

VO Numerische Mathematik
2017/18
Theoriefragen

https://github.com/Arkonos/NM-Theorie_1718

14. Februar 2018

1 Zahldarstellung, Rundung und Fehler

1.1. Wie werden ganze Zahlen binär abgespeichert?

S. 2

$$b_{N-1}b_{N-2}\dots b_1b_0 \cong b = \sum_{j=0}^{N-1} b_j 2^j, \quad b_j \in \{0, 1\}$$

Beispiel: 23_{10}

$$\begin{aligned} 10101_2 &= 1 \cdot 2^4 + 0 \cdot 2^3 + 0 \cdot 2^2 + 1 \cdot 2^1 + 1 \cdot 2^0 \\ &= 16 + 0 + 0 + 2 + 1 = 19_{10} \end{aligned}$$

$$:2 \quad \begin{array}{r} 19 \quad 9 \quad 4 \quad 2 \quad 1 \\ 1 \quad 1 \quad 0 \quad 0 \quad 1 \end{array} \rightarrow 10011_2$$

1.2. Wie werden Gleitpunktzahlen (doppelte Genauigkeit) binär abgespeichert?

S. 3

$$\begin{aligned} x &= (-1)^s \cdot m \cdot 2^e \\ x &\cong s \quad e_{11}e_{10}\dots e_0 \quad (m_0)m_1m_2\dots m_{51} \end{aligned}$$

s ... Vorzeichenbit $\in \{0, 1\}$
m ... Mantisse Normiert, $m_0 \stackrel{!}{=} 1$ wird weggelassen
e ... Exponent nach Abzug von b...Bias = 1023 (double)

	s	m	l
single 32 Bit	1	23	8
double 64 Bit	1	52	11

1.3. Wie werden Gleitpunktzahlen gerundet?

S. 6

Round to the nearest even.

...	m_M	m_{M+1}	m_{M+2}	m_{M+3}	...
	x	0	x	x	abrunden
	x	1	1	0	aufrunden
	x	1	0	1	aufrunden
	1	1	0	0	aufrunden
	0	1	0	0	abrunden

1.4. Wie ist der relative Rundungsfehler definiert?

S. 6

$$\frac{|\text{rd}(x) - x|}{|x|} \leq \frac{2^{-M-1} \cdot 2^e}{a \cdot 2^e} \stackrel{a \in [1,2)}{\leq} 2^{-M-1} =: \text{eps}$$

$\text{rd}(a)$ durch rounding to the nearest even

1.5. Wie groß ist die relative Maschinengenauigkeit eps für doppelt genaue Gleitpunktzahlen? Wie kann man eps experimentell bestimmen?

1.6. Was ist die relative/absolute Kondition eines Problems?

$$\begin{aligned} \frac{|f(\tilde{x}) - f(x)|}{|f(x)|} &\leq \kappa_{\text{rel.}} \cdot \varepsilon, \quad \kappa_{\text{rel.}} > 0 \\ |f(\tilde{x}) - f(x)| &\leq \kappa_{\text{abs.}} \cdot \delta, \quad \kappa_{\text{abs.}} > 0 \end{aligned}$$

Wenn $\kappa_{\text{rel.}}$ klein ist werden Inputfehler nicht übermäßig verstärkt und f gilt als gut konditioniert.

1.7. Was bedeuten die Begriffe Konsistenz und Konsistenzordnung?

Def 1.3 Konsistenz, Konsistenzordnung

Ein numerisches Verfahren f_h mit Diskretisierungsweite h zur Bestimmung einer Näherung von $f_h(x)$ an $f(x)$ ist konsistent, falls gilt

$$\|f_h(x) - f(x)\| \leq C h^p$$

wobei die Konstante $C > 0$ nicht von h abhängen darf und $p \geq 1$. Der Exponent p ist dann die Konsistenzordnung und es gilt $f_h \rightarrow f$ für $h \rightarrow 0$, falls f_h exakt ausgewertet wird.

1.8. Wodurch unterscheidet sich Konsistenz von Konvergenz?

$f_h \rightarrow f$ konvergiert für $h \rightarrow 0$ falls f_h exakt ausgewertet wird. Computer müssen jedoch runden, wodurch wir nur noch von Konsistenz reden können wenn das Verfahren nicht stabil ist.

1.9. Was bedeutet der Begriff Stabilität?

S. 15

Ein numerisches Verfahren f heißt stabil, falls bei der numerischen Auswertung $\tilde{f}(x)$ des Verfahrens Fehler wie Rundungsfehler, Abbruchfehler und Verfahrensfehler von Teilschritten nicht übermäßig verstärkt werden im Vergleich zu dem durch die relative Kondition κ_{rel} des Problems verursachten Fehler.

Für die Differenz zwischen numerischer Auswertung $\tilde{f}(x)$ mit obigen Fehlern und exakter Auswertung $f(x)$ gilt dann

$$\|\tilde{f}(x) - f(x)\| \leq C \cdot \kappa_{rel} \cdot \|f(x)\| \cdot \text{eps}, \quad C > 0 \text{ klein.}$$

Ein numerisches Verfahren ist insbesondere dann stabil, wenn alle seine Teilschritte gut konditioniert sind.

2 Numerische Differentiation

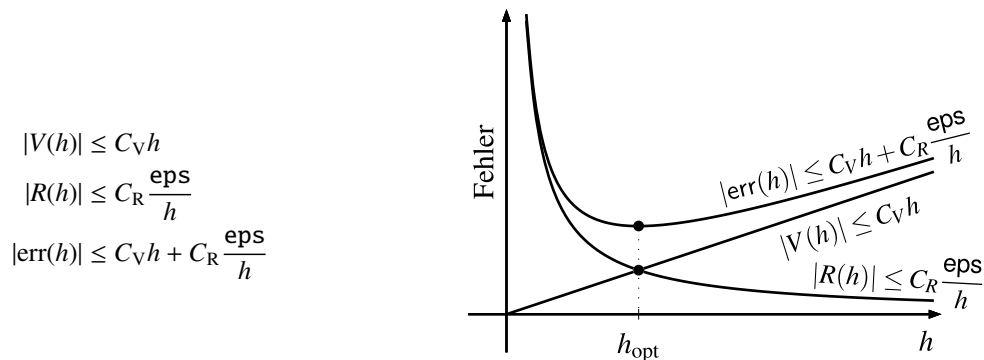
- 2.1. Wie wird mit Hilfe der Vorwärtsdifferenz eine differenzierbare Funktion f an der Stelle x differenziert? Wie groß ist h_{opt} ?

$$f'(x) = \frac{f(x+h) - f(x)}{h} - \frac{h}{2} f''(\xi), \quad h_{\text{opt}} = \sqrt{\text{eps}}$$

- 2.2. Wie wird mit Hilfe der zentralen Differenz eine differenzierbare Funktion f an der Stelle x differenziert? Wie groß ist h_{opt} ?

$$f'(x) = \frac{f(x+h) - f(x-h)}{2h} - \frac{h^2}{6} f'''(\xi), \quad h_{\text{opt}} = \sqrt[3]{\text{eps}}$$

- 2.3. Wie verhalten sich Verfahrensfehler und Rundungsfehler in Abhängigkeit von der Schrittweite h ? Machen Sie eine Skizze.



- 2.4. Wie lässt sich mit Hilfe eines logarithmischen Plots das Verhalten von Verfahrensfehler und Rundungsfehler ablesen? Wie kann man die optimale Schrittweite h_{opt} ablesen?

$$\log |\text{err}| = \log \left(C_V h + C_R \frac{\text{eps}}{h} \right) \approx \begin{cases} \log(C_R \text{eps} h^{-q}) = -q \log h + \log C_R + \log \text{eps}, & \text{links von } h_{\text{opt}}, \\ \log(C_V \text{eps} h^p) = p \log h + \log C_V, & \text{rechts von } h_{\text{opt}}. \end{cases}$$

Somit erhält man zwei Geraden der Form $y = kx + d$, wobei $k = -q$ und $k = p$ aus dem Plot abgelesen werden können.

Die optimale Schrittweite kann man im Schnittpunkt der beiden Geraden erkennen.

- 2.5. Wieso gilt bei der zentralen Differenz für den Verfahrensfehler $V(h) = O(h^2)$ statt $O(h)$?

Für die zentrale Differenz werden die Taylorpolynome für $f(x+h)$ und $f(x-h)$ gemittelt. Dabei heben sich die Terme zweiter Ordnung, $\frac{h^2}{2} f''(x)$, auf.

- 2.6. Wie wird die zweite Ableitung einer zweimal differenzierbaren Funktion an der Stelle x berechnet? Wie groß ist h_{opt} ?

$$f''(x) = \frac{f(x+h) - 2f(x) + f(x-h))}{h^2} - \frac{h^2}{12} f^{(4)}(\xi), \quad h_{\text{opt}} = \sqrt[4]{\text{eps}} = \sqrt[4]{\text{eps}}$$

- 2.7. Wie lässt sich die optimale Schrittweite h_{opt} aus dem Verfahrensfehler $V(h)$ und dem Rundungsfehler $R(h)$ bestimmen?

$$|\text{err}(h)| \leq C_V h + C_R \frac{\text{eps}}{h}$$

Der Fehler soll minimal sein und somit $C_V h + C_R \frac{\text{eps}}{h} = \min$. Dies führt zu $C_V h^2 + C_R \text{eps} = 0$ und somit

$$h_{\text{opt}} = \sqrt{\text{eps} \frac{C_V}{C_R}}.$$

Bei $f(x) \approx f'(x) \approx f''(x)$ gilt $C_V \approx C_R$ und somit

$$h_{\text{opt}} = \sqrt{\text{eps}}$$

- 2.8. Wie berechnet man die Jacobimatrix einer vektorwertigen Funktion $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ durch numerisches Differenzieren?

$$\text{Jacobimatrix: } \mathbf{J}(\mathbf{x}) = \mathbf{f}'(\mathbf{x}) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \dots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1} & \dots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$

Approximation der i-ten Spalte

$$\frac{\partial \mathbf{f}}{\partial x_i}(\mathbf{x}) = \begin{bmatrix} \frac{\partial f_1}{\partial x_i}(\mathbf{x}) \\ \vdots \\ \frac{\partial f_m}{\partial x_i}(\mathbf{x}) \end{bmatrix} \approx \frac{\mathbf{f}(x_1, \dots, x_{i-1}, x_i + h, x_{i+1}, \dots, x_n) - \mathbf{f}(\mathbf{x})}{h}$$

- 2.9. Wie berechnet man den Gradient einer skalaren Funktion $f : \mathbb{R}^n \rightarrow \mathbb{R}$ durch numerisches Differenzieren?

3 Interpolation

3.1. Wie werden die dividierten Differenzen berechnet?

Mit Hilfe eines Differenzenschema oder Differenzentableau.

Allgemeines Differenzenschema für 5 Punkte

x_i	y_i	δy_i	$\delta^2 y_i$	$\delta^3 y_i$	$\delta^4 y_i$
x_0	y_0				
x_1	y_1	$\delta y_0 = \frac{y_1 - y_0}{x_1 - x_0}$			
			$\delta^2 y_0 = \frac{\delta y_1 - \delta y_0}{x_2 - x_0}$		
x_2	y_2	$\delta y_1 = \frac{y_2 - y_1}{x_2 - x_1}$		$\delta^3 y_0 = \frac{\delta^2 y_1 - \delta^2 y_0}{x_3 - x_0}$	$\delta^4 y_0 = \frac{\delta^3 y_1 - \delta^3 y_0}{x_4 - x_0}$
			$\delta^2 y_1 = \frac{\delta y_2 - \delta y_1}{x_3 - x_1}$		
x_3	y_3	$\delta y_2 = \frac{y_3 - y_2}{x_3 - x_2}$		$\delta^3 y_1 = \frac{\delta^2 y_2 - \delta^2 y_1}{x_4 - x_1}$	
			$\delta^2 y_2 = \frac{\delta y_3 - \delta y_2}{x_4 - x_2}$		
x_4	y_4	$\delta y_3 = \frac{y_4 - y_3}{x_4 - x_3}$			

3.2. Wie ist das Newtonsche Interpolationspolynom definiert?

$$p(x) = y_0 + (x - x_0)\delta y_0 + (x - x_0)(x - x_1)\delta^2 y_0 + \cdots + (x - x_0)(x - x_1) \cdots (x - x_{n-1})\delta^n y_0$$

3.3. Wie wird mit dem Hornerschema ein Polynom $p(x) = a_0 + a_1x + \cdots + a_nx^n$ ausgewertet?

$$p(x) = a_0 + x(a_1 + x(a_2 + \cdots + x(a_{n-2} + x(a_{n-1} + x a_n)) \cdots))$$

3.4. Wie wird mit dem Hornerschema ein Newtonsches Interpolationspolynom ausgewertet?

	a_n	a_{n-1}	a_{n-2}	\cdots	a_1	a_0
x	0	$x \cdot b_n$	$x \cdot b_{n-1}$	\cdots	$x \cdot b_2$	$x \cdot b_1$
	b_n	b_{n-1}	b_{n-2}	\cdots	b_1	$b_0 = p(x)$

3.5. Wie sind die Lagrange-Polynome definiert? Welche Eigenschaften haben sie?

$$\ell_i(x) = \frac{(x - x_0)(x - x_1) \cdots \widehat{(x - x_i)} \cdots (x - x_n)}{(x_i - x_0)(x_i - x_1) \cdots \widehat{(x_i - x_i)} \cdots (x_i - x_n)}$$

$$\ell_i(x_j) = \delta_{ij} = \begin{cases} 1, & \text{falls } i = j, \\ 0, & \text{falls } i \neq j. \end{cases}$$

Wobei die überdachten Terme wegzulassen sind.

3.6. Wie wird mit Hilfe der Lagrange-Polynome das Langrangesche Interpolationspolynom berechnet?

$$p(x) = \sum_{i=0}^n y_i \ell_i(x), \quad p(x_j) = y_j$$

3.7. Erklären Sie die Begriffe Datenfehler, Verstärkungsfaktor, Lebesgue-Funktion und Lebesgue-Konstante in Zusammenhang mit der Polynominterpolation. Was ist die Kondition der Polynominterpolation?

An einer Stützstelle x_j kann ein Datenfehler ε_j auftreten. Führt man diesen in ein Lagrange-Polynom ein erhält man

$$\bar{p}(x) = \sum_{i=0}^{j-1} (y_i \ell_i(x)) + (y_j + \varepsilon_j) + \sum_{i=j+1}^n (y_i \ell_i(x)).$$

Der Fehler ε_i wird also um den Faktor $|\ell_i|$ verstärkt. Wenn alle Knoten mit Fehlern versehen sind ergibt sich

$$\bar{p}(x) = p(x) + \sum_{i=0}^n (\varepsilon_i \ell_i).$$

Falls die einzelnen Fehler durch $\varepsilon_i \leq M$ beschränkt sind erhält man für den absoluten Fehler die Abschätzung

$$\underbrace{|\bar{p}(x) - p(x)|}_{\text{abs. Fehler Output}} \leq M \underbrace{\sum_{i=0}^n |\ell_i(x)|}_{\kappa_{\text{abs}}(x)} = M \cdot \kappa_{\text{abs}}(x).$$

Wobei $\kappa_{\text{abs},i} = |\ell_i(x)|$ der Verstärkungsfaktor für den Datenfehler in der Stützstelle i und die Lebesgue-Funktion $\kappa_{\text{abs}} = \lambda_n(x) = \sum_{i=0}^n |\ell_i(x)|$ die absolute Kondition für die Polynominterpolation ist. Die schlechteste Konditionszahl $\lambda_n(x)$ im Intervall $[\min_i x_i, \max_i x_i]$ nennen wir Lebesgue-Konstante und erhalten sie durch

$$\max_{x \in [\min_i x_i, \max_i x_i]} \lambda_n(x) := \Lambda_n.$$

3.8. Was besagt der Satz über den Fehler des Interpolationspolynoms? Wie ist der Verfahrensfehler definiert?

Gegeben seien $n + 1$ verschiedene Stützstellen $x_i, i = 0, \dots, n$, in einem Intervall $[a, b]$ und eine $n + 1$ -mal stetig differenzierbare Funktion $f \in \mathcal{C}^{n+1}([a, b])$. Dann gilt für den Fehler $f(x) - p(x)$ des Interpolationspolynoms folgende Aussage:

Für alle $x \in [a, b]$ gibt es ein $\xi = \xi(x)$, das also von x abhängen kann, mit

$$\xi \in (\min\{x_0, \dots, x_n, x\}, \max\{x_0, \dots, x_n, x\})$$

sodass gilt:

$$f(x) - p(x) = (x - x_0)(x - x_1) \cdots (x - x_n) \frac{f^{(n+1)}(\xi)}{(n+1)!}.$$

Somit ergibt sich für den Verfahrensfehler

$$\prod(x) := |(x - x_0)(x - x_1) \cdots (x - x_n)|.$$

3.9. Wie sind die Tschebyscheff-Polynome definiert? Welche Eigenschaften haben sie?

Die durch Abbildung $T_n : [-1, 1] \rightarrow [-1, 1]$ mit

$$T_n(x) = \cos(n \arccos(x))$$

definierten Polynome heißen Tschebyscheff-Polynome. Für sie gilt:

- (1) T_n ist ein Polynom n -ten Grades in $x = \cos \phi$.
- (2) Rekursionsformel für Tschebyscheff-Polynome:
 $T_0(x) = 1, T_1(x) = x$ und $T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x), n = 1, 2, \dots$
- (3) $T_n(x) \leq 1$ für $x \in [-1, 1]$.
- (4) T_n ist eine gerade oder ungerade Funktion, je nachdem, ob n gerade oder ungerade ist.
- (5) T_n besitzt ganzzahlige Koeffizienten. Der führende Koeffizient ist 2^{n-1} für $n \geq 1$.
- (6) T_n nimmt für $n \geq 1$ im Intervall $[-1, 1]$ $(n+1)$ -mal die Werte ± 1 an, nämlich für $x = \cos \frac{k\pi}{n}, k = 0, 1, \dots, n$.
 Insbesondere gilt $T_n(1) = 1, T_n(-1) = (-1)^n$.
- (7) T_n hat n reelle Nullstellen in $[-1, 1]$, nämlich die Tschebyscheff-Knoten

$$t_k^{(n)} = \cos\left(\frac{2k-1}{2n}\pi\right), k = 1, \dots, n.$$

- 3.10. **Wie berechnet man die Knoten für die Tschebyscheff-Interpolationspolynome im Intervall $[-1, 1]$ bzw. $[a, b]$? Welche Vorteile hat die Verwendung von Tschebyscheff-Knoten im Vergleich zu äquidistanten Stützstellen.**
 T_n hat n reelle Nullstellen in $[-1, 1]$, nämlich die Tschebyscheff-Knoten

$$t_k^{(n)} = \cos\left(\frac{2k-1}{2n}\pi\right), k = 1, \dots, n.$$

Für $[a, b]$ wird die Transformation

$$[-1, 1] \rightarrow [a, b], t \rightarrow x(t) = \frac{1}{2}(a+b) + \frac{1}{2}(b-a)t$$

ausgeführt und man erhält die Tschebyscheff-Knoten in $[a, b]$

$$x_k^{(n)} = \frac{a+b}{2} + \frac{b-a}{2}t_k^{(n)}. \quad t_k^{(n)} \text{ ist } k\text{-ter Tschebyscheff-Knoten in } [-1, 1]$$

Die Tschebyscheff-Interpolation ist wesentlich besser konditioniert.

- 3.11. **Wie lässt sich das dividierte Differenzschema und das Newtonsche Interpolationspolynom verallgemeinern, falls in den Stützstellen auch noch Ableitungen vorgegeben sind?**

Durch Hermite-Interpolation. Idee: Interpolation durch die Punkte $(x_i, y_i), (x_i+h, y_i+hy'_i), i = 0, \dots, n$, und Grenzübergang $h \rightarrow 0$. Führt zu Differenzenschema in dem Punkte doppelt angeschrieben und Werte für die man mit 0 dividieren müsste durch die gegebenen Ableitungen ersetzt werden.

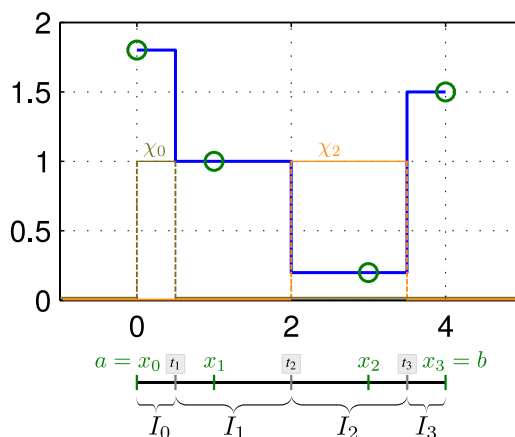
- 3.12. **Wie wird mit stückweise konstanten Funktionen interpoliert?**

Das Intervall $[a, b]$ wird in n Intervalle unterteilt, wobei die Intervallsgrenzen genau zwischen den Knoten liegen. Die Treppenfunktion definiert sich aus den Intervallen I_i und den Stützstellen x_i zu

$$s(x) = \sum_{i=0}^n y_i \chi_i(x)$$

mit

$$\chi_i(x) = \begin{cases} 1, & \text{falls } x \in I_i, \\ 0, & \text{sonst.} \end{cases}$$



3.13. Wie wird mit stetigen, stückweise linearen Funktionen interpoliert?

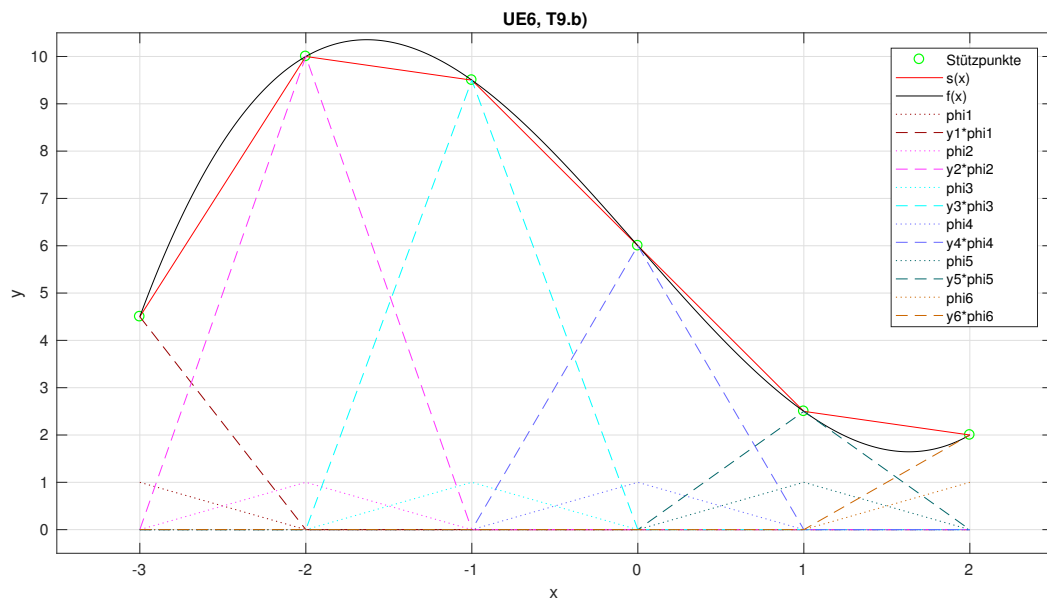
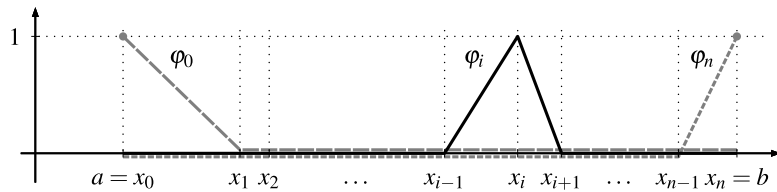
Wie oben, nur wird χ_i ersetzt durch die Hutfunktion

$$\varphi_i = \begin{cases} \frac{x-x_{i-1}}{x_i-x_{i-1}} & x \in [x_{i-1}, x_i], \quad \text{falls } i = 1, \dots, n \\ \frac{x_{i+1}-x}{x_{i+1}-x_i} & x \in [x_i, x_{i+1}], \quad \text{falls } i = 0, \dots, n-1 \\ 0, & \text{sonst} \end{cases}$$

was analog zu konstanten Funktionen auf

$$s(x) = \sum_{i=0}^n y_i \varphi_i(x)$$

führt.



3.14. Was sind Hutfunktionen und welche Eigenschaften haben sie?

Stückweise lineare Funktionen

$$\varphi_i = \begin{cases} \frac{x-x_{i-1}}{x_i-x_{i-1}} & x \in [x_{i-1}, x_i], \quad \text{falls } i = 1, \dots, n \\ \frac{x_{i+1}-x}{x_{i+1}-x_i} & x \in [x_i, x_{i+1}], \quad \text{falls } i = 0, \dots, n-1 \\ 0, & \text{sonst} \end{cases}$$

für die gelten

$$\phi_i(x_k) = \delta_{ij}.$$

3.15. Was für Eigenschaften besitzen kubische Splines? Was für Typen von kubischen Splines gibt es?

Eigenschaften:

- $s_i(x_i) = y_i$, $i = 0, \dots, n$ (Interpolationsbedingung)
- $s \in \mathcal{C}^2([a, b])$, d.h. zweimal stetig differenzierbar
- $s_i := s|_{[x_i, x_{i+1}]} \in \mathbb{P}_3([x_i, x_{i+1}])$, $i = 0, \dots, n-1$.

Typen:

- Natürlicher Spline
 $s''(x_0) = s''(x_n) = 0$
 Die Momente an beiden Enden sind 0.
- Eingespannter Spline (vollständiger Spline)
 $s'(x_0) = y'_0$, $s'(x_n) = y'_n$
 Die beiden Enden sind eingespannt und die Steigungen vorgegeben.
- Periodischer Spline
 $s(x_0) = s(x_n)$, $s'(x_0) = s'(x_n)$, $s''(x_0) = s''(x_n)$.
- Not a knot
 s''' ist stetig in x_1 und x_{n-1}
 Auf den ersten und letzten beiden Intervallen wird ein durchgehendes Polynom dritten Grades verwendet.

3.16. Wieso ist es besser durch viele Punkte einen kubischen Spline zu legen, statt ein Interpolationspolynom zu verwenden?

Polynome schaukeln sich am Rand auf und sind daher zur Interpolation ungeeignet, wenn man ein Polynom durch viele Punkte mit beliebigen, paarweise verschiedenen Abszissen legen will.

3.17. Wie wird auf einem rechteckigen Gitter zweidimensional interpoliert?

Analog zum eindimensionalen Fall benutzen wir zwei Lagrange-Polynome

$$\ell_i(x) = \frac{(x-x_0)(x-x_1) \cdots \widehat{(x-x_i)} \cdots (x-x_n)}{(x_i-x_0)(x_i-x_1) \cdots \widehat{(x_i-x_i)} \cdots (x_i-x_n)}$$

$$L_j(y) = \frac{(y-y_0)(y-y_1) \cdots \widehat{(y-y_j)} \cdots (y-y_m)}{(y_j-y_0)(y_j-y_1) \cdots \widehat{(y_j-y_j)} \cdots (y_j-y_m)}$$

um auf das Interpolationspolynom

$$p(x, y) = \sum_{i=0}^n \sum_{j=0}^m z_{ij} \ell_i(x) L_j(y)$$

zu kommen, wobei $\ell_i(x)L_j(y)$ das zur Stützstelle (x_i, y_j) gehörige Lagrange-Polynom ist. Wiederum gilt $\ell_i(x_i)L_j(y_j) = 1$ und $\ell_i(x_r)L_j(y_s) = 0$ für alle anderen Punkte $(x_r, y_s) \neq (x_i, y_j)$ des Gitters.

Nun kann man entweder primär von links nach rechts und dann von oben nach unten rechnen, oder umgekehrt, hier nur eine Art.

$p(x, y)$ wird umgeformt zu

$$p(x, y) = \sum_{j=0}^m \left(\sum_{i=0}^n z_{ij} \ell_i(x) \right) L_j(y) = \sum_{j=0}^m r_j(x) L_j(y).$$

$r_j(x)$ ist das eindeutige Interpolationspolynom vom Grad n durch die Punkte $(x_0, y_i, z_{0i}), \dots, (x_n, y_i, z_{ni})$.

3.18. **Wie wird die zweidimensionale, stetige, stückweise lineare Interpolierende auf einem rechteckigen Gitter bestimmt?**

Wir betrachten nur ein Rechteck $[x_i, x_{i+1}] \times [y_j, y_{j+1}]$ in dem der Punkt (x, y) liegt.

$$s(x, y) = z_{ij}\varphi_i(x)\Phi_j(y) + z_{i+1,i}\varphi_{i+1}(x)\Phi_j(y) + z_{i,j+1}\varphi_i(x)\Phi_{j+1}(y) + z_{i+1,j+1}\varphi_{i+1}(x)\Phi_{j+1}(y)$$

Wobei φ_i und Φ_j die Hutfunktionen in x- und y-Richtung sind.

4 Numerische Integration

4.1. Was bedeutet *Linearität* und *Positivität* des Integrals?

Eigenschaften des Integrals und der numerischen Approximation.

- Linearität: $\int_a^b (\alpha f(x) + \beta g(x)) dx = \alpha \int_a^b f(x) dx + \beta \int_a^b g(x) dx$
- Positivität: $f(x) \geq 0$ für alle $x \in [a, b] \Rightarrow \int_a^b f(x) dx \geq 0$.

4.2. Erklären Sie den Begriff *Quadraturformel*.

Eine Quadraturformel $Q = (b_i, c_i)_{i=1}^s$ ist eine Näherungsformel

$$I(g) = \int_0^1 g(t) dt \approx \sum_{i=1}^s b_i g(c_i) =: Q(g)$$

zur numerischen Berechnung eines Integrals auf dem Intervall $[0, 1]$. Die Zahlen b_1, \dots, b_s heißen Gewichte und c_1, \dots, c_s Knoten der Quadraturformel, s ist die Anzahl der Stufen. Für die Knoten wird

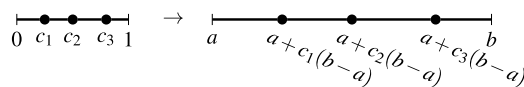
$$0 \leq c_1 < c_2 < \dots < c_s \leq 1$$

verlangt und für die Gewichte

$$b_1 + b_2 + \dots + b_s = 1.$$

Für ein beliebiges Intervall $[a, b]$ werden die Knoten von $[0, 1]$ nach $[a, b]$ transformiert und es gilt

$$I(g) = \int_a^b g(t) dt \approx (b-a) \sum_{i=1}^s b_i \underbrace{f(a + c_i(b-a))}_{g(c_i)} =: Q(f, [ab]).$$



4.3. Nennen Sie einige einfache Quadraturformeln inklusive Knoten und Gewichte.

Regel	s	c_i		b_i
Linksregel	1	0		1
Rechtsregel	1	1		1
Mittelpunktsregel	1	$\frac{1}{2}$		1
Trapezregel	2	0 1		$\frac{1}{2}$ $\frac{1}{2}$
Simpsonregel	3	0 $\frac{1}{2}$ 1		$\frac{1}{6}$ $\frac{4}{6}$ $\frac{1}{6}$

4.4. Erklären Sie den Begriff *zusammengesetzte Quadraturformel*.

Ist der Abstand $b - a$ sehr groß oder ändert sich der Integrand im Integrationsbereich $[a, b]$ rasch, zerlegt man $[a, b]$ in n Teilintervalle $[x_0, x_1], [x_1, x_2], \dots, [x_{n-1}, x_n]$ mit $a = x_0 < x_1 < x_2 < \dots < x_{n-1} < x_n = b$. Das Integral wird aufgespalten in eine Summe von n Teilintegralen über die einzelnen Teilintervalle:

$$\int_a^b f(x) dx = \int_{x_0}^{x_1} f(x) dx + \int_{x_1}^{x_2} f(x) dx + \dots + \int_{x_{n-1}}^{x_n} f(x) dx$$

Jedes dieser Teilintegrale $\int_{x_{k-1}}^{x_k} f(x) dx$ wird mit einer Quadraturformel numerisch berechnet.

Genauer Definiert mit der Definition aus 4.2:

Es sei eine Quadraturformel $Q = (b_i, c_i)_{i=1}^s$ und eine Unterteilung des Intervalls $[a, b]$ wie oben gegeben. Dann heißt eine Näherungsformel

$$\int_a^b f(x) dx \approx \sum_{k=1}^n Q(f, [x_{k-1}, x_k]) = \sum_{k=1}^n h_k \sum_{i=1}^s b_i f(x_{k-1} + c_i h_k) =: S_Q(f, x_0, \dots, x_n)$$

mit $h_k = x_k - x_{k-1}$ zusammengesetzte Quadraturformel oder Summe von Quadraturformeln. Für eine äquidistante Unterteilung von $[a, b]$ in n Teilintervalle, also

$$x_k = a + kh, \quad h = \frac{b-a}{n}, \quad k = 0, \dots, n,$$

haben wir folgende Näherungsformel

$$\int_a^b f(x) dx \approx h \sum_{k=1}^n \sum_{i=1}^s b_i f(x_{k-1} + c_i h) =: S_Q(f, h, [a, b]) = S_Q(f, h).$$

4.5. Wie erhält man Quadraturformeln mit Hilfe von Polynominterpolation?

Die zu integrierende Funktion g wird durch ein Polynom q ersetzt und dieses integriert. Die Stützstellen der Polynominterpolation sind die Knoten der Quadraturformel. Die Gewichte erhält man durch Integrieren der Lagrange-Polynome:

$$b_i = \int_0^1 \ell_i(t) dt, \quad \ell_i(t) = \prod_{j=1, j \neq i}^s \frac{t - c_j}{c_i - c_j}.$$

Für die Knoten c_i gibt es verschiedene Ansätze, zum Beispiel Abgeschlossene Newton-Cotes-Formeln

$$c_i = \frac{i-1}{s-1}, \quad i = 1, \dots, s.$$

Für bessere Näherungen verwendet man jedoch Tschebyscheff-Knoten was zu den Clenshaw-Curtis-Formeln führt:

$$c_i = \frac{1}{2} \sin\left(\frac{(2i-1)\pi}{2s}\right) + \frac{1}{2}$$

4.6. Erklären Sie den Begriff *Ordnung einer Quadraturformel*. Wie bestimmt man die Ordnung?

Eine Quadraturformel $Q = (b_i, c_i)_{i=1}^s$ hat Ordnung p genau dann, wenn alle Polynome q vom Grad $\leq p-1$ exakt integriert werden, also $I(q) = Q(q)$ gilt.

4.7. Was sind *Bedingungsgleichungen*?

Alternativ muss die Quadraturformel die Bedingungsgleichung erfüllen:

$$\sum_{i=1}^s b_i c_i^k = \frac{1}{k+1}, \quad k = 0, \dots, p-1$$

4.8. Erklären Sie die Begriffe *Fehler einer Quadraturformel* und *Fehlerkonstante*.

Gegeben sei ein Intervall $[a, b]$ und die Quadraturformel $Q(f, [a, b]) = h \sum_{i=1}^s b_i f(a + c_i h)$ mit Schrittweite $h = b - a$. Für stetige Funktionen $f \in \mathcal{C}([a, b])$ definiert man den Fehler der Quadraturformel als

$$E(f) = I(f) - Q(f) = \int_a^{a+h} f(x) dx - h \sum_{i=1}^s b_i f(a + c_i h).$$

- 4.9. **Was für Abschätzungen gelten für den Fehler einer Quadraturformel bzw. einer zusammengesetzten Quadraturformel? Was muss der Integrand f dabei erfüllen?**

Siehe 4.5 Fehler einer Quadraturformel, Satz 4.3 Fehler einer Quadraturformel und Satz 4.4 Fehler von zusammengesetzten Quadraturformeln.

- 4.10. **Was sind *symmetrische Quadraturformeln* und welche Eigenschaft besitzen sie?**

Eine Quadraturformel $b_i, c_i)_{i=1}^s$ heißt symmetrisch, wenn für die Knoten und Gewichte folgendes gilt:

$$c_i = 1 - c_{s+1-i}, \quad \text{und} \quad b_i = b_{s+1-i}, \quad i = 1, \dots, s.$$

Die Knoten und Gewichte sind also an $1/2$ gespiegelt. Die Ordnung einer symmetrischen Quadraturformel ist gerade.

- 4.11. **Was ist eine *Gaußsche Quadraturformel*? Welche Ordnung besitzen sie?**

Eine Quadraturformel $b_i, c_i)_{i=1}^s$ mit s Stufen heißt Gaußsche Quadraturformel, wenn die Knoten c_1, \dots, c_s die auf das Intervall $[0, 1]$ transformierte Nullstelle x_1, \dots, x_s des Legendre Polynoms P_s vom Grad s sind, also

$$c_i = \frac{1}{2} + \frac{x_i}{2}, \quad i = 1, \dots, s.$$

Die Gewichte b_i sind so bestimmt, dass sie die Bedingungsgleichung

$$\sum_{i=1}^s b_i c_i^k = \frac{1}{k+1}, \quad k = 0, \dots, s-1$$

erfüllen.

Die Gaußsche Quadraturformel mit s Stufen besitzt die Ordnung $p = 2s$.

- 4.12. **Wie groß kann die Ordnung einer Quadraturformel maximal sein?**

$$p = 2s$$

- 4.13. **Was gilt für die Gewichte einer Gaußschen Quadraturformel?**

Die Gewichte b_i sind so bestimmt, dass sie die Bedingungsgleichung

$$\sum_{i=1}^s b_i c_i^k = \frac{1}{k+1}, \quad k = 0, \dots, s-1$$

erfüllen.

- 4.14. **Wie funktioniert eine *Schrittweitensteuerung*? Erklären Sie die Begriffe *Fehlerkriterium* und *Fehlerschätzer*.**

Um Rechenzeit einzusparen will man äquidistanten Intervallen dort verfeinert, wo sich die Funktion f rasch ändert. Das Fehlerkriterium dient zur Entscheidung, ob ein betrachtetes Integrationsintervall $[\alpha, \beta]$, ein Teilintervall von $[a, b]$, weiter verfeinert werden muss, oder ob das Fehlerkriterium erfüllt ist. Dann wäre der Fehler $E(f, [\alpha, \beta])$ des Ergebnisses $Q(f, [\alpha, \beta])$ akzeptabel. Da man das genaue Ergebnis nicht kennt, muss man einen Fehlerschätzer verwenden um einen Wert zu erhalten den man mit der Toleranz TOL vergleichen kann.

- 4.15. **Erklären Sie den Begriff *Richardson-Extrapolation*. Wie berechnet man est und Q_{extr} .**

- 4.16. **Was passiert bei Integranden mit Singularitäten oder Singularitäten in den Ableitungen?**

Reduktion der Ordnung von Qc^p zu Qc^q mit $q < p$.

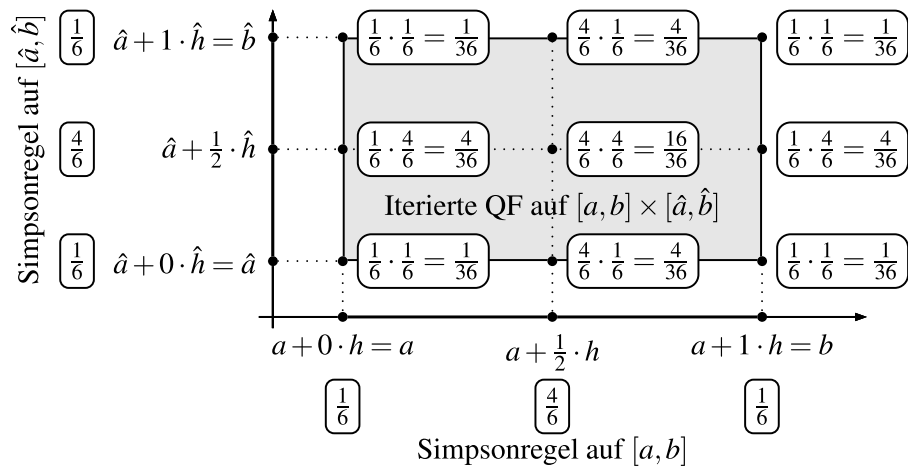
4.17. **Wie werden Doppelintegrale auf Rechtecken numerisch berechnet?**

Durch Quadraturformeln für Rechtecke.

Wir verwenden auf dem Rechteck $R = [a, b] \times [\hat{a}, \hat{b}]$ in x -Richtung einen Schritt der Quadraturformel $Q = (b_i, c_i)_{i=1}^s$ und in y -Richtung einen Schritt der Quadraturformel $\hat{Q} = (\hat{b}_i, \hat{c}_i)_{i=1}^{\hat{s}}$:

$$\iint_R f(x, y) \, dF = \int_a^b \int_{\hat{a}}^{\hat{b}} f(x, y) \, dy \, dx \approx h \sum_{i=1}^s b_i \int_{\hat{a}}^{\hat{b}} f(a + c_i h, y) \, dy \approx h \hat{h} \sum_{i=1}^s \sum_{j=1}^{\hat{s}} b_i \hat{b}_j f(a + c_i h, \hat{a} + \hat{c}_j \hat{h})$$

mit $h = b - a$ und $\hat{h} = \hat{b} - \hat{a}$. Man hat also eine Quadraturformel mit Gewichten b_i, \hat{b}_j , die f in den Punkten $(a + c_i h, \hat{a} + \hat{c}_j \hat{h})$ auswertet.



4.18. **Wie werden Doppelintegrale auf Dreiecken numerisch berechnet? Wie überprüft man die Ordnung einer Quadraturformel für Dreiecke?**

4.19. **Was sind baryzentrische Koordinaten?**

5 Lineare Algebra

5.1. Wie ist die LU-Zerlegung definiert?

Es sei A eine invertierbare $n \times n$ -Matrix. Dann existieren drei ebenfalls invertierbare Matrizen und zwar eine Permutationsmatrix P , eine untere Dreiecksmatrix L und eine obere Dreiecksmatrix U , d.h.

$$L = \begin{bmatrix} 1 & 0 & \dots & 0 \\ \ell_{21} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ \ell_{n1} & \dots & \ell_{n,n-1} & 1 \end{bmatrix}, \quad U = \begin{bmatrix} u_{11} & u_{12} & \dots & u_{1n} \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & u_{n-1,n} \\ 0 & \dots & 0 & u_{nn} \end{bmatrix}$$

mit $u_{ii} \neq 0$, $i = 1, \dots, n$, sodass

$$PA = LU$$

gilt. Diese Zerlegung heißt LU-Zerlegung. Die Multiplikation mit P von links entspricht einer Vertauschung der Zeilen von A wegen der Spaltenpivotsuche.

5.2. Was bedeutet Spaltenpivotsuche und wieso wird sie verwendet?

Für die LU-Zerlegung wählt man das betragsmäßig größte Element der Pivotspalte als Pivotelement. Dieser Vorgang wird als Spaltenpivotsuche bezeichnet.

Im k -ten Schritt der LU-Zerlegung heißt dies: Suche das betragsgrößte Element in der k -ten Spalte in den Zeilen $k, k+1, \dots, n$!

5.3. Wie erhält man die Permutationsmatrix bzw. den Permutationsvektor?

Permutationsmatrix P

- Starte mit der Einheitsmatrix
- Führe an dieser Matrix ebenfalls die entsprechend der Pivotsuche beim Gaußschen Algorithmus notwendige Zeilenvertauschung durch.

Permutationsvektor IP mit Länge n

- $IP[n] = (-1)^{\text{Anzahl Vertauschungen}}$
- $IP[k] = m$
Im k -ten Eliminationsschritt wird die Zeile m mit der aktuellen Zeile k vertauscht.
- Alternativ: Man erstellt Vektor $\begin{bmatrix} 1 & 2 & \dots & n \end{bmatrix}^T$ und wendet darauf die Vertauschen an.

5.4. Wie löst man die LU-Zerlegung lineare Gleichungssysteme $AX = b$?

Schritt 1: Vertauschen der Elemente von b mit Hilfe von $IP \Rightarrow \tilde{b}$

Schritt 2: Vorwärtssubstitution mit L und \tilde{b}

$$Ly = \tilde{b} \Rightarrow y$$

Schritt 3: Rückwärtssubstitution mit U und y

$$Ux = y \Rightarrow x$$

5.5. Wie berechnet man $\det A$ mit Hilfe der LU-Zerlegung?

$$\det A = u_{11} \cdot u_{22} \cdot \dots \cdot u_{nn} \cdot IP[n]$$

5.6. Wie groß ist der Rechenaufwand der LU-Zerlegung und für das Auflösen eines linearen Gleichungssystems?

- LU-Zerlegung: $\frac{n^3}{3} + O(n^2)$ Operationen
- Auflösen l.Gls.: $n^2 + O(n)$ Operationen

5.7. Wie ist die Cholesky-Zerlegung einer Matrix A definiert? Welche Eigenschaften muss A besitzen?

Cholesky-Zerlegung: $A = CC^T$ wobei der Cholesky-Faktor C eine untere Dreiecksmatrix von A ist.
Eigenschaften die A haben muss:

- reell: Die Koeffizienten von A sind reelle Zahlen;
- symmetrisch: $A = A^T$;
- positiv definit: Für alle $x \neq 0$ gilt $x^T A x > 0$.

5.8. Wie berechnet man die Cholesky-Zerlegung?

Für alle Spalten $k = 1, \dots, n$ von links beginnend:

$$\text{Diagonalelement } c_{kk} = \sqrt{a_{kk} - c_{k1}^2 - \dots - c_{k,k-1}^2}$$

Für alle Elemente $c_{ik}, i > k$, also unterhalb von c_{kk} :

$$c_{ik} = (a_{ik} - c_{i1}c_{k1} - \dots - c_{i,k-1}c_{k,k-1})/c_{kk}$$

5.9. Wie groß ist der Rechenaufwand einer Cholesky-Zerlegung?

$\frac{n^3}{6} + O(n^2)$, Hälfte der LU-Zerlegung dank Symmetrie von A.

5.10. Wie löst man mit der Cholesky-Zerlegung lineare Gleichungssysteme $Ax = b$?

$$Ax = b \iff C \underbrace{C^T x}_y = b$$

$$Cy = b \Rightarrow y$$

$$C^T x = y \Rightarrow x$$

5.11. Wie ist die rationale Cholesky-Zerlegung einer Matrix A definiert? Welche Eigenschaften muss A besitzen?

$$A = LDL^T$$

Wobei L eine untere Dreiecksmatrix mit Einsen in der Diagonale und D eine Diagonalmatrix mit $d_{ii} \neq 0$ ist.
A muss invertierbar und symmetrisch sein.

5.12. Wie berechnet man die rationale Cholesky-Zerlegung?

5.13. Wie löst man mit der rationalen Cholesky-Zerlegung lineare Gleichungssysteme $Ax = b$?

$$Ax = b \iff LDL^T x = b$$

5.14. Wie ist die Matrixnorm allgemein definiert?

Gegeben sei ein Vektorraum V. Eine Norm V ist eine Abbildung $\|\cdot\| : V \rightarrow \mathbb{R}$ mit den folgenden Eigenschaften:

- (i) $\|x\| \geq 0, \quad \|x\| = 0 \Rightarrow x = 0$,
- (ii) $\|\lambda x\| = |\lambda| \cdot \|x\|$,
- (iii) $\|x + y\| \leq \|x\| + \|y\|$. (Dreiecksungleichung)

5.15. **Wie berechnet man $\|A\|_1$, $\|A\|_2$ und $\|A\|_\infty$?**

- Es sei A eine $m \times n$ -Matrix. Die Matrixnorm für die 1-Norm

$$\|A\|_1 = \max_{1 \leq j \leq n} \left(\sum_{i=1}^m |a_{ij}| \right)$$

ist die maximale Spaltenbetragssumme. Merkregel: 1 steht wie eine Spalte.

- Die Matrixnorm für die ∞ -Norm

$$\|A\|_\infty = \max_{1 \leq i \leq m} \left(\sum_{j=1}^n |a_{ij}| \right)$$

ist die maximale Zeilenbetragssumme. Merkregel: ∞ liegt wie eine Zeile.

- Die Matrixnorm für die 2-Norm (euklidische Norm):

$$\|A\|_2 = \sqrt{\text{größter Eigenwert von } A^T A}$$

5.16. **Wie ist die Norm von A^{-1} definiert?**

$$\|A^{-1}\| := \left(\inf_{x \neq 0} \frac{\|Ax\|}{\|x\|} \right)^{-1} = \left(\inf_{\|y\|=1} \|Ay\| \right)^{-1}$$

5.17. **Wie ist die Kondition einer linear Abbildung bzw. einer Matrix A definiert?**

$$\kappa(A) := \|A\| \cdot \|A^{-1}\|$$

5.18. **Was gilt für den relativen Fehler der Lösung eines linearen Gleichungssystems?**

$$\frac{\|\tilde{x} - x\|}{\|x\|} \leq \frac{\kappa(A)}{1 - \varepsilon_A \cdot \kappa(A)} (\varepsilon_A + \varepsilon_b)$$

5.19. **Welche Eigenschaften besitzt die Kondition einer Matrix?**

- (1) $\kappa(A) \geq 1$.
- (2) Für $\alpha \neq 0$ ist $\kappa(\alpha A) = \kappa(A)$
- (3) $\kappa(A) = \frac{\sup_{\|y\|_2=1} \|Ay\|_2}{\inf_{\|y\|_2=1} \|Ay\|_2} = \frac{\sqrt{\lambda_{\max}}}{\sqrt{\lambda_{\min}}}$

5.20. **Welche Matrizen haben eine große/kleine Kondition?**

Groß: Hilbertmatrix, Vandermonde-Matrix

Klein: $A = \alpha I$, $\kappa(A) = \kappa(\alpha I) = \kappa(I) = 1$

Für eine orthogonale Matrix Q gilt $\kappa_2(Q) = 1$

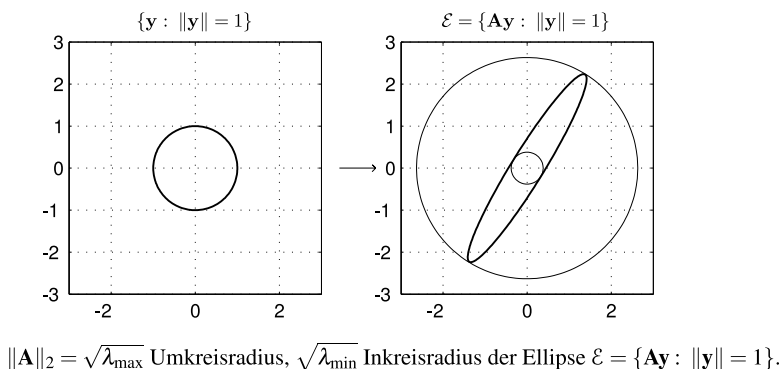
5.21. **Was ist die Kondition der Einheitsmatrix?**

$$\kappa(I) = 1$$

5.22. Wie lässt sich die Kondition einer Matrix geometrisch interpretieren?

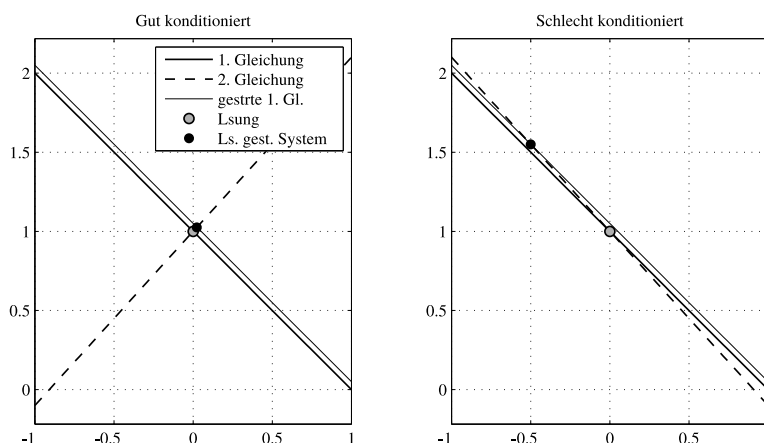
Verzerrungsfaktor: Es sei $\mathcal{E} = \{A \cdot y : \|y\|_2 = 1\}$ jene Ellipse, die durch Anwendung der Matrix auf A auf alle Vektoren mit Länge 1 erzeugt wird. Dann gilt:

$\kappa(A) = \text{Umkreisradius/Inkreisradius der Ellipse } \mathcal{E} = \text{Verhältnis der Hauptachsen von } \mathcal{E} = \text{Verzerrungsfaktor von A.}$



5.23. Wie lässt sich die Kondition eines linearen Gleichungssystems geometrisch interpretieren?

Eine Störung von b in einem $Ax=b$ System verschiebt die erste Gerade leicht. Die neue Lösung eines gut Konditionierten Problems liegt wie in der Abbildung zu sehen näher an einer ungestörten Lösung als die eines schlecht Konditionierten.



Gut und schlecht konditionierte lineare Gleichungssysteme.

5.24. Wie ist die QR-Zerlegung definiert? Wozu wird sie verwendet?

Es sei A eine $m \times n$ -Matrix mit $m \geq n$ und $\text{Rg } A = n$. Dann existiert eine orthogonale $m \times m$ -Matrix Q mit

$$A = QR \quad \text{und} \quad R = \begin{bmatrix} r_{11} & \cdots & r_{1n} \\ & \ddots & \vdots \\ 0 & & r_{nn} \\ 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{bmatrix}.$$

Die $m \times n$ -Matrix R besteht aus einer oberen $n \times n$ -Dreiecksmatrix und einem $(m-n) \times n$ Block aus Nullen darunter. Die Diagonalelemente r_{ii} sind alle von Null verschieden.

QR-Zerlegung wird verwendet um überbestimmte lineare Gleichungssysteme im Sinne der kleinsten Fehlerquadrate zu lösen.

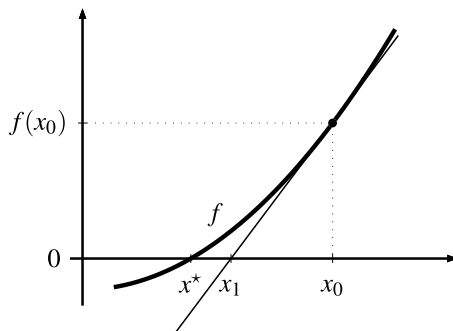
6 Nichtlineare Gleichungssysteme

6.1. Beschreiben Sie das Newton-Verfahren in einer Veränderlichen. Skizze!

Es sei $f : (a, b) \rightarrow \mathbb{R}$ eine stetig differenzierbare Funktion und x^* eine einfache Nullstelle von f , d.h.

$$f(x^*) = 0 \quad \text{und} \quad f'(x^*) \neq 0.$$

Weiters sei x_0 eine (grobe) Näherung an die gesuchte Nullstelle x^* und $f'(x_0) \neq 0$. Man bestimmt die Tangente im Punkt $(x_0, f(x_0))$ an den Graphen der Funktion f und berechnet deren Nullstelle x_1 . Die Nullstelle x_1 der Tangente ist eine neue und hoffentlich bessere Näherung an die Nullstelle x^* von f .



Eine Iteration mit dem Newtonverfahren.

Nun gilt

$$\frac{f(x_0) - 0}{x_0 - x_1} = f'(x_0) \Rightarrow x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}$$

Allgemein Formuliert:

$$x_0 = \text{Startwert mit } f'(x_0) \neq 0$$

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}, \quad k \geq 0$$

6.2. Wie ist das Newton-Verfahren in mehreren Veränderlichen definiert?

$$x^{(0)} = \text{Startwert mit } f'(x^{(0)}) \text{ invertierbar.}$$

$$\text{löse } f'(x^{(k)}) \cdot \Delta^{(k)} = -f(x^{(k)}), \quad \text{mit } \Delta x^{(0)} = x^{(1)} - x^{(0)}$$

$$x^{(k+1)} = x^{(k)} + \Delta x^{(k)}, \quad k \geq 0$$

Falls $\|\Delta x^{(k)}\| \leq \text{TOL}$ wird das Verfahren abgebrochen und $x^{(k+1)}$ als numerisches Ergebnis Akzeptiert.

6.3. Wie schnell konvergiert das Newton-Verfahren? Welche Voraussetzungen muss der Startwert erfüllen?

Die Funktion f sei in einer Umgebung von x^* zweimal stetig differenzierbar und es gelte $f(x^*) = 0$ und $\det f'(x^*) \neq 0$. Dann gilt

$$\|x^{(k+1)} - x^*\| \leq C \cdot \|x^{(k)} - x^*\|^2.$$

6.4. Wie ist das Gauß-Newton-Verfahren definiert?

$$x^{(0)} = \text{Startwert}$$

$$\text{löse } f'(x^{(k)})^T f'(x^{(k)}) \cdot \Delta^{(k)} = -f(x^{(k)})^T f(x^{(k)}), \quad \text{mit } \Delta x^{(0)} = x^{(1)} - x^{(0)}$$

$$x^{(k+1)} = x^{(k)} + \Delta x^{(k)}, \quad k \geq 0$$

Falls $\|\Delta x^{(k)}\| \leq \text{TOL}$ wird das Verfahren abgebrochen und $x^{(k+1)}$ als numerisches Ergebnis Akzeptiert.

6.5. Wie ist das Konvergenzverhalten des Gauß-Newton-Verfahrens?

Es gibt Konstanten $C_1 > 0$ und $C_2 > 0$, sodass in der Nähe von x^* gilt:

$$\|x^{(k+1)} - x^*\| \leq C_1 \cdot f(x^*) \cdot \|x^{(k)} - x^*\| + C_2 \cdot \|x^{(k)} - x^*\|^2.$$

Was liest man daraus?

- (1) Die Konvergenz ist im Allgemeinen nur linear wegen $C_1 \cdot f(x^*) \cdot \|x^{(k)} - x^*\|$.
- (2) Falls $f(x^*) = 0$ (Die Gleichungen sind in diesem Fall kompatibel, d.h. nicht widersprüchlich), dann ist die Konvergenz sogar quadratisch, weil $C_1 \cdot f(x^*) \cdot \|x^{(k)} - x^*\|$ wegfällt.
- (3) Falls $f(x^*)$ zu groß ist, dann divergiert das Verfahren.

7 Ausgleichsrechnung und Approximation

7.1. Was bedeutet der Begriff *Methode der kleinsten Fehlerquadrate*?

Die Summe der Fehlerquadrate

$$\sum_{i=0}^m (f(x_i) - y_i)^2$$

soll minimal sein.

7.2. Was sind Gründe eine Funktion oder ein Polynom nicht direkt durch die Punkte hindurch zu legen, sondern dazwischen hindurch?

Die Werte y_i sind mit Fehlern behaftet (Rundungsfehler, Messfehler, Modellfehler). Zudem sind meist mehr Datenpunkte gegeben, als für eine eindeutige Bestimmung der Modellparameter a der Modellfunktion f notwendig sind. Somit gibt es im Allgemeinen keine Modellfunktion f , die exakt durch sämtliche Datenpunkte hindurchpasst.

7.3. Was bedeutet der Begriff *Modellfunktion*?

Die Modellfunktion

$$f: \mathbb{R}^d \times \mathbb{R}^{n+1} \rightarrow \mathbb{R}: y = f(x_1, \dots, x_d, a_0, \dots, a_n) = f(x, a_0, \dots, a_n)$$

hängt von den Inputdaten $x = (x_1, \dots, x_d)$ und von sogenannten Modellparametern $a = (a_0, \dots, a_n)$ ab und ist eine (möglichst einfache) Funktion, die den Zusammenhang zwischen den Inputdaten x_i und den Outputdaten y_i beschreiben soll.

7.4. Erklären Sie den Begriff *lineare Ausgleichsrechnung*. Wie lässt sich hier die Modellfunktion schreiben?

Bei einer linearen Ausgleichsrechnung hängt die Modellfunktion f für festes x nur linear von den Modellparametern a_0, \dots, a_n ab. Dadurch lässt sich f als Linearkombination aus linear unabhängigen (aber nicht unbedingt linearen) Funktionen f_0, \dots, f_n schreiben.

$$Fa = y$$

Sollte f linear unabhängig sein, kann man obige Gleichung im Sinne der kleinsten Fehlerquadrate gelöst werden.

$$\text{Suche } a_0, \dots, a_n, \text{ sodass } \|Fa - y\|^2 = \sum_{i=0}^m (F_{i*}a - y_i)^2 \text{ minimal.}$$

7.5. Erklären Sie den Begriff *nichtlineare Ausgleichsrechnung*.

Bei einer nichtlinearen Ausgleichsrechnung hängt die Modellfunktion f für festes x nichtlinear von den Modellparametern a_0, \dots, a_n ab. Setzt man die Datenpunkte $(x_0, y_0), \dots, (x_m, y_m)$ in die Modellfunktion ein erhält man ein überbestimmtes, nichtlineares Gleichungssystem

$$\begin{aligned} f(x_0, a_0, \dots, a_n) - y_0 &= 0 \\ f(x_1, a_0, \dots, a_n) - y_1 &= 0 \\ &\vdots \\ f(x_m, a_0, \dots, a_n) - y_m &= 0 \end{aligned}$$

7.6. Worin besteht der Unterschied zwischen *linearer* und *nichtlinearer* Ausgleichsrechnung?

In der linearen/nichtlinearen Abhängigkeit der Modellfunktion f von den Modellparametern a_0, \dots, a_n .

7.7. Sie legen eine Ausgleichsparabel zwischen den Punkten hindurch. Ist dies lineare oder nichtlineare Ausgleichsrechnung?

linear

7.8. Welche zwei gegensätzlichen Ziele werden bei den *glättenden Splines* betrachtet bzw. optimiert?

- (1) Die Splinefunktion s soll "möglichst gut zwischen den Datenpunkten hindurchpassen.
D.h. es soll die Summe $\sum_{i=0}^n (y_i - s(x_i))^2$ der Fehlerquadrate möglichst klein sein.
- (2) Die Splinefunktion s soll "möglichst glatt sein, d.h. die potentielle Energie bzw. die Gesamtkrümmung bzw. das Integral $\int_{x_0}^{x_n} (s''(x))^2$ soll möglichst klein sein.

7.9. Was bedeutet ein Wert des Glättungsparameters λ nahe 0 bzw. nahe 1?

Obige widersprüchliche Ziele werden zum Optimierungsproblem

$$\lambda \cdot \sum_{i=0}^n (y_i - s(x_i))^2 + (1 - \lambda) \cdot \int_{x_0}^{x_n} (s''(x))^2 \rightarrow \min, \quad \lambda \in (0, 1]$$

Verknüpft, wobei λ zwischen ihnen Gewichtet.

- λ nahe bei 0 bedeutet, dass die Glattheit des Splines sehr wichtig ist.
- λ nahe bei 1 bedeutet, dass kleine Fehlerquadrate wichtiger sind.

Man wählt also zwischen $s \rightarrow$ Ausgleichsgerade für $\lambda \rightarrow 0$ und ungeglätteten Spline für $\lambda \rightarrow 1$.

7.10. Was passiert, wenn man $\lambda = 1$ nimmt?

Bei $\lambda = 0$ wäre jede Funktion s mit $s'' = 0$ Lösung des Optimierungsproblems, da dann der Fehler völlig egal wäre. Daher wird der Fall nicht betrachtet.

Allerdings gilt: $s \rightarrow$ Ausgleichsgerade für $\lambda \rightarrow 0$.

7.11. Was passiert, wenn man λ gegen 0 gehen lässt?

Bei $\lambda = 1$ ist die potentielle Energie bzw. Glattheit egal und nur die Fehlerquadrate wichtig. In diesem Fall erhalten wir wieder den üblichen Spline, der exakt durch die Datenpunkte geht.