

I GRUNDLEGENDE KONZEPTE

S. BARTELS, 23.10.2013

I.A. **Aufgabenstellung.** Die Numerik beschäftigt sich mit der praktischen Berechnung mathematischer Objekte wie beispielsweise

$$\int_0^1 e^{-x^2} dx, \quad \sin(20), \quad \min_{x \in [0,1]} F(x), \quad f(x) = 0, \quad Ax = b, \quad Ax = \lambda x.$$

Abstrakt lässt sich dies als Auswertung einer Abbildung formulieren.

Definition I.1. Eine mathematische Aufgabe besteht in der Auswertung einer Funktion $\phi : X \rightarrow Y$ bei $x \in X$.

Dabei ist zum Beispiel $\phi(x) = A^{-1}x$ oder $\phi(x) = \sin(x)$. Viele der oben aufgeführten Objekte sind nicht durch geschlossene Formeln definiert und können eventuell nur *approximativ* angegeben werden. Zudem stehen auf Computern nur endlich viele sogenannte *Maschinenzahlen* zur Verfügung, so dass nicht jede reelle Zahl exakt eingegeben und elementare Rechenoperationen wie $1/3$ nur näherungsweise bestimmt werden können. Dies führt auf *Rundungsfehler*. Weitere Fehlerquellen sind *Modellfehler*, die bei der vereinfachten mathematischen Beschreibung eines realen Vorgangs auftreten, sowie *Datenfehler*, die durch Messungen verursacht sein können. Viele dieser Ungenauigkeiten sind unvermeidbar und daher ist es in der Regel weder notwendig noch sinnvoll, ein mathematisches Problem exakt zu lösen. Durch *approximatives Lösen* lässt sich der *Rechenaufwand* häufig erheblich verringern. Die Berechnung der Determinante einer Matrix $A \in \mathbb{R}^{n \times n}$ mit dem Laplaceschen Entwicklungssatz führt beispielsweise auf $n!$ Rechenoperationen, was für große Dimensionen n kaum in vertretbarer Zeit zu realisieren ist. Häufig lässt sich jedoch mit polynomialem Aufwand zumindest approximativ eine Faktorisierung $A \approx LR$ mit Dreiecksmatrizen $L, R \in \mathbb{R}^{n \times n}$ konstruieren, mit deren Hilfe sich die Determinante $\det A \approx \det L \det R$ mit einem Aufwand bestimmen lässt, der mit n vergleichbar ist. Das Lösen linearer Gleichungssysteme $Ax = b$ ist damit eng verbunden. In der Praxis wird in der Regel nicht die inverse Matrix A^{-1} explizit bestimmt, sondern das Gleichungssystem direkt gelöst. Der Ausdruck $x = A^{-1}b$ steht daher in der Numerik für die Lösung des Gleichungssystems $Ax = b$ und nicht für die Multiplikation von b mit A^{-1} . In der Vorlesung werden die folgenden für die Numerik typischen Fragestellungen diskutiert:

- Berechenbarkeit von Problemen (Algorithmik)
- Einfluss von Störungen (Stabilität)
- Fehler zwischen berechneter und exakter Lösung (Konvergenz)

- Aufwand von Verfahren (Komplexität)

Ein wichtiges Ziel ist es, einen guten Kompromiss zwischen Genauigkeit und Aufwand eines Verfahrens zu erreichen. Dies wird für folgende Probleme untersucht:

- Lineare Gleichungssysteme
- Eigenwertprobleme
- Interpolation von Funktionen
- Integration von Funktionen
- Nullstellensuche und Optimierung

I.B. Kondition und Stabilität. Wir betrachten ein Beispiel, das die Auswirkungen von Störungen auf die Lösung eines Problems illustriert.

Beispiel I.2. Für jedes $\varepsilon \in \mathbb{R} \setminus \{0\}$ ist die eindeutige Lösung des Gleichungssystems

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 + \varepsilon \end{bmatrix} x = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

gegeben durch $x = [2, 0]^\top$. Wir nehmen an, dass ε sehr klein ist und stören die rechte Seite in der zweiten Komponente, das heißt wir betrachten

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 + \varepsilon \end{bmatrix} \tilde{x} = \begin{bmatrix} 2 \\ 2 + \varepsilon \end{bmatrix}.$$

Die eindeutige Lösung ist gegeben durch $\tilde{x} = [1, 1]^\top$. Obwohl die Störung beliebig klein ist, unterscheiden sich die Lösungen x und \tilde{x} sehr stark.

Die Auswirkungen von Störungen auf die Lösung eines Problems führen auf den Begriff der Konditionierung.

Definition I.3. Eine mathematische Aufgabe $\phi(x)$ heißt schlecht konditioniert (an der Stelle x), wenn kleine Störungen der Daten große relative Fehler in der Lösung verursachen, das heißt, wenn eine Störung \tilde{x} existiert mit

$$\frac{|\phi(\tilde{x}) - \phi(x)|}{|\phi(x)|} \gg \frac{|\tilde{x} - x|}{|x|},$$

wobei $x \neq 0$ und $\phi(x) \neq 0$ gelte. Andernfalls heißt die Aufgabe gut konditioniert.

Die Relation $a \gg b$ bedeutet, dass a wesentlich größer ist als b , zum Beispiel $a \geq 100b$. Was als wesentlich größer gilt, ist im Allgemeinen jedoch problemabhängig. Die Multiplikation zweier Zahlen ist gut konditioniert.

Satz I.4 (Konditionierung der Multiplikation). Die Aufgabe $\phi(x, y) = xy$ ist gut konditioniert, in dem Sinne, dass für $x, y \in \mathbb{R}$ mit $x, y \neq 0$ und folglich $\phi(x, y) \neq 0$ sowie Störungen $\tilde{x}, \tilde{y} \in \mathbb{R}$ die relativen Fehler

$$\varepsilon_\phi = \frac{|\phi(\tilde{x}, \tilde{y}) - \phi(x, y)|}{|\phi(x, y)|}, \quad \varepsilon_x = \frac{|\tilde{x} - x|}{|x|}, \quad \varepsilon_y = \frac{|\tilde{y} - y|}{|y|}$$

die Abschätzung

$$\varepsilon_\phi \leq \varepsilon_x + \varepsilon_y + \varepsilon_x \varepsilon_y$$

erfüllen. Sind ε_x und ε_y klein, ist demnach der relative Fehler ε_ϕ ebenfalls klein.

Beweis. Es gilt

$$\varepsilon_\phi = \frac{|\tilde{x}\tilde{y} - xy|}{|xy|} = \frac{|(\tilde{x} - x)\tilde{y} + x(\tilde{y} - y)|}{|xy|} \leq \frac{|\tilde{x} - x|}{|x|} \frac{|\tilde{y} - y + y|}{|y|} + \frac{|\tilde{y} - y|}{|y|}$$

und dies impliziert die Behauptung. \square

Bemerkung I.5. Ebenfalls gut konditioniert sind die Addition zweier positiver oder zweier negativer Zahlen und die Inversion von Null verschiedener Zahlen. Schlecht konditioniert ist hingegen die Subtraktion nahezu gleich großer Zahlen, wie unten gezeigt wird.

Die gute Konditionierung einer Aufgabe ist offensichtlich notwendig, um das Problem numerisch sinnvoll lösen zu können, da Rundungsfehler andernfalls große Fehler verursachen könnten.

Definition I.6. Ein Verfahren oder Algorithmus zur (näherungsweise) Lösung einer Aufgabe ϕ ist eine Abbildung $\tilde{\phi} : X \rightarrow Y$, die durch die Hintereinanderausführung elementarer, möglicherweise rundungsfehlerbehafteter Rechenoperationen definiert ist, im einfachsten Fall

$$\tilde{\phi} = f_J \circ f_{J-1} \circ \cdots \circ f_1.$$

Beispiel I.7. (i) Die Aufgabe $\phi(x) = x^4$ lässt sich realisieren durch $\tilde{\phi} = f \circ f$, mit der vom Rechner bereitgestellten Multiplikation $f(x) = x \boxtimes x$.
(ii) Die Wurzel $\phi(x) = \sqrt{x}$ einer Zahl $x > 0$ ist nach Heron gegeben als Grenzwert jeder Folge $z_{n+1} = (z_n + x/z_n)/2$ mit $z_0 > 0$. Damit kann $\tilde{\phi}$ als J -malige Anwendung der Iterationsvorschrift definiert werden.

Für eine Aufgabe sind in der Regel verschiedene Verfahren denkbar, aber selbst wenn die Aufgabe gut konditioniert ist, führen nicht alle Verfahren auf gute Ergebnisse, da sich Rundungsfehler im Laufe der Ausführung eines Verfahrens unterschiedlich auswirken können.

Beispiel I.8. Die durch die Funktion

$$\phi(x) = \frac{1}{x} - \frac{1}{x+1} = \frac{1}{x(x+1)}$$

definierte Aufgabe ist für große Zahlen $x \in \mathbb{R}$ gut konditioniert, denn für eine Störung $\tilde{x} = (1 + \varepsilon_x)x$ mit einer kleinen Zahl ε_x erhalten wir

$$\phi(x) - \phi(\tilde{x}) = \frac{(1 + \varepsilon_x)x((1 + \varepsilon_x)x + 1) - x(x + 1)}{(1 + \varepsilon_x)x((1 + \varepsilon_x)x + 1)x(x + 1)} \approx \frac{2\varepsilon_x x^2}{x^4}.$$

Damit folgt für den relativen Fehler $\varepsilon_\phi \leq 2\varepsilon_x$, sofern $x \geq 1$ gilt. Die numerische Realisierung kann über die Verfahren

$$\tilde{\phi}_1(x) = \left(\frac{1}{x}\right) - \left(\frac{1}{x+1}\right), \quad \tilde{\phi}_2(x) = \frac{1}{(x(x+1))}$$

erfolgen, wobei die Klammerung die Reihenfolge der Ausführung der Operationen festlegt. Numerische Experimente zeigen, dass $\tilde{\phi}_1$ und $\tilde{\phi}_2$ für große Zahlen x stark voneinander abweichen.

Definition I.9. Ein Algorithmus $\tilde{\phi}$ heißt instabil, wenn es eine Störung \tilde{x} von x gibt, so dass der durch Rundungsfehler und Störungen verursachte relative Fehler erheblich größer ist als der nur durch die Störung verursachte Fehler, das heißt, falls $\phi(x) \neq 0$ und

$$\frac{|\tilde{\phi}(\tilde{x}) - \phi(x)|}{|\phi(x)|} \gg \frac{|\phi(\tilde{x}) - \phi(x)|}{|\phi(x)|}.$$

Ein Algorithmus heißt stabil, wenn er nicht instabil ist.

Bemerkung I.10. Notwendig für die Stabilität eines Algorithmus ist, dass jeder einzelne Rechenschritt eine gut konditionierte Aufgabe ist.

Der obige Algorithmus $\tilde{\phi}_1$ ist instabil aufgrund sogenannter *Auslöschungseffekte*, die bei der Subtraktion nahezu gleich großer Zahlen auftreten.

Beispiel I.11. Für $x = 0.677354$ und $y = 0.677335$ ist $\phi(x, y) = x - y = 0.000019 = 0.19 \cdot 10^{-4}$. Für die Störung $\tilde{x} = (1 + \varepsilon_x)x$ mit $\varepsilon_x = 1.0 \cdot 10^{-4}$ folgt

$$\varepsilon_\phi = \frac{|\phi(\tilde{x}, y) - \phi(x, y)|}{|\phi(x, y)|} = \frac{\varepsilon_x x}{x - y} = \frac{0.677354 \cdot 10^{-4}}{0.19 \cdot 10^{-4}} = 3.565021$$

Die Störung von 0.01% bewirkt also einen relativen Fehler von über 350%.

Die Subtraktion nahezu gleich großer Zahlen ist eine schlecht konditionierte Aufgabe.

Bemerkung I.12. Der durch Rundung und näherungsweise Lösen verursachte Fehler bei der Lösung einer Aufgabe lässt sich mittels der Konditionierung der Aufgabe und der Stabilität des Verfahrens abschätzen, denn es gilt

$$|\phi(x) - \tilde{\phi}(\tilde{x})| \leq |\phi(x) - \phi(\tilde{x})| + |\phi(\tilde{x}) - \tilde{\phi}(\tilde{x})|.$$

I.C. Aufwand. Neben der Stabilität eines numerischen Verfahrens ist der Rechenaufwand eine wichtige Größe.

Definition I.13. Für eine Aufgabe $\phi: \mathbb{R}^n \rightarrow \mathbb{R}^m$ und ein zugehöriges Verfahren $\tilde{\phi}: \mathbb{R}^n \rightarrow \mathbb{R}^m$ ist der Aufwand die Anzahl der benötigten Rechenoperationen von $\tilde{\phi}$.

Eine exakte Bestimmung des Aufwands ist in der Regel nicht notwendig und es wird die Abhängigkeit von der Problemgröße n untersucht. Dabei ist die sogenannte Landau-Notation hilfreich.

Definition I.14. Die Folge $(a_n)_{n \in \mathbb{N}}$ ist von der Ordnung der Folge $(b_n)_{n \in \mathbb{N}}$, falls Zahlen $C > 0$ und $N \in \mathbb{N}$ existieren, so dass $|a_n| \leq C|b_n|$ für alle $n \geq N$ gilt. In diesem Fall schreiben wir $a_n = \mathcal{O}(b_n)$.

Für den Aufwand a_n eines Verfahrens ist interessant, ob dieser von einer polynomialen Ordnung n^p ist.

Beispiel I.15. (i) Die Multiplikation eines Vektors $x \in \mathbb{R}^n$ mit einer fixierten Zahl $a \in \mathbb{R}$ ist von der Ordnung $\mathcal{O}(n)$.

(ii) Das Gaußsche Verfahren zur Lösung eines linearen Gleichungssystems besitzt den Aufwand $\mathcal{O}(n^3)$, während die Cramersche Regel auf einen Aufwand der Ordnung $\mathcal{O}(n!)$ führt.