenwertes der Matrix $\boldsymbol{Q_v}$ an. $\boxed{1}$

Frage 1: Das Produkt orthogonaler Matrizen ist orthogonal.

$$(\boldsymbol{Q_1}\,\boldsymbol{Q_2})\,(\boldsymbol{Q_1}\,\boldsymbol{Q_2})^{\boldsymbol{T}} = \boldsymbol{Q_1}\,(\boldsymbol{Q_2}\,\boldsymbol{Q_2^T})\,\boldsymbol{Q_1^T} = \boldsymbol{Q_1}\,\boldsymbol{Q_1^T} = \boldsymbol{I}$$

Frage 2: "falsche" Norm. Gilt nur für die $\boldsymbol{2}$-Norm.

Frage 3: Orthogonale Matrizen bilden auch für Matrizen mit Rangdefekt ein stabiles Verfahren zur Berechnung der $\boldsymbol{QR}$-Zerlegung

Frage 4: $\boldsymbol{Q}\,\boldsymbol{x} = \boldsymbol{\lambda}\,\boldsymbol{x}$: für $\|\cdot\|_{\boldsymbol{2}}$ gilt

$$\|\boldsymbol{x}\|_{\boldsymbol{2}} = \|\boldsymbol{Q}\,\boldsymbol{x}\|_{\boldsymbol{2}} = \|\boldsymbol{\lambda}\,\boldsymbol{x}\|_{\boldsymbol{2}} = |\boldsymbol{\lambda}|\;\|\boldsymbol{x}\|_{\boldsymbol{2}} \text{ also } |\boldsymbol{\lambda}| = \boldsymbol{1}$$

Verständnisfragen VF-4

Es seien $A \in \mathbb{R}^{m\times n}$, mit $\mathbf{Rang}(A) = n < m$, und $b \in \mathbb{R}^m$. Weiter seien $Q \in \mathbb{R}^{m\times m}$ eine orthogonale Matrix und $R \in \mathbb{R}^{m\times n}$ eine obere Dreiecksmatrix so, dass $Q\,A = R = \begin{pmatrix} \widetilde{R} \\ 0 \end{pmatrix}$ gilt, mit $\widetilde{R} \in \mathbb{R}^{n\times n}$. Es sei $x^* = \arg\min_{x\in\mathbb{R}^n} \|A\,x - b\|_2 \in \mathbb{R}^n$. Weiter sei $\Theta \in \left[0, \frac{\pi}{2}\right)$ der Winkel zwischen $A\,x^*$ und b.

- **w** Je kleiner der Winkel Θ, desto kleiner ist die Größe $\frac{\|A\,x^* - b\|_2}{\|b\|_2}$.
- **f** Es gilt $\tilde{R}\,x^* = Q^T\,b$.
- **w** Die Matrix $\tilde{R}$ kann man über Givens-Rotationen bestimmen.
- **f** Es gilt $\det\widetilde{R} = \det A$.

Verständnisfragen VF-4

Es seien $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, mit $\mathbf{Rang}(\boldsymbol{A}) = n < m$, und $\boldsymbol{b} \in \mathbb{R}^{m}$.
Weiter seien $\boldsymbol{Q} \in \mathbb{R}^{m \times m}$ eine orthogonale Matrix und $\boldsymbol{R} \in \mathbb{R}^{m \times n}$ eine obere Dreiecksmatrix so, dass $\boldsymbol{Q}\,\boldsymbol{A} = \boldsymbol{R} = \begin{pmatrix} \widetilde{\boldsymbol{R}} \\ \mathbf{0} \end{pmatrix}$ gilt, mit $\widetilde{\boldsymbol{R}} \in \mathbb{R}^{n \times n}$. Es sei $\boldsymbol{x}^* = \arg\min_{x \in \mathbb{R}^n} \|\boldsymbol{A}\,\boldsymbol{x} - \boldsymbol{b}\|_2 \in \mathbb{R}^n$.
Weiter sei $\boldsymbol{\Theta} \in [\mathbf{0}, \frac{\pi}{2})$ der Winkel zwischen $\boldsymbol{A}\,\boldsymbol{x}^*$ und $\boldsymbol{b}$.

- $\boxed{w}$ Je kleiner der Winkel $\boldsymbol{\Theta}$, desto kleiner ist die Größe $\frac{\|\boldsymbol{A}\,\boldsymbol{x}^* - \boldsymbol{b}\|_2}{\|\boldsymbol{b}\|_2}$.
- $\boxed{f}$ Es gilt $\tilde{\boldsymbol{R}}\,\boldsymbol{x}^* = \boldsymbol{Q}^T\,\boldsymbol{b}$.
- $\boxed{w}$ Die Matrix $\tilde{\boldsymbol{R}}$ kann man über Givens-Rotationen bestimmen.
- $\boxed{f}$ Es gilt $\mathbf{det}\widetilde{\boldsymbol{R}} = \mathbf{det}\boldsymbol{A}$.

Frage 1: $\sin\boldsymbol{\theta} = \dfrac{\|\boldsymbol{A}\,\boldsymbol{x}^* - \boldsymbol{b}\|_2}{\|\boldsymbol{b}\|_2}$

Frage 2: "doppelt" falsch.

a) Dimensionen passen nicht $\widetilde{\boldsymbol{R}}\,\boldsymbol{x}^* \in \mathbb{R}^{n}$, $\boldsymbol{Q}^{T}\,\boldsymbol{b} \in \mathbb{R}^{m}$

b) $\boldsymbol{Q}\,\boldsymbol{A} = \boldsymbol{R}$: Aber mit $\begin{pmatrix} \boldsymbol{b}_1 \\ \boldsymbol{b}_2 \end{pmatrix} = \boldsymbol{Q}\,\boldsymbol{b}$ plus $\widetilde{\boldsymbol{R}}\,\boldsymbol{x}^* = \boldsymbol{b}_1$

Frage 3: Givens-Rotationen sind orthogonale Transformationen

Frage 4: $\mathbf{det}\boldsymbol{A}$ ist nicht-quadratische Matrizen nicht definiert.

Verständnisfragen VF-4

w Es gilt $\|A\,x - b\|_2 = \|R\,x - Q\,b\|_2$ für beliebiges $x \in \mathbb{R}^n$.

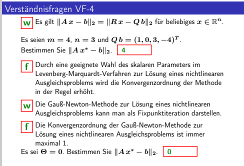
Es seien $m = 4$, $n = 3$ und $Q\,b = (1, 0, 3, -4)^T$.
Bestimmen Sie $\|A\,x^* - b\|_2$. **4**

f Durch eine geeignete Wahl des skalaren Parameters im Levenberg-Marquardt-Verfahren zur Lösung eines nichtlinearen Ausgleichsproblems wird die Konvergenzordnung der Methode in der Regel erhöht.

w Die Gauß-Newton-Methode zur Lösung eines nichtlinearen Ausgleichsproblems kann man als Fixpunktiteration darstellen.

f Die Konvergenzordnung der Gauß-Newton-Methode zur Lösung eines nichtlinearen Ausgleichsproblems ist immer maximal 1.

Es sei $\Theta = 0$. Bestimmen Sie $\|A\,x^* - b\|_2$. **0**



Frage 1: $\|A\,x - b\|_2 = \|Q(A\,x - b)\|_2 = \|R\,x - Q\,b\|_2$

Frage 2: Es gilt: $\|A\,x^* - b\|_2 = \|R\,x^* - Q\,b\|_2 = \left\| \begin{pmatrix} \widetilde{R}\,x^* \\ 0 \end{pmatrix} - \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \right\|_2.$

$\widetilde{R}\,x^* = b_1$, $\|b_2\|_2$ Residuum. Hier $b_2 = (-4)$.

Frage 3: Der Parameter dient zur Dämpfung der Korrektur. Ordnung bleibt. Konvergiert möglicherweise langsamer, dafür aber auch dann, wenn Gauß-Newton nicht konvergiert.

Frage 4: Formal $x^{k+1} = x^k - \left(F'(x^k)^T\, F'(x^k)\right)^{-1}\, F'(x^k)^T\, x^k$

Frage 5: Kann auch höher sein.
Falls $\|F(x^*)\| = 0$ mindestens 2 (lokal!)

Frage 6: Mit der erste Frage in diesem Abschnitt folgt $0 = \sin 0 = \dfrac{\|A\,x^* - b\|_2}{\|b\|_2}$, also

$\|A\,x^* - b\|_2 = 0.$

Verständnisfragen VF-5

Es seien $\mathbf{\Phi} : \mathbb{R}^n \to \mathbb{R}^n$ stetig differenzierbar und $\boldsymbol{x}^*$ so, dass $\mathbf{\Phi}(\boldsymbol{x}^*) = \boldsymbol{x}^*$ gilt. Für $\boldsymbol{x}_0 \in \mathbb{R}^n$ wird die Fixpunktiteration $\boldsymbol{x}_{k+1} = \mathbf{\Phi}(\boldsymbol{x}_k)$, $k = 0, 1, 2, \ldots$ definiert.
Weiter sei $\mathbf{\Phi}'(\boldsymbol{x})$ die Ableitung von $\mathbf{\Phi}$ an der Stelle $\boldsymbol{x}$.
Für $n = 1$ sei außerdem $\Phi_1(x) := \frac{1}{4}x^2 - 1$.

- **f** Es gilt $\|\mathbf{\Phi}'(\boldsymbol{x}^*)\| < 1$.
- **f** Die Konvergenzordung der Fixpunktiteration ist maximal 2.
- **f** Das Fixpunktproblem $\Phi_1(x) = x$ hat eine eindeutige Lösung $x^* \in \mathbb{R}$.
- **w** Für Φ_1 sind auf dem Intervall $[-1, 0]$ alle Voraussetzungen des Banachschen Fixpunktsatzes erfüllt.

Verständnisfragen VF-5

Es seien $\Phi : \mathbb{R}^n \to \mathbb{R}^n$ stetig differenzierbar und x^* so, dass $\Phi(x^*) = x^*$ gilt. Für $x_0 \in \mathbb{R}^n$ wird die Fixpunktiteration $x_{k+1} = \Phi(x_k)$, $k = 0, 1, 2, \ldots$ definiert.
Weiter sei $\Phi'(x)$ die Ableitung von Φ an der Stelle x.
Für $n = 1$ sei außerdem $\Phi_1(x) := \frac{1}{4}x^2 - 1$.

f Es gilt $\|\Phi'(x^*)\| < 1$.

f Die Konvergenzordung der Fixpunktiteration ist maximal 2.

f Das Fixpunktproblem $\Phi_1(x) = x$ hat eine eindeutige Lösung $x^* \in \mathbb{R}$.

w Für Φ_1 sind auf dem Intervall $[-1, 0]$ alle Voraussetzungen des Banachschen Fixpunktsatzes erfüllt.

Frage 1: Es gibt auch "abstoßende" Fixpunkte

Frage 2: $\Phi(x) = x^p$ hat Konvergenzordnung p, denn $x^* = 0$,

$$\Phi'(x^*) = \ldots = \Phi^{(p-1)}(x^*) = 0$$

und

$$\Phi^{(p)}(x^*) = p! \neq 0.$$

Frage 3: $\Phi_1(x) = \frac{1}{4}x^2 - 1 = x$ hat genau 2 Lösungen.

$$x^2 - 4x - 4 = 0 \quad \Leftrightarrow \quad x_{1,2} = 2 \pm \sqrt{8}$$

Frage 4: Φ_1 ist auf $[-1, 0]$ monoton fallend, also $\Phi_1([-1, 0]) \subset [-1, -\frac{3}{4}]$

$\Phi_1'(x) = \frac{1}{2}x$ also $|\Phi_1'(x)| \leq \frac{1}{2}$ auf $[-1, 0]$.

Verständnisfragen VF-5

Wir betrachten die Fixpunktiteration zur Bestimmung einer Lösung $x^* < 0$ des Fixpunktproblems $\Phi_1(x) = x$, mit einem Startwert x_0 aus einer hinreichend kleinen Umgebung von x^*. Geben Sie die Konvergenzordnung dieser Methode an. $\boxed{1}$

$\boxed{f}$ Bei der Sekantenmethode zur Bestimmung einer Nullstelle einer skalaren Funktion f, müssen die Startwerte x_0, x_1 so gewählt werden, dass $f(x_0)\, f(x_1) < 0$ gilt.

$\boxed{w}$ Es sei $f(x) = x^2 - 3$. Das auf f angewandte Newton-Verfahren konvergiert für jeden Startwert $x_0 > 0$ gegen die Nullstelle $x^* > 0$ dieser Funktion.

$\boxed{f}$ Eine Dämpfungsstrategie beim Newton-Verfahren zur Bestimmung einer Nullstelle kann man nur bei skalarwertigen Funktionen $f : \mathbb{R} \to \mathbb{R}$ anwenden.

Verständnisfragen VF-5

Wir betrachten die Fixpunktiteration zur Bestimmung einer Lösung $x^* < 0$ des Fixpunktproblems $\Phi_1(x) = x$, mit einem Startwert x_0 aus einer hinreichend kleinen Umgebung von x^*. Geben Sie die Konvergenzordnung dieser Methode an. $\boxed{1}$

- $\boxed{\text{f}}$ Bei der Sekantenmethode zur Bestimmung einer Nullstelle einer skalaren Funktion f, müssen die Startwerte x_0, x_1 so gewählt werden, dass $f(x_0)\,f(x_1) < 0$ gilt.
- $\boxed{\text{w}}$ Es sei $f(x) = x^2 - 3$. Das auf f angewandte Newton-Verfahren konvergiert für jeden Startwert $x_0 > 0$ gegen die Nullstelle $x^* > 0$ dieser Funktion.
- $\boxed{\text{f}}$ Eine Dämpfungsstrategie beim Newton-Verfahren zur Bestimmung einer Nullstelle kann man nur bei skalarwertigen Funktionen $f : \mathbb{R} \to \mathbb{R}$ anwenden.

Frage 1: $\Phi'(x)$ hat nur $x = 0$ als Nullstelle $\Rightarrow \Phi'(x^*) \neq 0$
$\Rightarrow$ Konvergenzordnung maximal 1. Es folgt genau 1

Frage 2: Muss nicht sein. Man verliert den Einschluss ja auch (vorübergehend) während der Iteration

Frage 3: Siehe Beispiel 5.24.

$$\begin{aligned}
0 < x_0 < \sqrt{3} \quad &\rightarrow \quad \sqrt{3} < x_1 \\
x_i = \sqrt{3} \quad &\rightarrow \quad x_{i+1} = \sqrt{3}, \quad i = 0, 1, 2, \ldots \\
\sqrt{3} < x_i \quad &\rightarrow \quad \sqrt{3} < x_{i+i} < x_i, \quad i = 0, 1, 2, \ldots
\end{aligned}$$

Frage 4: Geht mehrdimensional genauso.

Verständnisfragen VF-5

$\boxed{\text{f}}$ Es seien $f : \mathbb{R} \to \mathbb{R}$ zweimal stetig differenzierbar, und $f(x^*) = 0$, $f'(x^*) \neq 0$.
Weiter sei x_0 so gewählt, dass die Newton-Methode

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$$

mit Startwert x_0 gegen x^* konvergiert.
Dann gilt: $|x^* - x_k| \approx (x_{k+1} - x_k)^2$ für k hinreichend groß.

Es seien $n = 1$ und $\Phi(x) = e^{-\frac{1}{2}x}$.
Wir betrachten das Fixpunktproblem auf dem Intervall $[0, 1]$.
Geben Sie eine scharfe obere Schranke für die Lipschitzkonstante $L < 1$ aus dem Banachschen Fixpunktsatz an. $\boxed{0.5}$

Verständnisfragen VF-5

f Es seien $f : \mathbb{R} \to \mathbb{R}$ zweimal stetig differenzierbar, und $f(x^*) = 0$, $f'(x^*) \neq 0$.
Weiter sei x_0 so gewählt, dass die Newton-Methode

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$$

mit Startwert x_0 gegen x^* konvergiert.
Dann gilt: $|x^* - x_k| \approx (x_{k+1} - x_k)^2$ für k hinreichend groß.

Es seien $n = 1$ und $\Phi(x) = e^{-\frac{1}{2}x}$.
Wir betrachten das Fixpunktproblem auf dem Intervall $[0, 1]$.
Geben Sie eine scharfe obere Schranke für die Lipschitzkonstante $L < 1$ aus dem Banachschen Fixpunktsatz an. 0.5

Frage 1: Richtig wäre

$$|x^* - x_k| \approx |x_{k+1} - x_k|$$

oder

$$|x_{k+1} - x^*| \leq c\,|x_k - x^*|^2.$$

Frage 2: $[0, 1]$ wird durch $e^{-\frac{1}{2}x}$ auf $[e^{-\frac{1}{2}}, 1]$ abgebildet.

$$\Phi'(x) = -\frac{1}{2}e^{-\frac{1}{2}x} \to |\Phi'(x)| \leq \frac{1}{2} = 0.5 \text{ auf } [0, 1]$$

Verständnisfragen VF-6

Es sei $P(f \mid x_0, \ldots, x_n)$ das Lagrange–Interpolationspolynom zu den Daten $(x_0, f(x_0)), \ldots, (x_n, f(x_n))$ mit $a = x_0 < \cdot < x_n = b$. Weiter sei $[x_0, \ldots, x_n]f$ die dividierte Differenz der Ordnung n von f.

- **f** Es gilt für alle $x \in \mathbb{R}$ $P(f \mid x_0, \ldots, x_n)(x) = (x - x_n)[x_0, \ldots, x_n]f + P(f \mid x_0, \ldots, x_{n-1})(x)$.
- **w** Es sei Π_n der Raum der reellen Polynome vom Grad maximal n. Die Knotenpolynome $\omega_k(x) := (x - x_0) \cdots (x - x_{k-1})$, $k = 1, \ldots, n$, $\omega_0(x) := 1$, bilden eine Basis des Raumes Π_n.
- **w** Die Auswertung des Interpolationspolynoms in der monomialen Basis $P(f \mid x_0, \ldots, x_n)(x) = \sum_{k=0}^{n} a_k\, x^k$ ist für numerische Zwecke ungünstig, weil das Problem (Auswertung) bezüglich der Koeffizienten a_k oft schlecht konditioniert ist.
- **w** Der Fehler $\max_{x\in[a,b]} |P(f \mid x_0, \ldots, x_n) - f(x)|$ hängt von der Wahl der Stützstellen ab.

Es sei $f(x) = 3\,x^2 + 2$. $[x_0, x_1, x_2, x_3]f$ ist gleich **0**

Verständnisfragen VF-6

Es sei $P(f \mid x_0, \ldots, x_n)$ das Lagrange–Interpolationspolynom zu den Daten $(x_0, f(x_0)), \ldots, (x_n, f(x_n))$ mit $a = x_0 < \cdot < x_n = b$. Weiter sei $[x_0, \ldots, x_n]f$ die dividierte Differenz der Ordnung n von f.

- $\boxed{f}$ Es gilt für alle $x \in \mathbb{R}$ $P(f \mid x_0, \ldots, x_n)(x) = (x - x_n)[x_0, \ldots, x_n]f + P(f \mid x_0, \ldots, x_{n-1})(x)$.
- $\boxed{w}$ Es sei Π_n der Raum der reellen Polynome vom Grad maximal n. Die Knotenpolynome $\omega_k(x) := (x - x_0) \cdots (x - x_{k-1})$, $k = 1, \ldots, n$, $\omega_0(x) := 1$, bilden eine Basis des Raumes Π_n.
- $\boxed{w}$ Die Auswertung des Interpolationspolynoms in der monomialen Basis $P(f \mid x_0, \ldots, x_n)(x) = \sum_{k=0}^{n} a_k x^k$ ist für numerische Zwecke ungünstig, weil das Problem (Auswertung) bezüglich der Koeffizienten a_k oft schlecht konditioniert ist.
- $\boxed{w}$ Der Fehler $\max_{x \in [a,b]} |P(f \mid x_0, \ldots, x_n) - f(x)|$ hängt von der Wahl der Stützstellen ab.

Es sei $f(x) = 3x^2 + 2$. $[x_0, x_1, x_2, x_3]f$ ist gleich $\boxed{0}$

Frage 1: a) Polynomgrad nur $n - 1$

b) Richtig wäre es wenn $(x - x_n)$ durch $\prod_{i=0}^{n-i}(x - x_i)$ ersetzt wird.

Frage 2: Das ist die Newtonsche Basis

Frage 3: Beispiel Kondition bzgl. a_0.

$$\frac{\partial P(x)}{\partial a_0} = 1 \Rightarrow \kappa_{rel,a_0} = \left| \frac{a_0}{P(x)} \right|$$

a_0 ist groß, $|P(x)|$ klein, daher ist κ_{rel,a_0} groß.

Frage 4: Wir hatten z. B. das Beispiel von Runge und

$$P_n(x) - f(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!} \prod_{i=0}^{n} (x - x_i). \text{ Abhg. von } x_i.$$

Frage 5: $[x_0, x_1, x_2, x_3]f$ ist der führende Koeffizient von $P(f \mid x_0, \ldots, x_3)$, d.h. der Koeffizient von x^4, also 0.

Verständnisfragen VF-6

Es sei $f \in C[a,b]$. Das Integral $I(f) = \int_a^b f(x)\,dx$ soll numerisch approximiert werden durch eine Newton-Cotes-Formel $I_m(f) = (b-a)\sum_{j=0}^m w_j\, f(x_j)$ mit $a \le x_0 < \ldots < x_m \le b$. Weiter sei $I_{m,n}(f)$ die aus $I_m(f)$ konstruierte summierte Quadraturformel auf den Teilintervallen $[t_{j-1}, t_j]$, $j = 1, \ldots, n$, mit $t_j = a + j\,h$, $j = 0, 1, \ldots, n$, $h = (b-a)/n$.

- **w** Es sei $I_2(f)$ die Simpson-Regel. Dann gilt $|I_{2,n}(f) - I(f)| \to 0$ für $n \to \infty$.
- **w** Es gilt $I_m(p) = I(p)$ für alle Polynome p vom Grad maximal m.
- **f** Es gilt $I_{1,n}(p) = I(p)$ für alle Polynome p vom Grad maximal n.
- **f** Bei den Newton-Cotes-Formeln hängen die Gewichte w_j von dem Interval $[a,b]$ ab.

Berechnen Sie eine Approximation von $\int_0^2 x^5\,dx$ mit Hilfe der summierten Trapezregel $I_{1,2}(f)$. **17**

Verständnisfragen VF-6

Es sei $f \in C[a,b]$. Das Integral $I(f) = \int_a^b f(x)\,dx$ soll numerisch approximiert werden durch eine Newton-Cotes-Formel $I_m(f) = (b-a)\sum_{j=0}^m w_j\, f(x_j)$ mit $a \le x_0 < \ldots < x_m \le b$. Weiter sei $I_{m,n}(f)$ die aus $I_m(f)$ konstruierte summierte Quadraturformel auf den Teilintervallen $[t_{j-1}, t_j]$, $j = 1, \ldots, n$, mit $t_j = a + j\,h$, $j = 0, 1, \ldots, n$, $h = (b-a)/n$.

- $\boxed{\text{w}}$ Es sei $I_2(f)$ die Simpson-Regel. Dann gilt $|I_{2,n}(f) - I(f)| \to 0$ für $n \to \infty$.
- $\boxed{\text{w}}$ Es gilt $I_m(p) = I(p)$ für alle Polynome p vom Grad maximal m.
- $\boxed{\text{f}}$ Es gilt $I_{1,n}(p) = I(p)$ für alle Polynome p vom Grad maximal n.
- $\boxed{\text{f}}$ Bei den Newton-Cotes-Formeln hängen die Gewichte w_j von dem Interval $[a,b]$ ab.

Berechnen Sie eine Approximation von $\int_0^2 x^5\,dx$ mit Hilfe der summierten Trapezregel $I_{1,2}(f)$. $\boxed{17}$

Frage 1: Stetigkeit reicht bereits für die Konvergenz der summierten Regeln. Mit $f \in C^4[a,b]$ hätten wir hier sogar $\mathcal{O}(h^4)$, $h = (b-a)/n$

Frage 2: Der Exaktheitsgrad von Newton-Cotes-Formeln ist m oder $m+1$, also ist m immer OK.

Frage 3: Summation ändert **nicht** den Exaktheitsgrad

Frage 4: Weil wir die Darstellung $h\sum_i w_i f(x_i)$ wählen, sind die Gewichte nicht intervallabhängig.

Frage 5: $\frac{1}{2}(0^5 + 2\cdot 1^5 + 2^5) = \frac{34}{2} = 17$