Numerische Mathematik
WS 2016/2017

Priv. Doz. Dr. Maria Charina

Kapitel 0: Einleitung

Lineare Algebra, Analysis,... beschäftigen sich mit der

(*I*) (eindeutigen) Lösbarkeit eines mathematischen Problems.

Numerische Mathematik studiert

(*II*) Konditionierung dieses Problems,

(*III*) Numerische Algorithmen zur approximativen Lösung dieses Problems, Stabilität von Algorithmen,

(*IV*) Effizienz von Algorithmen.

Beispiel zu (*I*) und (*III*)

Fundamental-Satz der linearen Algebra: Jedes Polynom $n-$ten Grades hat genau n komplexe Nullstellen.

Gegeben: $p_1(x) = x^3 + x^2 - 2$

Lösungsmethode: Raten einer der Nullstellen

$$p_1(x) = x^3 + x^2 - 2 = (x-1)(x+1+i)(x+1-i)$$

Gegeben: $p_2(x) = x^3 + x^2 - 2.1$

$$\text{Raten} \quad \Longrightarrow \quad p_2(x) = x^3 + x^2 - 2.1 = ????$$

Lösungsmethode: numerische Algorithmen

Beispiel zu (*II*)

Problem A: löse

$$\begin{pmatrix} 1.2969 & 0.8648 \\ 0.2161 & 0.1441 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0.8642 \\ 0.1440 \end{pmatrix}$$

hat die exakte Lösung $(x_1, x_2)^T = (2, -2)^T$.

Problem $\tilde{A}$: löse

$$\begin{pmatrix} 1.2969 & 0.8648 \\ 0.2161 & 0.1441 \end{pmatrix} \begin{pmatrix} \widetilde{x}_1 \\ \widetilde{x}_2 \end{pmatrix} = \begin{pmatrix} 0.8642 - 0.00000001 \\ 0.1440 + 0.00000001 \end{pmatrix}$$

hat die exakte Lösung $(\widetilde{x}_1, \widetilde{x}_2)^T = (0.9911, -0.4870)^T$.

$\Rightarrow$ Das Problem A ist schlecht konditioniert.

Beispiel zu *II-III*

Problem: Sei $f : \mathbb{R} \to \mathbb{R}$, $f(x) = cos(x)$. Berechne $f'(1)$.

Exakte Lösung ist $f'(1) = -0.8414709...$

MATLAB-Programm (Algorithmus):

```
h=1;
for i=1:55
    res=(cos(1+h)-cos(1))/h;
    h=h/2;
end
```

Näherungslösung: für i=55, ergibt sich $f'(1) \approx \text{res} = 0$

Ist dieses Problem schlecht konditioniert, oder ist der Algorithmus instabil?

Beispiel zu (IV)

Gegeben: invertierbare $A \in \mathbb{R}^{n \times n}$ und $b = (b_1, \ldots, b_n)^T \in \mathbb{R}^n$.

Gesucht: $x = (x_1, \ldots, x_n)^T \in \mathbb{R}^n$ mit $Ax = b$.

For $j = 1, \ldots, n$, berechne

$$x_j = \frac{\det \begin{bmatrix} a_{1,1} & \ldots & a_{1,j-1} & b_1 & a_{1,j+1} & \ldots & a_{1,n} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{n,1} & \ldots & a_{n,j-1} & b_n & a_{n,j+1} & \ldots & a_{n,n} \end{bmatrix}}{\det(A)}$$

Dieser Algorithmus benötigt $2^n \cdot n!$ Rechenoperationen, ist nicht effizient.

Kapitel 1: Funktionsauswertungen

Problem 1: Sei $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^m$. Berechne $f(x)$, $x \in \Omega$.

1.1 Relative Konditionszahlen

Satz+Definition: Sei

$$f = \begin{pmatrix} f_1 \\ \vdots \\ f_m \end{pmatrix} : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^m, \quad \Omega \text{ offen und konvex},$$

zweimal stetig differenzierbar. Dann gilt, für $x = (x_1, \ldots, x_n) \in \Omega$, $x_j \neq 0$, und $f_i(x) \neq 0$, $i = 1, \ldots, m$,

$$\left| \frac{f_i(x) - f_i(\tilde{x})}{f_i(x)} \right| \leq \sum_{j=1}^{n} k_{i,j}(x) \left| \frac{x_j - \tilde{x}_j}{x_j} \right| + o\left(\left| \frac{x_j - \tilde{x}_j}{x_j} \right| \right), \quad |x_j - \tilde{x}_j| \to 0,$$

mit den relativen Konditionszahlen

$$k_{i,j}(x) = \left| \frac{\partial f_i(x)}{\partial x_j} \cdot \frac{x_j}{f_i(x)} \right|.$$

Bezeichnung: Man nennt das Problem 1 schlecht konditioniert, wenn ein $|k_{i,j}(x)| >> 1$ ist; andernfalls gut konditioniert. Im Fall $|k_{i,j}(x)| < 1$ spricht man von Fehlerdämpfung und im Fall $|k_{i,j}(x)| > 1$ von Fehlerverstärkung.

1.2 Konditionierung arithmetischer Grundoperationen

- Die Addition $y = f(x_1, x_2) = x_1 + x_2$, $x_1, x_2 \in \mathbb{R} \setminus \{0\}$, mit

$$\begin{aligned} k_1 &= \frac{\partial f}{\partial x_1}(x) \cdot \frac{x_1}{f(x)} = 1 \cdot \frac{x_1}{x_1 + x_2} \\ k_2 &= \frac{\partial f}{\partial x_2}(x) \cdot \frac{x_2}{f(x)} = 1 \cdot \frac{x_2}{x_1 + x_2} \end{aligned}$$

  ist schlecht konditioniert für $x_1 \approx -x_2$.

- Die Multiplikation $y = f(x_1, x_2) = x_1 \cdot x_2$ mit

$$\begin{aligned} k_1 &= \frac{\partial f}{\partial x_1}(x) \cdot \frac{x_1}{f(x)} = x_2 \cdot \frac{x_1}{x_1 \cdot x_2} = 1 \\ k_2 &= \frac{\partial f}{\partial x_2}(x) \cdot \frac{x_2}{f(x)} = x_1 \cdot \frac{x_2}{x_1 \cdot x_2} = 1 \end{aligned}$$

  ist generell gut konditioniert.

Beispiel: Seien $x_1, x_2 \in \mathbb{R} \setminus \{0\}$. Das Problem der Berechnung von $f(x_1, x_2) = x_1^2 - x_2^2$ mit

$$\begin{aligned} k_1 &= \frac{\partial f}{\partial x_1}(x) \cdot \frac{x_1}{f(x)} = 2x_1 \cdot \frac{x_1}{x_1^2 - x_2^2} = \frac{2}{1 - \left(\frac{x_2}{x_1}\right)^2} \\ k_2 &= \frac{\partial f}{\partial x_2}(x) \cdot \frac{x_2}{f(x)} = 2x_2 \cdot \frac{x_2}{x_1 \cdot x_2} = \frac{2}{1 - \left(\frac{x_1}{x_2}\right)^2} \end{aligned}$$

ist schlecht konditioniert für $x_1 \approx \pm x_2$.

1.3 Rundungsfehler und Gleitkommaarithmetik

Definition: Die Menge $fl = fl(B, m, s)$, $B \in \mathbb{N}$, $B \geq 2$, $m, s \in \mathbb{N}$, besteht aus Gleitkommazahlen der Form

$$\pm \left(\sum_{j=1}^{m} d_j \cdot B^{-j} \right) \times B^E, \qquad E = \pm \sum_{k=0}^{s-1} e_k \cdot B^k,$$

mit $d_j, e_k \in \{0, \ldots, B-1\}$ und $d_1 \neq 0$.

Bemerkung: Die Menge fl ist endlich und

$$fl \subset D(fl) := [x_{min}, x_{negmax}] \cup \{0\} \cup [x_{posmin}, x_{max}]$$

mit $x_{min,max} = \mp(1 - B^{-m})B^{B^s - 1}$ und $x_{neqmax,posmin} = \mp B^{-B^s}$.

Definition: Seien $B \in \mathbb{N}$ gerade, $m, s \in \mathbb{N}$. Der Rundungsoperator

$$\mathrm{rd} : D(fl) \setminus \{0\} \to fl(B, m, s)$$

für

$$x = \pm \left(\sum_{j=1}^{\infty} d_j \cdot B^{-j} \right) \times B^E \in D(fl) \setminus \{0\}$$

ist definiert durch

$$\mathrm{rd}(x) = \begin{cases} \pm \left(\sum_{j=1}^{m} d_j \cdot B^{-j} \right) \times B^E & \text{, falls} \quad d_{m+1} < \frac{B}{2}, \\ \pm \left(\sum_{j=1}^{m} d_j \cdot B^{-j} + B^{-m} \right) \times B^E & \text{, sonst.} \end{cases}$$

Satz: Seien $B \in \mathbb{N}$ gerade, $m, s \in \mathbb{N}$. Dann gilt für $x \in D(fl) \setminus \{0\}$

$$\left| \frac{\mathrm{rd}(x) - x}{x} \right| \leq \frac{1}{2} B^{-m+1} := \mathrm{eps}.$$

Die Zahl eps wird die Machinengenauigkeit genannt.

Korollar: Es gilt $\mathrm{rd}(x) = x(1 + \delta_x) \quad \text{mit} \quad |\delta_x| \leq \mathrm{eps}$.

Definition: Bei dem Standardmodell der Gleitkommaarithmetik sind die Gleitkommaoperationen

$$\{\oplus, \ominus, \otimes, \oslash\} : fl(B, m, s) \times fl(B, m, s) \to fl(B, m, s)$$

definiert durch

$$x \oplus y := \mathrm{rd}(x + y) = (x + y)(1 + \delta), \quad |\delta| \leq \mathrm{eps}, \quad u.s.w.$$

Die Gleitkommaoperationen sind weder associativ noch distributiv.

1.4 Stabilität von Algorithmen

Definition: Ein Algorithmus zur Auswertung der Funktion

$$f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^m$$

ist eine Zerlegung

$$f = \phi^{(r)} \circ \phi^{(r-1)} \ldots \circ \phi^{(0)}, \quad r \in \mathbb{N},$$

von f in elementare Grundoperationen oder Standardfunktionen $\phi^{(i)}$.

Definition+Satz: Ein Algoritmus heißt (numerisch) stabil, wenn der im Verlaufe der Ausführung des Algorithmus akkumulierte Rundungsfehler den durch die Konditionierung des Problems bedingten unvermeidbaren Problemfehler aus 1.1 nicht übersteigt.

Der Algorithmus heißt stabil, falls alle $\phi^{(i)}$ gut konditioniert sind.