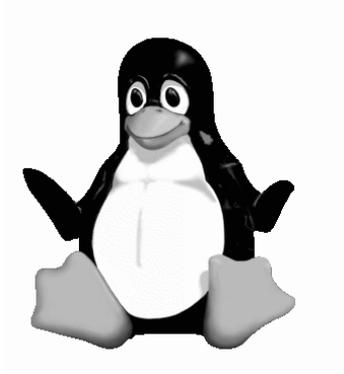


Numerik

Dr. Stephan

WS 1999/2000

Stand: 15. Oktober 2001



Gewidmet allen freizeitlosen Mathematikern.

Dieses Dokument ist unsere Mitschrift der Vorlesung "Numerik I" von Dr. Stephan im Wintersemester 1999/2000.

Die Nummerierung der Lemmata, Sätze, etc. haben wir teilweise etwas unseren Normen angepaßt.

Für eventuelle Fehler aller Art übernehmen wir keine Verantwortung !

Vielen Dank allen Entwicklern von $\text{T}_\text{E}\text{X}$, X-Fig, GnuPlot und Linux !

Tobias Müller, Heike Trusch

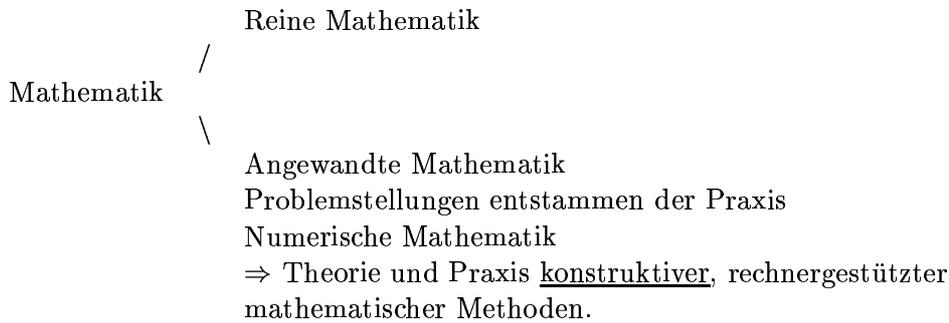
Inhaltsverzeichnis

Standortbestimmung und Einleitung	5
Standortbestimmung und Einleitung	5
konstruktive Methode	5
Algorithmus	5
Absoluter und relativer Fehler	7
Maschinenzahlen	8
Normweise Konditionsanalyse	9
1 Interpolation durch Polynome	10
Existenz eines approximierenden Polynoms	10
Lagrange Interpolationspolynom	10
Newton-Interpolationspolynom	11
Interpolation bei äquidistanten Knoten	12
Zusammenhang zwischen vorwärts Differenz und dividierter Differenz	13
Interpolationsfehler	14
Knotenpolynom	16
Hermite-Interpolation	16
Stückweise polynomiale Interpolation	18
Interpolation mit kubischen Splines	20
2 Lineare Gleichungssysteme, direkte Verfahren	24
2.1 Vorbemerkung	24
2.2 Gauß 'scher Algorithmus	25
"Methoden zur Rechenstabilität"	25
positiv definit	31
3 Vektor- und Matrixnormen, Fehlerabschätzungen für LGS, iterative Verfahren	33
3.1 Vektor- und Matrixnormen	33
3.2 Iterative Verfahren zur Lösung von LGS	38
$A^m \rightarrow 0$	38
Konvergenz des SOR-Verfahrens	39
4 Nichtlineare Gleichungssysteme	42
4.1 Das Banach'sche Verfahren	42
4.2 Das Newton-Verfahren	43
4.3 Nichtlineare GSV, ESV und SOR-Verfahren	44
Methode des steilsten Abstiegs	45
konjugierte Gradienten Methode (CG)	46
5 Approximation	48
Existenz und Eindeutigkeit der besten Approximation	49
(Gram-Schmidt-Orthonormalisierung)	51

Inhaltsverzeichnis

6	Numerische Quadratur	54
	Trapez-Regel	54
	Newton-Cotes-Formeln	55
	Satz von Peano	56
7	Zusammengesetzte Quadraturformeln	60
	Zusammengesetzte Trapez- und Simpsonregel	60
	Gauß-Quadratur für beliebiges Intervall	63
	Numerische Lösung von Systemen von nichtlinearen Gleichungen	65
7.1	Picard Iteration	65
	Banach-Fixpunkt-Satz	65
7.2	Newton-Verfahren	69
	lokale Konvergenz des Newton-Verfahrens	69
7.3	Bisections-Methode	72
8	Matrix-Eigenwertaufgaben	73
8.1	Vorbemerkungen, Eigenwertabschätzungen	73
8.2	Die Verfahren von Wilkinson und Householder	74
	8.2.1 Wilkinson-Algorithmus	75
	8.2.2 Verfahren von Householder	77
8.3	Berechnung von Eigenwerten und Eigenvektoren von Hessenberg-Matrizen	79
8.4	von Mises'sche Iterationsverfahren (1929)	82
8.5	Das Jacobi-Verfahren (1846!)	84
8.6	QR-Verfahren (Francis, 1961)	85

Inhaltsverzeichnis



Definition: Eine konstruktive Methode zur Lösung eines mathematischen Problems ist eine Folge von Anweisungen über die Durchführung von logischen und arithmetischen Operationen.

Definition: Ein Algorithmus ist die Spezifikation einer endlichen Folge von Operationen mit reellen Zahlen, die jedem m-Tupel $a = (a_1, \dots, a_m)$ von reellen Zahlen aus einem gewissen Bereich D des \mathbb{R}^n , den Daten, ein n-Tupel $x = (x_1, \dots, x_n)^T$ von reellen Zahlen, die Ergebnisse, zuordnet:

$$x = r(a) \Leftrightarrow \begin{cases} x_1 = r_1(a_1, \dots, a_m) \\ \vdots \\ x_n = r_n(a_1, \dots, a_m) \end{cases}$$

Die Ergebnisse entstehen dabei aus den Daten über eine Folge von Zwischenergebnissen y_i

$$\begin{aligned} y_1 &= z_1(a_1, \dots, a_m) = z_1(a) \\ y_2 &= z_2(a, y_1) \\ y_3 &= z_3(a, y_1, y_2) \\ &\vdots \\ y_I &= z_I(a, y_1, \dots, y_{I-1}) \end{aligned}$$

wobei z_i "elementare Abbildungen" sind. Die Ergebnisse $x_\nu (\nu = 1, \dots, n)$ sind eine spezielle Auswahl der $y_i (1 \leq i_\nu \leq I, \nu = 1, \dots, n)$

Implementierung eines Algorithmus:

Daten a	↔	a' Maschinendaten
↓ y		↓ r' Implementierung
Ergebnisse x		x' Maschinenergebnis
$ a - a' < \epsilon$	⇒	$ x - x' ?$

Ökonomiebedingungen

1. bezüglich der verfügbaren Hilfsmittel (Einfachheit)
2. Rechenaufwand / Rechenzeit (Effizienz, Komplexität)
3. Konvergenzgeschwindigkeit
4. möglichst großer Anwendungsbereich (Flexibilität)

Inhaltsverzeichnis

- | | |
|---|-------------------------|
| 5. Voraussetzungen | (Zuverlässigkeit) |
| 6. Unempfindlichkeit gegen Rundungsfehler | (numerische Stabilität) |
| 7. Unempfindlichkeit gegenüber Datenfehlern | (gute Kondition) |

Beispiel:

$$\begin{aligned}
 I &= \int_0^1 \frac{x^4}{x+5} dx \\
 I_n &= \int_0^1 \frac{x^n}{x+5} dx \Rightarrow \lim_{n \rightarrow \infty} I_n = 0 \\
 &= \int_0^1 \frac{x^n + 5x^{n-1}}{x+5} dx - 5 \int_0^1 \frac{x^{n-1}}{x+5} dx \\
 &= \int_0^1 x^{n-1} - 5 \int_0^1 \frac{x^{n-1}}{x+5} dx \\
 &= \frac{1}{n} - 5 \int_0^1 \frac{x^{n-1}}{x+5} dx
 \end{aligned}$$

$I_0 = \log \frac{6}{5}$	Maschinenalgorithmus
$I_1 = 1 - 5 \cdot I_0$	$I'_0 = \underline{0,182}$
$I_2 = \frac{1}{2} - 5 \cdot I_1$	$I'_1 = 1 - 5 \cdot I'_0 = \underline{0,090}$
$I_3 =$	$I'_2 = 1,500 - 5 \cdot I'_1 = \underline{0,050}$
$I_4 =$	$I'_3 = 0,333 - 5 \cdot I'_2 = \underline{0,083}$
	$I'_4 = 0,250 - 5 \cdot I'_3 = -0,165$

Laut TI85: $I_4 = 0,034306329555$
 _:=gültige Stellen

Anderer Algorithmus:

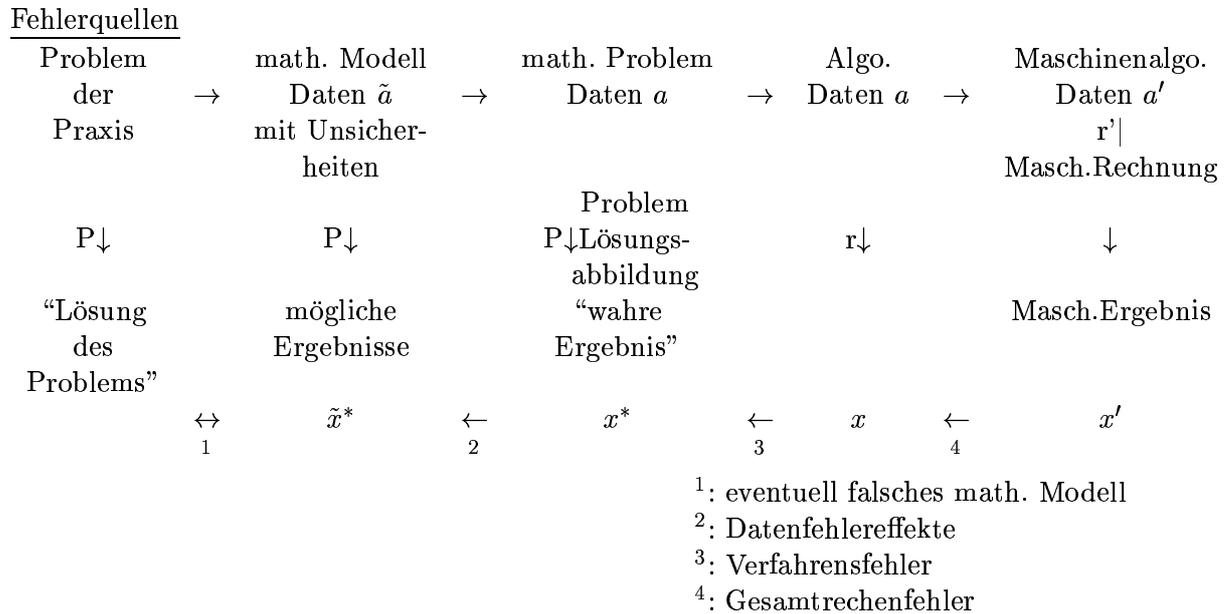
Sei i_{10} näherungsweise 0.

$i_{10} = 0$	$i'_{10} = 0$
$i_9 = \frac{1}{5} * (\frac{1}{10} - i_{10})$	$i'_9 = 0,020$
$i_8 = \frac{1}{5} * (\frac{1}{9} - i_9)$	$i'_8 = 0,018$
$i_7 = \frac{1}{5} * (\frac{1}{8} - i_8)$	$i'_7 = 0,021$
$i_6 = \frac{1}{5} * (\frac{1}{7} - i_7)$	$i'_6 = 0,024$
$i_5 = \frac{1}{5} * (\frac{1}{6} - i_6)$	$i'_5 = 0,029$
$i_4 = \frac{1}{5} * (\frac{1}{5} - i_5)$	$i'_4 = \underline{\underline{0,034}}$

gutartig !

Hier wird der Fehler bei jedem Schritt verringert ($\frac{1}{5}$)

Inhaltsverzeichnis



Definition: Absoluter und relativer Fehler

Seien a, \hat{a} reelle Zahlen.

1. \hat{a} heißt Näherung für a , bis auf einen Fehler, der dem Betrage nach kleiner oder gleich ϵ ist :gdw $|\hat{a} - a| \leq \epsilon$
 $\hat{a} - a$ heißt wahrer Fehler
2. $a - \hat{a}$ heißt die Korrektur von \hat{a} als Näherung für a
3. $|a - \hat{a}|$ heißt absoluter Fehler von \hat{a} als Näherung für a
4. wenn $a \neq 0$ ist, heißt $|\frac{a-\hat{a}}{a}|$ relativer Fehler von \hat{a} als Näherung für a (üblicherweise in Prozent)
5. \hat{a} besitzt als Näherung für a N gültige Dezimalen gdw \hat{a} ist als Näherung für a auf N Dezimalen genau gdw $a - \hat{a} \leq \frac{1}{2} * 10^{-N}$ ($N \in \mathbb{Z}$)
6. \hat{a} sei als Näherung für a auf N Dezimalen genau. Die Ziffern von \hat{a} , die in der Position 10^{-N} und davot stehen, heißen signifikante Ziffern (oder wesentliche) von \hat{a} als Näherung für a , wobei führende Nullen nicht mitgezählt werden

Bemerkung:

- Anzahl der gültigen Dezimalen ist scharfe Abschätzung für absolute Fehler
- Anzahl der signifikanten Ziffern ist grobes Maß für den relativen Fehler

Inhaltsverzeichnis

Bemerkung:

Subtraktion annähernd gleich großer Zahlen führt zu einem Verlust von signifikanten Ziffern.

Beispiel: Lösung von $x^2 - 26x + 1 = 0$

$$x^2 + p * x + q = 0 \Rightarrow x_{1,2} = -\frac{p}{2} \pm \sqrt{\left(\frac{p}{2}\right)^2}$$

$$\text{Satz von Vieta: } x_2 = \frac{q}{x_1}$$

1. Maschinentalgorithmus:

$$p = -26$$

$$u = -\frac{p}{2} = 13$$

$$v = u^2 = 13^2 = 169$$

$$w = u^2 - q = 169$$

$$r = \sqrt{w} = 12,961$$

$$\hat{x}_1 = u + r = \underline{25,961}$$

$$\hat{x}_2 = u - r = \underline{0,039}$$

5 signifikante Ziffern

2 signifikante Ziffern

2. Maschinentalgorithmus:

⋮

$$\hat{x}_2 = \frac{q}{\hat{x}_1} = \frac{1}{25,961} = 0,038519$$

5 signifikante Ziffern

Maschinenzahlen

Die mit einer bestimmten Codierung darstellbaren Zahlen bilden die Menge \mathbb{M} der Maschinenzahlen.

1. Festpunktdarstellung
2. Gleitpunktdarstellung einer Zahl x zur Basis B .
(engl.: floating point representation)

$x = \sigma * m * B^e$, wobei $2 \leq B \in \mathbb{N}$ Basis des Ziffernsystems

$m = (0.a_1 a_2 \dots a_N)_B$ $0 \leq a_k \leq B - 1$ Mantisse von x

N Stellenzahl (Länge) der Mantisse

$e \in \mathbb{Z}$, $L \leq e \leq U$ Exponent $\sigma = \text{sign } x \in \{-1, 1\}$

Die Gleitpunkt-Darstellung heißt normalisiert gdw. $a_1 \neq 0$ (oder $m = 0$ falls $x = 0$).

Wir brauchen eine Abbildung

$\gamma : \mathbb{R} \rightarrow \mathbb{M}$ (Reduktionsabbildung)

1. Abschneiden
2. Runden

$$\mathbb{R} \ni x \mapsto \gamma x =: \tilde{x} \text{ mit } |\tilde{x} - x| = \min_{\tilde{y} \in \mathbb{M}} |\tilde{y} - x|$$

Inhaltsverzeichnis

In den Fällen, wo es zwei von x gleichweit entfernte \tilde{y} gibt, sollte man so runden, daß die letzte Ziffer eine gerade Zahl ist.

Gegeben $\mathbb{M} = \mathbb{M}(B, N, L, U)$.

Dann ist $X_{max} = (0.a_1 \cdots a_n)_B * B^U$, $a_k = B - 1$, $k = 1, \dots, N$

$X_{min} = (0.1)_B * B^L$ die größte bzw. kleinste positive Zahl in \mathbb{M} .

Gegeben γ , so heißt die bestmögliche Schranke für den relativen Fehler von γ .

$$\epsilon_M := \sup_{\substack{x \in \mathbb{R} \\ |x| \in [x_{min}, x_{max}]}} \left| \frac{x - \gamma x}{x} \right|$$

die Maschinengenauigkeit des Rechners.

Dann gilt: $\forall x \in \mathbb{R}, |x| \in [x_{min}, x_{max}] \exists \substack{\delta \in \mathbb{R} \\ |\delta| \leq \epsilon_M} \gamma x = x(1 + \delta)$

Beispiel: Für $\gamma =$ “Runden” gilt $\epsilon_M = \frac{\frac{1}{2}B^{1-N}}{\frac{1}{2}B^{1-N}} \leq \frac{1}{2}B^{1-N}$

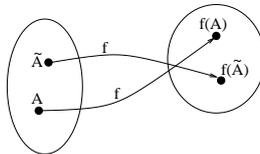
Normweise Konditionsanalyse

Betrachte die folgende Aufgabe:

Gegeben: $A \in U$ }
 Gesucht: $x = f(A)$ } =: Problem (f, A)

Das Problem (f, A) heißt wohlgestellt bzgl. absoluter Datenänderungen gemessen in der Norm $\| \cdot \|$ gdw.

$$\exists \delta > 0 \exists L_{abs} \geq 0 \forall \tilde{A} \in \overline{K} \substack{\| \cdot \|} \substack{(A, \delta) \leq U} \|f(\tilde{A}) - f(A)\| \leq L_{abs} \|\tilde{A} - A\|. \quad (1)$$



Das Problem (f, A) heißt wohlgestellt bzgl. relativer Datenänderungen gdw.

$$A \neq 0, f(A) \neq 0 \text{ und } \exists \delta \exists L_{rel} \geq 0 \forall \tilde{A} \in \overline{K} \substack{\| \cdot \|} \substack{(A, \delta \|A\|) \leq U} \frac{\|f(\tilde{A}) - f(A)\|}{\|f(A)\|} \leq L_{rel} \frac{\|\tilde{A} - A\|}{\|A\|} \quad (2)$$

Die kleinste Konstante L_{abs} mit (1) wird mit $L_{abs}(\delta)$ bezeichnet.

Die kleinste Konstante L_{rel} mit (2) wird mit $L_{rel}(\delta)$ bezeichnet.

Definition: Die Kondition des wohlgestellten Problems (f, A) bzgl. absoluter (relativer) Datenänderungen ist die Zahl

$$K_{abs}(f, A) := \lim_{\delta \downarrow 0} L_{abs}(\delta)$$

$$K_{rel}(f, A) := \lim_{\delta \downarrow 0} L_{rel}(\delta)$$

Folgerung: Es gilt näherungsweise $\|f(\tilde{A}) - f(A)\| \leq K_{abs}(f, A) \|\tilde{A} - A\|$

$$\frac{\|f(\tilde{A}) - f(A)\|}{\|f(A)\|} \leq K_{rel}(f, A) \frac{\|\tilde{A} - A\|}{\|A\|}$$

Gut konditioniert bzgl. absoluter Datenänderungen: $K_{abs}(f, A) \approx 1$.

Schlecht konditioniert bzgl. absoluter Datenänderungen: $K_{abs}(f, A) \gg 1$.

1 Interpolation durch Polynome

Aufgabe: Gegeben: Funktionswerte $f(x_0), f(x_1), \dots, f(x_n)$

$f_\nu := f(x_\nu)$ ($0 \leq \nu \leq n$) in paarweise verschiedenen Knoten.

$x_0 < x_1 < \dots < x_\nu < \dots < x_n \in \mathbb{R}$.

Finde ein Polynom vom Grade $\leq n$ mit $p(x_\nu) = f_\nu$ ($0 \leq \nu \leq n$).

Theorem 1.1 Seien $f_0, f_1, \dots, f_n \in \mathbb{R}$ und $x_0, x_1, \dots, x_n \in \mathbb{R}$ gegeben mit $x_i \neq x_j$ für ein $i \neq j$. Dann existiert genau ein Polynom p_n vom Grad $\leq n$ mit $p_n(x_\nu) = f_\nu$ ($\nu = 0, \dots, n$)

Beweis:

a) Existenz: Konstruktion von Polynomen

$\mathbb{P}_n = \{g : \mathbb{R} \rightarrow \mathbb{C}, g \text{ Polynom, Grad } g \leq n\}$

$l_\nu \in \mathbb{P}_n$ mit $l_\nu(x_i) = \delta_{\nu i} = \begin{cases} 1, & i = \nu \\ 0, & i \neq \nu \end{cases}$ ($\nu = 0, \dots, n$)

$$\mathbb{P}_n \ni l_\nu(x) := \prod_{i=0, i \neq \nu}^n \frac{x-x_i}{x_\nu-x_i} = \frac{\overset{\text{Kronecker-Symbol}}{(x-x_0)(x-x_1)\dots(x-x_{\nu-1})(x-x_{\nu+1})\dots(x-x_n)}}{(x_\nu-x_0)(x_\nu-x_1)\dots(x_\nu-x_{\nu-1})(x_\nu-x_{\nu+1})\dots(x_\nu-x_n)}$$

$$\boxed{p_n(x) := \sum_{\nu=0}^n f_\nu l_\nu(x)}$$

Lagrange Interpolationspolynom

$$\Rightarrow p_n(x_k) = \sum_{\nu=0}^n f_\nu l_\nu(x_k) = f_k \underbrace{l_k(x_k)} = 1 = f_k$$

$\Rightarrow p_n$ interpoliert f_ν in x_ν

b) Eindeutigkeit:

Seien $p, q \in \mathbb{P}_n$, die die Interpolationsaufgabe erfüllen.

$z := p - q \in \mathbb{P}_n$ hat $n+1$ Nullstellen x_ν ($\nu = 0, 1, \dots, n$)

(denn $p(x_\nu) = f_\nu = q(x_\nu)$ ($\nu = 0, 1, \dots, n$))

$\Rightarrow z \equiv 0$ nach dem Fundamentalsatz der Algebra.

Beispiel:

$f(x) := x * \sin(\pi x)$, $x_\nu := -1 + \frac{\nu}{2}$, $\nu = 0, \dots, 4$

ν	0	1	2	3	4
x_ν	-1	$-\frac{1}{2}$	0	$\frac{1}{2}$	1
f_ν	0	$\frac{1}{2}$	0	$\frac{1}{2}$	0

$$\begin{aligned} p_4(x) &= f_3 * l_3(x) + f_1 * l_1(x) \\ &= \frac{1}{2} * \underbrace{\frac{(x-x_0)(x-x_1)(x-x_2)(x-x_4)}{(x_3-x_0)(x_3-x_1)(x_3-x_2)(x_3-x_4)}}_{l_3} + \frac{1}{2} * \underbrace{\frac{(x-x_0)(x-x_2)(x-x_3)(x-x_4)}{(x_1-x_0)(x_1-x_2)(x_1-x_3)(x_1-x_4)}}_{l_1} \\ &= \frac{1}{2} * \frac{(x+1)(x+\frac{1}{2})x(x-1)}{(\frac{1}{2}+1)(\frac{1}{2}+\frac{1}{2})\frac{1}{2}(\frac{1}{2}-1)} + \frac{1}{2} * \frac{(x+1)x(x-\frac{1}{2})(x-1)}{(-\frac{1}{2}+1)(-\frac{1}{2})(-\frac{1}{2}-\frac{1}{2})(-\frac{1}{2}-1)} \\ &= \frac{8}{3}x^2(1-x^2) \end{aligned}$$

Andere Möglichkeit:

1 Interpolation durch Polynome

Definiton: (Dividierte Differenzen)

$$f[x_\nu, \dots, x_{\nu+k}] := \begin{cases} f[x_\nu] (= f(x_\nu) = f_\nu) & , k = 0 \\ \frac{f[x_{\nu+1}, \dots, x_{\nu+k}] - f[x_\nu, \dots, x_{\nu+k-1}]}{x_{\nu+k} - x_\nu} & , k > 0 \end{cases} \quad (2)$$

für $0 \leq k, \nu; \nu + k \leq n$ **Dividierte Differenzen**

$$f[x_\nu, x_{\nu+1}] = \frac{f[x_{\nu+1}] - f[x_\nu]}{x_{\nu+1} - x_\nu} = \frac{f(x_{\nu+1}) - f(x_\nu)}{x_{\nu+1} - x_\nu}$$

Tabelle für dividierte Differenzen:

$k = 0$	x_0 $f[x_0]$	x_1 $f[x_1]$	x_2 $f[x_2]$	x_3 $f[x_3]$	\dots	x_{n-1} $f[x_{n-1}]$	x_n $f[x_n]$
$k = 1$	$f[x_0, x_1]$		$f[x_1, x_2]$		$f[x_2, x_3]$		\dots
$k = 2$	$f[x_0, x_1, x_2]$			$f[x_1, x_2, x_3]$			$f[x_{n-2}, x_{n-1}, x_n]$
$k = 3$	$f[x_0, x_1, x_2, x_3]$					$f[x_{n-3}, \dots, x_n]$	
\vdots							
$k = n$	$f[x_0, x_1, \dots, x_n]$						

Vorausgesetzt $x_\nu, \nu = 0, \dots, n$ sind alle paarweise verschieden, dann sind $f[x_0, \dots, x_{\nu+k}]$ rekursiv durch (2) definiert $0 \leq k, \nu; \nu + k \leq n$.

Theorem 1.2 Seien x_0, x_1, \dots, x_n paarweise verschieden, dann gilt:

$$p_n(x) = \sum_{k=0}^n f[x_0, \dots, x_k] * \prod_{j=1}^k (x - x_{k-j}) \quad (3)$$

Newton-Interpolationspolynom

Beweis: (durch Induktion über n)

$$n = 0 \Rightarrow p_0(x) = f[x_0] * 1 = f(x_0) = f_0 \quad \checkmark$$

Annahme: Behauptung (3) gilt für $n - 1 \geq 0$

$$\Rightarrow p_{n-1}(x) = \sum_{k=0}^{n-1} f[x_0, \dots, x_k] * \prod_{j=1}^k (x - x_{k-j})$$

interpoliert f_0, f_1, \dots, f_{n-1} in den Knoten x_0, \dots, x_{n-1}

\Rightarrow das (eindeutig bestimmte) Interpolationspolynom p_n kann geschrieben werden:

$$p_n(x) = p_{n-1}(x) + a * \underbrace{\prod_{j=1}^n (x - x_{n-j})}_{\neq 0 \text{ für } x_n}$$

(wobei $a \in \mathbb{R}$ noch bestimmt werden muß)

Wähle a so, daß p_n in x_n interpoliert !

Berechnung von a :

$$\frac{d^n}{dx^n} p_n(x) + p_{n-1}(x) + a * n! = \frac{d^n}{dx^n} p_n(x) \stackrel{(1)}{=} \sum_{j=0}^n f_j \frac{d^n}{dx^n} f_j(x)$$

$$= n! \sum_{j=0}^n f_j (\prod_{i=0, i \neq j}^n (x_j - x_i))^{-1}$$

$$\stackrel{\text{Lemma 1}}{=} n! f[x_0, \dots, x_n]$$

□

1 Interpolation durch Polynome

Beispiel: (dividierte Differenzen)

$x_0 = 0$	1		
	$\frac{2}{3}$		
$x_1 = 1,5$	2	$\underline{-0,26667}$	
	0		$\underline{0,022222}$
$x_2 = 2,5$	2	$-0,16667$	
	$-\frac{1}{2}$		
$x_3 = 4,5$	1		

(z.B. $-0.266667 = \frac{0-\frac{2}{3}}{2.5-0}$)
 Zahlenpaare für x_i gegeben!
 $\prod_{x=a}^b f(x) = 1$, für $b < a$

⇒ Newton Interpolationspolynom

$$\begin{aligned}
 p_x(x) &= f[x_0] + f[x_0, x_1](x - x_0) + f[x_0, x_1, x_2](x - x_0)(x - x_1) + \\
 &\quad + f[x_0, x_1, x_2, x_3](x - x_0)(x - x_1)(x - x_2) \\
 &= \underbrace{f[x_0]}_1 + \underbrace{\{(x-x_0) f[x_0, x_1]\}}_{\frac{2}{3}} + \underbrace{(x-x_1) \{f[x_0, x_1, x_2]\}}_{-0,26667} + \underbrace{(x-x_2) f[x_0, x_1, x_2, x_3]}_{0.02222}
 \end{aligned}$$

⇒ $p_3(2, 1) = 2,0528$

Lemma 1.1 Seien x_ν paarweise verschieden, $0 \leq k \leq n$:

$$f[x_0, \dots, x_k] = \sum_{\nu=0}^k f(x_\nu) * \left(\prod_{j=0, j \neq \nu}^k (x_\nu - x_j) \right)^{-1}$$

Beweis: (durch Induktion über k)

$k = 0$: $f_0 = f(x_0) = f[x_0]$ ✓

Annahme: Behauptung gilt für $k - 1$

$$\begin{aligned}
 f[x_0, \dots, x_k] &= \frac{1}{x_k - x_0} \{ \underbrace{f[x_1, \dots, x_k]} - \underbrace{f[x_0, \dots, x_{k-1}]} \} \\
 &= \frac{1}{x_k - x_0} \{ \sum_{\nu=0}^{k-1} f(x_{\nu+1}) * \prod_{j=0, j \neq \nu}^{k-1} (x_{\nu+1} - x_{j+1})^{-1} - \\
 &\quad - \sum_{\nu=0}^{k-1} f(x_\nu) * \prod_{j=0, j \neq \nu}^{k-1} (x_\nu - x_j)^{-1} \}
 \end{aligned}$$

$$f[x_1, \dots, x_{1+(k-1)}] = \sum_{\nu=0}^{k-1} f(x_{\nu-1}) * \left(\prod_{j=0, j \neq \nu}^{\nu-1} (x_{\nu+1} - x_{j+1}) \right)^{-1}$$

$$\begin{aligned}
 \Rightarrow f[x_0, \dots, x_k] &= \\
 &= \frac{1}{x_k - x_0} \left\{ \frac{f(x_k)}{\prod_{j=1}^k} + \sum_{\nu=0}^{k-1} f(x_\nu) \left[\frac{1}{\prod_{j=1, j \neq \nu}^k (x_\nu - x_j)} - \frac{1}{\prod_{j=0, j \neq \nu}^{k-1} (x_\nu - x_j)} \right] - \frac{f(x_0)}{\prod_{j=0, j \neq \nu}^{k-1} (x_0 - x_j)} \right\} \\
 &= \frac{f_k}{\prod_{j=0, j \neq \nu}^k (x_k - x_j)} + \sum_{\nu=1}^{k-1} f_\nu \frac{1}{\prod_{j=0, j \neq \nu}^k (x_\nu - x_j)} \underbrace{\frac{x_\nu - x_0 - (x_\nu - x_k)}{x_k - x_0}}_1 + \frac{f_0}{\prod_{j=0, j \neq \nu}^k (x_0 - x_j)} \\
 &= \sum_{\nu=0}^k f_\nu * \frac{1}{\prod_{j=0, j \neq \nu}^k (x_\nu - x_j)} \quad \square
 \end{aligned}$$

Interpolation bei äquidistanten Knoten

$$x_i = a + ih, \quad i = 0, \dots, N, \quad N = \frac{b-a}{h}$$

Variablentransformation:

$$s(x) = \frac{x-x_0}{h} \Rightarrow x(s) = x_0 + s * h$$

1 Interpolation durch Polynome

$$f(x) = f(x_0 + s * h) =: f_s$$

Aufgabe: Berechne Interpolationspolynom

$$p_n(x) \text{ für } f(x) \text{ in } x_k, \dots, x_{k+n}$$

Es ist keine Tabelle der "dividierten Differenzen" nötig. Es genügt eine Differenzen-Tabelle zu berechnen.

"Vorwärtsgerichtete Differenz" $\Delta^i f_s := \begin{cases} f_s & (i = 0) \\ \Delta(\Delta^{i-1} f_s) = \Delta^{i-1} f_{s+1} - \Delta^{i-1} f_s & (i > 0) \end{cases}$

Lemma 1.2 (Zusammenhang zwischen vorwärts Differenz und dividierter Differenz)

Für alle $i \geq 0 : f[x_k, \dots, x_{k+i}] = \frac{1}{i!h^i} \Delta^i f_k$ (4)

Beweis: (durch Induktion)

$i = 0 : f[x_k] = f(x_k) = f_k = \Delta^0 f_k$ ✓

Angenommen (4) gilt für $i = n \geq 0$, dann

$$\begin{aligned} f[x_k, \dots, x_{k+n+1}] &= \frac{f[x_{k+1}, \dots, x_{k+n+1}] - f[x_k, \dots, x_{k+n}]}{x_{k+n+1} - x_k} \\ &= \frac{(\Delta^n f_{k+1} / n!h^n) - (\Delta^n f_k / n!h^n)}{(n+1)h} \\ &= \frac{\Delta^{n+1} f_k}{(n+1)!h^{n+1}} \end{aligned}$$

□

Interpolationspolynom für $f(x)$ in x_k, \dots, x_{k+n}

$$P_n(x) = \sum_{i=0}^n \underbrace{f[x_k, \dots, x_{k+i}]}_{\frac{1}{i!h^i} \Delta^i f_k} \prod_{j=0}^{i-1} (x - x_{k+j})$$
 Newton

$$x - x_{k+j} = x_0 + s * h - [x_0 + (k + j)h] = (s - k - j)h$$

$$P_n(x) = P_n(x_0 + s * h) = \sum_{i=0}^n \Delta^i f_k \prod_{j=0}^{i-1} \frac{s-k-j}{j+1} = \sum_{i=0}^n \Delta^i f_k \binom{s-k}{i} \tag{5}$$

Binomial Funktion: $\binom{y}{i} := \begin{cases} 1 & , i = 0 \\ \frac{y(y-1)\dots(y-i+1)}{1*2*\dots*i} \left(= \prod_{j=0}^{i-1} \frac{y-j}{j+1} \right) & , i > 0 \end{cases}$

Beispiel:

$x_0 = 20$	$f_0 \downarrow$ <u>.34202</u>	\searrow	$\Delta f_0 \downarrow$ <u>.15798</u>	\searrow	$\Delta^2 f_0 \downarrow$ <u>-.01519</u>	\searrow	$\Delta^3 f_0 \downarrow$ <u>-.00435</u>
$x_1 = 30$.5	\swarrow	.14279	\searrow	-.01954	\swarrow	
$x_2 = 40$.642	\swarrow	.12325	\swarrow			
$x_3 = 50$.76604	\swarrow					

(Differenzen der beiden Vorgänger, unten-oben)

$$\begin{aligned} P_3(x) \sum_{i=0}^3 \Delta^i f_0 \prod_{j=0}^{i-1} \frac{s-j}{j+1} &= f_0 + \Delta f_0 * s + \Delta^2 f_0 * s * \frac{s-1}{2} + \Delta^3 f_0 * s * \frac{s-1}{2} * \frac{s-2}{3} \\ &= f_0 + s[\Delta f_0 + \frac{s-1}{2} \{ \Delta^2 f_0 + \frac{s-2}{3} \Delta^3 f_0 \}] \end{aligned}$$

$$P_3(36) = 0,58778$$

(da $x = x_0 + sh \Rightarrow s = \frac{x-x_0}{x_1-x_0}$, hier: $s = 1.6$)

1 Interpolation durch Polynome

Interpolationsfehler

Theorem 1.3

Sei $I := [a, b] \in \mathbb{R}$, $a < b$, $x_0, \dots, x_n \in I$ mit $x_j \neq x_i, \forall i \neq j$ und $w(x) = \prod_{\nu=0}^n (x - x_\nu)$
Knotenpolynom

Sei $f \in C^{n+1}(I)$ und $p_n(x)$ das zugehörige Interpolationspolynom.

Dann gilt:

$$\forall x \in I \exists \xi(x) \in I f(x) - p_n(x) = \frac{w(x)}{(n+1)!} f^{(n+1)}(\xi(x)) \quad (6)$$

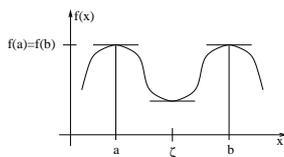
Beweis:

Sei $I \ni x \neq x_0, x_1, \dots, x_n$ (da sonst (6) trivial), x fest, $c := \frac{f(x) - p_n(x)}{\omega(x)}$

$F(t) := f(t) - p_n(t) - c * \omega(t)$, $t \in I$, hat $n + 2$ Nullstellen in I ,
 nämlich $t = x_0, x_1, \dots, x_n$ und $t = x$.

Satz von Rolle: Sei $f(x)$ stetig in $a \leq x \leq b$ und differenzierbar in $a < x < b$.

Falls $f(a) = f(b)$, dann existiert mindestens ein Punkt ζ zwischen a und b ,
 so daß $f'(\zeta) = 0$



$\Rightarrow \frac{dF}{dt}$ hat nach dem Satz von Rolle in den offenen Intervallen
 (zwischen den x'_ν s und x) $n + 1$ Nullstellen.

$\Rightarrow \frac{d^2 F}{dt^2}$ hat n Nullstellen und so fort.

$\Rightarrow \frac{d^{n+1} F}{dt^{n+1}}$ hat eine Nullstelle $\xi(x) \in I$ (gar im Inneren von I).

$\Rightarrow 0 = \frac{d^{n+1}}{dt^{n+1}} [f(t) - p_n(t) - c * \omega(t)]|_{t=\xi(x)} = f^{(n+1)}(\xi(x)) - c * (n + 1)!$

Da $\frac{d^{n+1}}{dt^{n+1}} p_n(t) = 0$ und $\frac{d^{n+1}}{dt^{n+1}} \omega(t) = (n + 1)! \Rightarrow c = \frac{f^{(n+1)}(\xi(x))}{(n+1)!}$

c einsetzen liefert (6) □

Für $f \in C^{n+1}(I)$ können wir obiges Theorem benutzen, um den Interpolationsfehler abzuschätzen:

$$\|f - p_n\|_\infty := \max_{x \in I} |f(x) - p_n(x)| \leq \frac{1}{(n+1)!} \|\omega\|_\infty * \|f^{(n+1)}\|_\infty \quad (7)$$

$(\omega \leq \frac{n!}{4} h^{(n+1)}, h := \max_{i=1, \dots, n} (x_i - x_{i-1}))$

Beispiel: $f(x) = \sin(x)$. Interpoliere $f(x)$ durch ein quadratisches Polynom

in den Knoten $x_0 = -h$, $x_1 = 0$, $x_2 = h$

Wie muß h gewählt werden, damit der zugehörige

Interpolationsfehler $< 5 * 10^{-8}$ bleibt ?

$$f(x) - p_2(x) = \frac{(x-x_0)(x-x_1)(x-x_2)}{3!} f^{(3)}(\xi)$$

$$\omega(x) = (x - x_0)(x - x_1)(x - x_2) = (x + h)(x)(x - h) = x^3 - h^2 x$$

1 Interpolation durch Polynome

notwendig für Maximum von $\omega(x) : \omega'(x) = 0 \Rightarrow 3x^2 - h^2 = 0 \Rightarrow x = \pm \frac{h}{\sqrt{3}}$

$$|\omega(\pm \frac{h}{\sqrt{3}})| = |\frac{h^3}{3\sqrt{3}} - \frac{h^3}{\sqrt{3}}| = \frac{2}{3} \frac{h^3}{\sqrt{3}}$$

Ferner $|f^{(3)}(\xi)| \leq 1$

Folglich $\max_{-h \leq x \leq h} |f(x) - p(x)| \leq \frac{1}{3!} \frac{2}{3} \frac{h^3}{\sqrt{3}} * 1 = \frac{\sqrt{3}}{27} h^3 < 5 * 10^{-8}$

$\Rightarrow \underline{\underline{h \approx 0,01}}$

Frage: Gilt immer $\lim_{n \rightarrow \infty} \|f - p_n\|_{\infty} = 0$?

Nein !

Gegenbeispiel: $f(x) = \frac{1}{1+x^2}, \quad -5 \leq x \leq 5$

In (7) können wir im Allgemeinen nicht die Konstante $\|f^{(n+1)}\|_{\infty}$ verringern, aber durch eine spezielle Wahl der Knoten können wir $\|\omega\|_{\infty}$ minimieren. Der Einfachheit halber betrachten wir Interpolationen im Einheitsintervall $J = [-1, 1]$. Der allgemeine Fall kann immer darauf zurückgeführt werden:

$$\phi(x) = \frac{1}{2}(a+b) + \frac{1}{2}(b-a) * x$$

erfüllt $I = \phi(J)$ wobei $I = [a, b]$ und ferner

$$\|f - p_n\|_{I, \infty} = \|f(\phi(\cdot)) - p_n(\phi(\cdot))\|_{J, \infty}$$

$$\mathbb{P}_n = \{p : \mathbb{R} \rightarrow \mathbb{C}, p \text{ Polynom vom Grad } \leq n\}$$

Definition: (Tschebyscheff-Polynome)

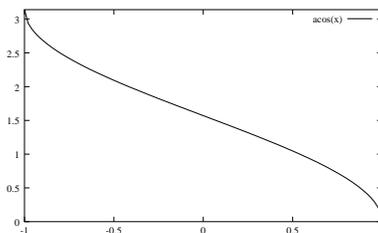
Die Polynome $T_n \in \mathbb{P}_n$ mit

$$T_n(x) = \cos(n * \arccos(x)) \tag{8}$$

heißen die Tschebyscheff-Polynome (1. Ordnung).

T_n sind Polynome, denn $T_0(x) = 1, T_1(x) = x$ und mit $\Phi := \arccos(x)$

$$\begin{aligned} T_{n+1}(x) + T_{n-1}(x) &= \cos(n+1)\Phi + \cos(n-1)\Phi \\ &= 2 * \cos n\Phi \cos \Phi = 2T_n(x) * x \end{aligned} \tag{9}$$



das heißt $T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x)$

Folglich mit (9) und (10) sehen wir, daß T_n Polynome sind.

Nullstellen von T_n :

$n * \arccos x_k = (k + \frac{1}{2})\pi, \quad k \text{ ganzzahlig}, x_k \in J = [-1, 1]$

$$\stackrel{(ii)}{\Rightarrow} x_k = \cos \frac{k+\frac{1}{2}}{n} \pi, \quad k = 0, 1, 2, \dots, n-1$$

(durch Symmetrie des Cosinus und seiner Periode 2π erhalten wir für alle anderen ganzzahligen k keine weiteren x_k)

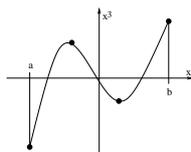
Extremwerte von T_n :

1 Interpolation durch Polynome

(8) $\Rightarrow |T_n(x)| \leq 1 \Rightarrow T_n(y_k)$ ist maximal oder minimal für $y_k \in J$ mit $|T_n(y_k)| = 1$

$$\Rightarrow n * \arccos y_k = k * \pi \stackrel{(12)}{\Rightarrow} y_k = \cos \frac{k * \pi}{n}, \quad 0 \leq k \leq n.$$

Dies sind schon $n + 1$ verschiedene y_k , denn als Polynom n^{ten} Grades kann T_n nicht mehr Extrema haben (2 Randextrema + $(n-1)$ innere Extrema)



Theorem 1.4 (Knotenpolynom)

Das Knotenpolynom $\omega(x) = \prod_{\nu=0}^n (x - x_\nu) \in \mathbb{P}_{n+1}$ mit der kleinsten ∞ -Norm ist das Polynom

$$\omega_{n+1}(x) := 2^{-n} T_{n+1}(x) \text{ mit } \|\omega_{n+1}\|_{J, \infty} = 2^{-n}$$

Der Beweis benutzt das

Lemma 1.3 Sei $\mathbb{Q}_n := \{q_n \in \mathbb{P}_n; \frac{d^n}{dx^n} q_n(x) = n!\}$.

Dann gilt:

$$\min_{q_n \in \mathbb{Q}_n} \|q_n\|_{J, \infty} = \|2^{1-n} T_n\|_{J, \infty} = 2^{1-n} \quad (13)$$

Beweis: (zum obigen Theorem)

$$\omega_{n+1} \in \mathbb{Q}_{n+1} \text{ (da } \frac{d^{n+1}}{dx^{n+1}} \omega_{n+1} = (n+1)!)$$

\Rightarrow Die Behauptung folgt durch Anwendung des vorigen Lemmas. □

Beweis: (zum obigen Lemma)

Aus (9) und (10) $\Rightarrow T_n$ hat den führenden Koeffizienten 2^{n-1}

$$\Rightarrow 2^{1-n} * T_n \in \mathbb{Q}_n$$

$$\Rightarrow \min_{q_n} \|q_n\|_{J, \infty} \leq 2^{1-n}$$

Angenommen $\exists q_n \in \mathbb{Q}_n$ mit $\|q_n\| \leq 2^{1-n}$ ♡

$$\text{Setze } q := q_n - 2^{1-n} T_n$$

$$\Rightarrow q(y_k) = q_n(y_k) - 2^{1-n} T_n(y_k) \begin{cases} < 0 & \text{für gerades } k \quad (T_n(y_k) = 1) \\ > 0 & \text{für ungerades } k \quad (T_n(y_k) = -1) \end{cases}$$

$q \in C(J) \Rightarrow q$ hat in J n Vorzeichenwechsel, das heißt mindestens n Nullstellen,

aber $q \in \mathbb{P}_{n-1} \Rightarrow q \equiv 0 \Rightarrow$ Widerspruch zur Annahme ♡. □

$q \in C \hat{=} q$ stetig

Aus (7) folgt bei Interpolation in den Tschebyscheff-Knoten $x_k = \cos \frac{k+\frac{1}{2}}{n} \pi$, $k = 0, \dots, n$ in J mit $f \in C^{n+1}(J)$:

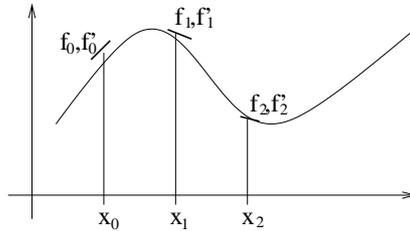
$$\|f - p_n\|_{J, \infty} \leq \frac{1}{2^n (n+1)!} \|f^{(n+1)}\|_{J, \infty} \quad (14)$$

Hermite-Interpolation

$I = [a, b] \in \mathbb{R}$, mit den Daten: $f(x_\nu) =: f_\nu$, $f'(x_\nu) =: f'_\nu$, $\nu = 0, 1, \dots, n$

1 Interpolation durch Polynome

einer glatten Funktion f können wir auf "Informationsbewahrende Art" interpolieren.



Theorem 1.5 Es existiert genau ein Hermite-Interpolationspolynom $h_{2n+1}(x_\nu) \in \mathbb{P}_{2n+1}$ mit $h_{2n+1}(x_\nu) = f_\nu$, $h'_{2n+1}(x_\nu) = f'_\nu$ ($0 \leq \nu \leq n$) für paarweise verschiedene Knoten x_ν , $\nu = 0, 1, \dots, n$

Beweis:

Eindeutigkeit

Annahme: Es gibt 2 Polynome g_{2n+1} , h_{2n+1} mit (15)

$$\Rightarrow \underbrace{h_{2n+1} - g_{2n+1}} \in \mathbb{P}_{2n+1}$$

hat $n + 1$ doppelte Nullstellen, denn

$$h_{2n+1}(x_\nu) = f_\nu = g_{2n+1}(x_\nu)$$

$$\text{und } h'_{2n+1}(x_\nu) = f'_\nu = g'_{2n+1}(x_\nu)$$

$$\Rightarrow h_{2n+1}(x) - g_{2n+1}(x) \equiv 0 \text{ (nach dem Fundamentalsatz der Algebra)}$$

$$\Rightarrow \text{Eindeutigkeit} \quad \checkmark$$

Existenz

$$h_{2n+1}(x) := \sum_{\nu=0}^n f_\nu k_\nu(x) + \sum_{\nu=0}^n f'_\nu m_\nu(x) \in \mathbb{P}_{2n+1}$$

$$\text{mit } k_\nu(x) := [1 - 2 * l'_\nu(x_\nu)(x - x_\nu)] l_\nu^2(x) \qquad l_\nu(x) := \prod_{i=0, i \neq \nu}^n \frac{x - x_i}{x_\nu - x_i} \in \mathbb{P}_n$$

$$m_\nu(x) := (x - x_\nu) \underbrace{l_\nu^2(x)}_{\in \mathbb{P}} \qquad \text{Lagrange Polynom}$$

$$\text{Es gilt: } k_\nu(x_\mu) = \delta_{\mu\nu} = m'_\nu(x_\mu)$$

$$\|k'_\nu(x_\mu) = m_\nu(x_\mu) = 0$$

$$(0 \leq \mu \leq n, 0 \leq \nu \leq n)$$

$$\delta_{\mu\nu} = \begin{cases} 1 & , \mu = \nu \\ 0 & , \mu \neq \nu \end{cases} \quad (\text{da } l_\nu(x_\mu) = \delta_{\mu\nu})$$

$$h_{2n+1}(x_k) \stackrel{(16)}{=} \underbrace{\sum_{\nu=0}^n f_\nu k_\nu(x_k)}_{f_k \underbrace{k_k(x_k)}_1} + \underbrace{\sum_{\nu=0}^n f'_\nu m_\nu(x_k)}_{=0} = f_k \quad \checkmark$$

Nachweis von (15)

$$h'_{2n+1}(x_\mu) \stackrel{(16)}{=} \sum_{\nu=0}^n f_\nu \underbrace{k'_\nu(x_\nu)}_{=0} + \underbrace{\sum_{\nu=0}^n f'_\nu m'_\nu(x_\mu)}_{f_\mu * 1} = f'_\mu \quad \checkmark \quad \square$$

1 Interpolation durch Polynome

Beispiel: $f(x) := \sin(x)$ $x_0 = 0$, $x_1 = \frac{\pi}{2}$

Finde zugehöriges Hermite-Interpolations-Polynom

x_ν	0	$\frac{\pi}{2}$
$f_\nu = \sin(x_\nu)$	0	1
$f'(x_\nu) = \cos(x_\nu) = f'_\nu$	1	0

$$\nu = 0, 1 = n$$

$$h_3(x) \stackrel{(16)}{=} \sum_{\nu=0}^1 f_\nu k_\nu(x) + \sum_{\nu=0}^1 f'_\nu m_\nu(x) = f_1 k_1(x) + f'_0 m_0(x)$$

$$\begin{aligned} k_1(x) &= [1 - 2l'_1(x_1)(x - x_1)]l_1^2(x) \\ &= [1 - 2\frac{2}{\pi}(x - \frac{\pi}{2})] \left(\frac{x-0}{\frac{\pi}{2}-0}\right)^2 \\ &= \frac{4}{\pi^2}x^2(3 - \frac{4}{\pi}x) \end{aligned}$$

$$l_1(x) = \frac{x-x_0}{x_1-x_0} = \left(\frac{x-0}{\frac{\pi}{2}-0}\right)$$

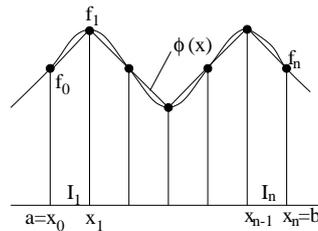
$$\begin{aligned} m_0(x) &= (x - x_0)l_0^2(x) \\ &= (x - 0) \left(\frac{x-\frac{\pi}{2}}{0-\frac{\pi}{2}}\right)^2 = \frac{4}{\pi^2}x(x - \frac{\pi}{2})^2 \end{aligned}$$

$$l_0(x) = \frac{x-x_1}{x_0-x_1} = \left(\frac{x-\frac{\pi}{2}}{0-\frac{\pi}{2}}\right)$$

$$h_3 = 1 * k_1(x) + 1 * m_0(x) = \frac{4}{\pi^2}x[3x - \frac{4}{\pi}x^2 + (x - \frac{\pi}{2})^2]$$

Stückweise polynomiale Interpolation Zerlegung von $I = [a, b]$ in (der Einfachheit halber) äquidistanter Teilintervalle

$I_i := [x_{i-1}, x_i]$, Knoten: $x_i = a + ih$ ($0 \leq i \leq n$), $h := \frac{b-a}{n}$



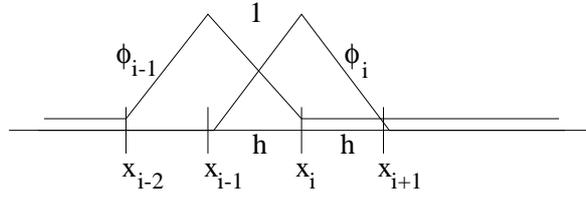
Lineare Interpolation auf jedem I_ν : Die Lösung für f_0, f_1, \dots, f_n kann sofort angegeben werden mit Hilfe sogenannter Hut-Funktionen ϕ_i .

Stückweise linear interpolierende:

$$\phi(x) := \sum_{i=0}^n f_i \phi_i(x) \text{ mit } \phi_i(x) = \begin{cases} \frac{x-x_{i-1}}{h} & ; x \in I_i = (x_{i-1}, x_i) \\ \frac{x_{i+1}-x}{h} & ; x \in I_{i+1} = (x_i, x_{i+1}) \\ 0 & ; \text{sonst} \end{cases} \quad (18)$$

Hutfunktion (Basisfunktion)

1 Interpolation durch Polynome



$$\begin{aligned} \phi|_{I_k} &= f_{k-1}\phi_{k-1}(x) + f_k\phi_k(x)|_{x \in I_k} \\ \phi(x_{k-1}) &= f_{k-1} \underbrace{\phi_{k-1}(x_{k-1})}_{=1} + f_k \underbrace{\phi_k(x_{k-1})}_{=0} = f_{k-1} \\ \phi(x_k) &= f_{k-1} \underbrace{\phi_{k-1}(x_k)}_{=0} + f_k \underbrace{\phi_k(x_k)}_{=1} = f_k \end{aligned}$$

Theorem 1.6 (Abschätzung)

Sei $f \in C^2(I)$ und $\phi \in \mathbb{P}_n$ wie in (18), x_i äquidistant, $f_i := f(x_i)$, dann gilt:

1. $\|f - \phi\|_{I,\infty} = \max_i \|f - \phi\|_{I_i,\infty} \leq \frac{h^2}{8} \|f''\|_{I,\infty}$
2. $\|f' - \phi'\|_{I,\infty} \leq \frac{h}{2} \|f''\|_{I,\infty}$

$f \in C^{n+1}(I)$, $p_n \in \mathbb{P}_n$
Interpolierbar

$$\|f - p_n\|_{I_i,\infty} \leq \frac{1}{(n+1)!} * \|\omega\|_{I_i,\infty} * \|f^{(n+1)}\|_{\infty, I_i}$$

(siehe Seite 14)

Beweis:

$$\begin{aligned} 1. \quad \|f - \phi\|_{I_i,\infty} &\leq \frac{1}{2!} \|I_i,\infty\| \overbrace{\|(x - x_i)(x - x_{i-1})\|_{I_i,\infty}}^{\omega} \\ 0 &\stackrel{!}{=} \frac{d\omega}{dx} = 2x - x_{i-1} - x_i \\ \Rightarrow x^* &= \frac{x_{i-1} + x_i}{2} \\ \omega(x^*) &= \frac{(x_{i-1} - x_i)}{2} * \frac{(x_i - x_{i-1})}{2} = \frac{h^2}{4} = \|\omega\|_{I_i,\infty} \\ \Rightarrow \|f - \phi\|_{I_i,\infty} &\leq \frac{h^2}{8} \|f''\|_{I_i,\infty} \\ \Rightarrow \max_{i=0,\dots,n} \|f - \phi\|_{I_i,\infty} &\leq \frac{h^2}{8} \underbrace{\max_{x \in I} \|f''(x)\|_{I_i,\infty}}_{\max_{x \in I} |f''(x)| = \|f''\|_{I,\infty}} \end{aligned}$$

2. In I_i gilt ($x \in I_i$):

$$\begin{aligned} h[f'(x) - \phi'(x)] &= h[f'(x) - \sum_{i=0}^n f_i \phi'_i(x)] \\ &= h[f'(x) - f_{i-1} \underbrace{\phi'_{i-1}(x)}_{-\frac{1}{h}} - f_i \underbrace{\phi'_i(x)}_{\frac{1}{h}}] \\ &= h * f'(x) - h * \frac{f_i - f_{i-1}}{h} \\ &= h f'(x) - \underbrace{\int_{x_{i-1}}^{x_i} f'(t) dt}_{f(x_i) - f(x_{i-1})} \\ &= \underbrace{(x_i - x_{i-1})}_{h} f'(x) - \underbrace{x_i f'(x_i) + x_{i-1} f'(x_{i-1}) + t f'(t)|_{t=x_{i-1}}^{x_i}}_{=0} - \int_{x_{i-1}}^{x_i} f'(t) dt \end{aligned}$$

1 Interpolation durch Polynome

$$\begin{aligned}
 &= \underbrace{(x_i - x_{i-1})f'(x) - x_i f'(x_i) + x_{i-1} f'(x_{i-1}) + t f'(t)|_{t=x_{i-1}}^{x_i}}_{-x_i \int_x^{x_i} f''(t) dt - x_{i-1} \int_{x_{i-1}}^x f''(t) dt} - \int_{x_{i-1}}^{x_i} f'(t) dt \\
 &= (x_i - x_{i-1})f'(x) - x_i f'(x_i) + x_{i-1} f'(x_{i-1}) + \underbrace{t f'(t)|_{t=x_{i-1}}^{x_i} - \int_{x_{i-1}}^{x_i} f'(t) dt}_{+ \int_{x_{i-1}}^{x_i} t f''(t) dt} \\
 &= h[f'(x) - \phi'(x)] \\
 &= \int_x^{x_i} (t - x_i) f''(t) dt + \int_{x_{i-1}}^x (t - x_{i-1}) f''(t) dt \\
 h|f'(x) - \phi'(x)| &\leq \|f''\|_{\infty, I_i} \int_x^{x_i} (x_i - t) dt + \|f''\|_{I_i, \infty} \left\{ \frac{1}{2}((x - x_{i-1})^2 + (x_i - x)^2) \right\} \\
 &\leq \frac{h^2}{2} \|f''\|_{I_i, \infty} \\
 &\text{dividieren durch } h \Rightarrow \text{Lösung.} \quad \square
 \end{aligned}$$

Interpolation mit kubischen Splines

Definition 1.1 Sei $I = [a, b]$, $f \in C^1[I]$, $x_i = a + ih$, $I_i := [x_{i-1}, x_i]$
 $\psi : I \rightarrow \mathbb{R}$ heißt {vollständiger} kubischer Spline zu f : \Leftrightarrow

1. $\psi \in C^2(I)$
2. $\psi \in \mathbb{P}_3(I_i)$, $i = 1, \dots, n$
3. $\psi(x_i) = f(x_i) =: f_i$ ($0 \leq i \leq n$)
4. $\{\psi'(a) = f'(a), \psi'(b) = f'(b)\}$

Bedingung (4) sichert die Eindeutigkeit.

Statt (4): $\psi''(a) = \psi''(b) = 0$ natürlicher Spline. (mit weniger guten Approximationseigenschaften, da er weniger Informationen über f hat.)

$$\Psi(x)|_{I_i} \stackrel{(2)}{=} a_i + b_i x + c_i x^2 + d_i x^3$$

Bedingungen:

$$\Psi \in C^2(I) \text{ wegen (1)}$$

$$\Psi(x_i) = f_i \text{ wegen (3)}$$

$$\Psi'(a) = f'(a), \Psi'(b) = f'(b), \text{ wegen (4)}$$

\Rightarrow insgesamt: $4n$ Unbekannte (Freiheitsgrad)

$$\text{Bedingungen: } 3(n-1) + (n+1) + 2 = 4n$$

Theorem 1.7 Seien x_i, f_i wie in der Definition von Seite 20.

Dann existiert genau ein vollständiger kubischer Spline Ψ zu f .

“Ich will Sie jetzt nicht mit dem Beweis dieses Satzes quälen !”

Theorem 1.8 Sei $f \in C^4(I)$, $I = [a, b]$ $x_i = a + i * h$, $h = |I_i|$, dann gilt:

Es existiert genau eine Konstante $c > 0$, so daß

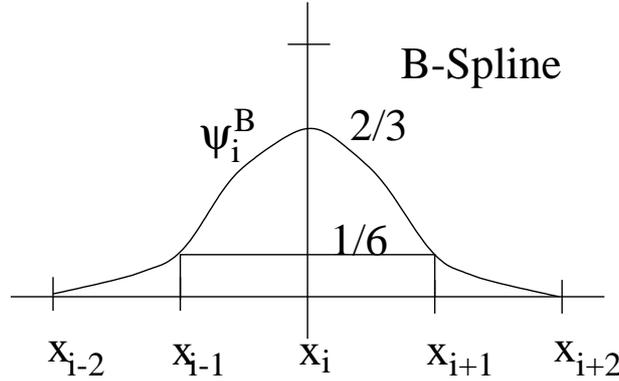
$$\|f^{(i)} - \Psi^{(i)}\|_{I, \infty} \leq c * h^{4-i} \quad (i = 0, 1, 2, 3)$$

1 Interpolation durch Polynome

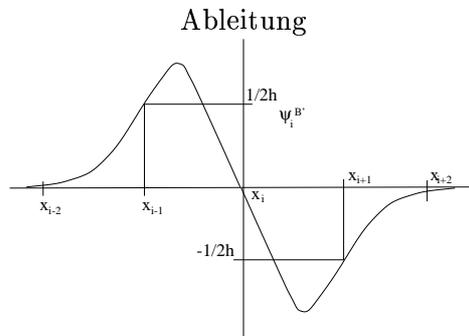
Fortgesetztes Gitter $x_i = a + i * h \ i \in \mathbb{Z}, I_i = [x_{i-1}, x_i]$

$S_n^1 := \{ \phi \in C(\mathbb{R}), \phi|_{I_i} \text{ linear } \forall i \in \mathbb{Z} \}$
 (Raum der stückweise linear stetigen Funktionen.)

$S_n^3 := \{ \Psi \in C^2(\mathbb{R}), \Psi|_{I_i} \in \mathbb{P}_3, \forall i \in \mathbb{Z} \}$
 (Raum der Spline-Funktionen.)



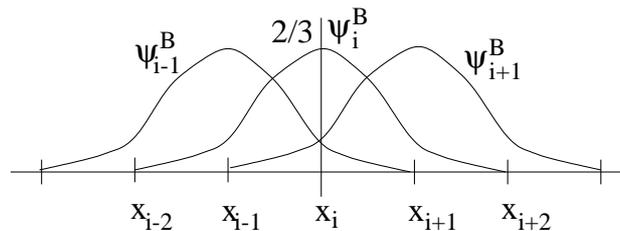
$$\Psi_i^B(x) = \begin{cases} 0 & , x \leq x_{i-2} \\ \frac{(x-x_{i-2})^3}{6h^3} & , I_{i-1} = (x_{i-2}, x_{i-1}) \\ \frac{(x_i-x)^3}{2h^3} - \frac{(x_i-x)^2}{h^2} + \frac{2}{3} & , I_i = (x_{i-1}, x_i) \\ \frac{(x-x_i)^3}{2h^3} - \frac{(x-x_i)^2}{h^2} + \frac{2}{3} & , I_{i+1} = (x_i, x_{i+1}) \\ \frac{(x_{i+2}-x)^3}{6h^3} & , I_{i+2} = (x_{i+1}, x_{i+2}) \\ 0 & , x \geq x_{i+2} \end{cases}$$



Lemma 1.4

Sei $\Psi \in S_n^3(I)$. Dann existiert genau ein $A \in \mathbb{R}^{n+3}$, $A^T = (A_{-1}, A_0, \dots, A_n, A_{n+1})$ mit

$$\Psi(x) = \sum_{\nu=-1}^{n+1} A_\nu \Psi_\nu^B(x)|_{I=[x_0, \dots, x_n]}$$



1 Interpolation durch Polynome

Beispiel:

$f(x) = \sin(x) - 2 * \sin^3(x)$ in $I = [0, \pi]$.

Finde den kubischen Spline Ψ für f mit dem Knoten $x_k = \frac{k*\pi}{4}$, $k = 0, 1, 2, 3, 4$, als Linearkombination geeigneter B-Splines.

$$\begin{aligned} & x_0, \dots, x_4 \\ \Psi(x_\nu) &= f(x_\nu) \\ \Psi'(x_0) &= f'(x_0) \\ \Psi'(x_4) &= f'(x_4) \\ & h = \frac{\pi}{4} \end{aligned}$$

$$\begin{array}{ccccccc} f'(0) & f(0) & f(\frac{\pi}{4}) & f(\frac{\pi}{2}) & f(\frac{3}{4}\pi) & f(\pi) & f'(\pi) \\ 1 & 0 & 0 & -1 & 0 & 0 & -1 \end{array}$$

$$\frac{1}{6} \begin{pmatrix} 4 & 2 & & & & & \\ 1 & 4 & 1 & & & & \\ & 1 & 4 & 1 & & & \\ & & 1 & 4 & 1 & & \\ & & & 2 & 4 & & \end{pmatrix} \begin{pmatrix} A_0 \\ A_1 \\ A_2 \\ A_3 \\ A_4 \end{pmatrix} = \begin{pmatrix} f(0) + \frac{h}{3}f'(0) \\ f(\frac{\pi}{4}) \\ f(\frac{\pi}{2}) \\ f(\frac{3}{4}\pi) \\ f(\pi) - \frac{h}{3}f'(\pi) \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{\pi}{12} \\ 0 \\ -1 \\ 0 \\ \frac{\pi}{12} \end{pmatrix}$$

$$\left. \begin{array}{l} A_{-1} = A_1 - 2hf'(0) \\ A_{n+1} = A_{n-1} + 2hf'(\pi) \end{array} \right\} \Rightarrow \begin{array}{l} A_{-1} = \frac{1}{2} - \frac{\pi}{24} - \frac{\pi}{2} \\ A_5 = \frac{1}{2} - \frac{\pi}{24} - \frac{\pi}{2} \end{array}$$

$$\Rightarrow A_1 = A_3 = \frac{1}{2} - \frac{\pi}{24}, A_0 = A_4 = -\frac{1}{4} + \frac{7}{48}\pi, A_2 = \frac{\pi}{48} - \frac{7}{4}$$

$$\Psi(x) = \sum_{\nu=-1}^5 A_\nu \Psi_\nu^B(x)$$

2 Lineare Gleichungssysteme, direkte Verfahren

2.1 Vorbemerkung

$$A = (a_{ik})_{\substack{i=1,\dots,m \\ k=1,\dots,n}} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m1} & \cdots & a_{mn} \end{pmatrix} \text{ mit } a_{ik} \in \mathbb{R}$$

das heißt $A \in \mathbb{R}^{m \times n}$ bzw. A ist (reelle) (m, n) -Matrix.
(eventuell auch $a_{ik} \in \mathbb{C}$, d.h. $A \in \mathbb{C}^{m \times n}$).

$$\vec{r} = (r_i)_{i=1,\dots,m} = \begin{pmatrix} r_1 \\ \cdots \\ r_m \end{pmatrix} \in \mathbb{R}^m$$

Lineares Gleichungssystem

$$A\vec{x} = \vec{r}, \text{ d.h. } \vec{x} = \begin{pmatrix} x_1 \\ \cdots \\ x_n \end{pmatrix} \in \mathbb{R}^n \quad (2.1)$$

gesucht, bzw.

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= r_1 \\ &\vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= r_m \end{aligned} \quad (2.2)$$

Zugehöriges homogenes System:

$$A\vec{x} = \vec{0} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} \in \mathbb{R}^m \quad (2.3)$$

Lineare Algebra:

Hat (2.3) genau ρ linear unabhängige Lösungen (z.B.) $\vec{x}^{(1)}, \dots, \vec{x}^{(\rho)} \in \mathbb{R}^n$ und existiert eine spezielle Lösung $\vec{x}^{(0)}$ von (2.1), so ergibt sich jede Lösung von (2.1) in der Form $\vec{x}^{(0)} + c_1\vec{x}^{(1)} + \cdots + c_\rho\vec{x}^{(\rho)}$ ($c_i \in \mathbb{R}$ beliebig)

Speziell $m = n$, d.h. $A \in \mathbb{R}^{n \times n}$ quadratische Matrix, $D := \det(A)$ (auch $|A|$ üblich).

$D \neq 0 \Leftrightarrow$ (2.3) hat nur die (sogenannte) triviale Lösung $\vec{x} = \vec{0} \Leftrightarrow$ (2.1) ist für jedes $r \in \mathbb{R}^m$ eindeutig lösbar.

$D = 0 \Leftrightarrow$ (2.3) hat $(0 <) \rho \leq n$ linear unabhängige Lösungen und (2.1) hat ∞ viele Lösungen oder ist unlösbar.

$D \neq 0$: $\vec{x} = A^{-1}\vec{r}$ in geschlossener Form darstellbar.

Cramersche Regel:

$$x_k \frac{D_k}{D}$$

D_k aus D durch Ersetzen der k -ten Spalte von D durch \vec{r} ; für Praxis ungünstig.

2.2 Gauß'scher Algorithmus

Zunächst betrachten wir ein LGS (2.2) mit $m = n$, $\det A \neq 0$:

Sei $a_{11} \neq 0$: Neue i -te Zeile = alte i -te Zeile $-\frac{a_{i1}}{a_{11}}$ ($i = 2, \dots, n$).

$$\left. \begin{array}{l} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = r_1 \\ 0 + a'_{22}x_2 + \dots + a'_{2n}x_n = r'_2 \\ \vdots \\ 0 + a'_{n2}x_2 + \dots + a'_{nn}x_n = r'_n \end{array} \right\} \quad (2.4)$$

Verfahren auf $(n-1) \times (n-1)$ LGS fortsetzen. ($a'_{22} \neq 0$, sonst eventuell Zeilenvertauschung.)

Schließlich LGS mit Δ -Matrix

$$\left. \begin{array}{l} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = r_1 \\ a'_{22}x_2 + \dots + a'_{2n}x_n = r'_2 \\ \vdots \\ a_{nn}^{(n-1)}x_n = r_n^{(n-1)} \end{array} \right\} \quad (2.4)'$$

die x_i in der Reihenfolge x_n, x_{n-1}, \dots, x_1 rückwärts bestimmen.

Bemerkung:

Der Gauß-Algorithmus hat in dieser Form $\cong \frac{n^3}{3}$ Operationen. Die Cramersche Regel mit Entwicklung der Determinanten D, D_1, \dots, D_n benötigt $\cong \frac{n^4}{3}$ Operationen.

Methoden zur Rechenstabilität

Pivotwahl

Elemente von A haben bereits gleiche Größenordnung, sei zum

Beispiel $\sigma < |a_{11}| \ll |a_{i1}|$ ($i = 2, \dots, n$)

Gauß in der Form (2.4)/(2.4)' möglich. (sogenannte "natürliche Pivotwahl", aber große Fehlerfortpflanzung bei Division), daher

1. Partielle Pivotwahl

In der 1. Spalte von A (bzw. in den weiteren Schritten im jeweiligen "Rest"-LGS) wird das betragsgrößte Element ("Pivotelement") gesucht und (jeweils) die 1. Zeile mit der Zeile des Pivotelements vertauscht.

2. Totale Pivotwahl

In A (bzw. den jeweiligen "Restmatrizen") wird das betragsgrößte Element gesucht und die jeweilige erste Spalte und 1. Zeile mit Spalte und Zeile des Pivotelements vertauscht. (bei Spaltenvertauschung auf Gesamtmatrix ausdehnen).

Bemerkung:

(2) im allgemeinen stabiler, aber aufwendiger, daher meistens nur (1).

weitere Bemerkungen zu Gauß

1. Berechnung der Determinante $\det A \neq 0$, 2 Operationen bei Gauß

2 Lineare Gleichungssysteme, direkte Verfahren

- a) Addition bzw. Subtraktion von Vielfachen von Zeilen und anderen Zeilen \Rightarrow Determinante bleibt erhalten
- b) Zeilen- und (eventuell) Spaltenvertauschungen \Rightarrow je Vertauschung Änderung von $\det A$ um Faktor -1

Aus (2.4)' folgt mit zusätzlichen Vertauschungen:

$$\det A = (-1)^\delta a_{11} a'_{22} \cdots a_{nn}^{(n-1)} \quad \delta = \text{Gesamtzahl der Vertauschungen} \quad (2.5)$$

2. Mehrere rechte Seiten, Inversenbestimmung

Schema für Gauß

$$\begin{array}{ccc|c|c} a_{11} & \cdots & a_{1n} & r_1 & s_1 \\ \vdots & & \vdots & \vdots & \vdots \\ a_{n1} & \cdots & a_{nn} & r_n & s_n \end{array}$$

Zum Beispiel mit partieller Pivotwahl führt der Algorithmus über (2.4) schließlich auf

$$\begin{array}{ccc|c|c|c} a_{11} & \cdots & \cdots & a_{1n} & r'_1 & s_1 \\ & a_{22} & & a'_{2n} & r'_2 & s'_2 \\ & & \ddots & \vdots & \vdots & \vdots \\ & & & a_{nn}^{(n-1)} & r_n^{(n-1)} & s_n^{(n-1)} \end{array}$$

Man kann auch eine neue rechte Seite \vec{s} einfügen ohne die Dreieckszerlegung von A nochmals durchzuführen.

Beispiel: Bestimmung von $A^{-1}(= x)$ mittels Matrizen bzw. LGS $A * x = E$, d.h.

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nn} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

3. Die LR-Zerlegung

Satz 2.1 Jede reguläre Matrix A besitzt eine verallgemeinernde LR-Zerlegung, das heißt zu A existiert eine Permutationsmatrix, eine untere Dreiecksmatrix L und eine obere Dreiecksmatrix R so, daß gilt $PA = LR$.

Vergleich:(Gauß-Algorithmus)

$(a_{11} \neq 0) \Rightarrow$ neue i-te Zeile = alte i-te Zeile $- \frac{a_{i1}}{a_{11}}$ (1. Zeile)

$$\begin{array}{cccccc} a_{11}x_1 + & a_{12}x_2 + & \cdots & + a_{1n}x_n & = & r_1 \\ \vdots & & & & & \vdots \\ a_{n1}x_1 + & \cdots & & + a_{nn}x_n & = & r_n \end{array}$$

2 Lineare Gleichungssysteme, direkte Verfahren

5. Der Fall $\det A = 0$, rechteckige Matrizen, Rangbestimmung
 Gauß entsprechend durchführen, eventuell totale Pivotwahl bis nur Nullzeilen übrigbleiben.

Beispiel:

$$A = \begin{pmatrix} 1 & 1 & 2 & 0 \\ 1 & 2 & 3 & 6 \\ 2 & 3 & 5 & 6 \end{pmatrix}, \text{Rang } A = \text{maximale Zahl linear}$$

unabhängiger Zeilen/ Spalten, hier: $\text{Rang}(A) = 2$
 Rang (A) bleibt gegen Gauß-Operationen invariant.

6. LGS mit positiv definiten Matrizen
 $A \in \mathbb{R}^{n \times n}$ sei symmetrisch, d.h. $\forall i, k \ a_{ik} = a_{ki} \Rightarrow A = A^t$ wichtiger Spezialfall für Anwendungen.

Definition:

Die symmetrische Matrix A ist positiv (semi) definit:

$$\Leftrightarrow \forall \vec{x} \neq \vec{0} \quad \vec{x}^t (A\vec{x}) > (\geq) 0$$

Für positiv definite Matrizen ist die Hauptdiagonale positiv:

$$a_{ii} = \vec{e}_i^t * A * \vec{e}_i > 0, \quad \vec{e}_i := \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}$$

Verallgemeinerung (Kriterium von Schur), für Anwendungen allerdings unhandlich:

$$A \text{ ist genau dann positiv definit, wenn } a_{ii} > 0, \quad \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} > 0,$$

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ \vdots & & \vdots \\ \vdots & \dots & a_{33} \end{vmatrix} > 0, \dots, \det(A) > 0 \text{ gilt.}$$

$A\vec{x} = \vec{r}$ mit A positiv definit sei gegeben.

L-R Zerlegung mit $P = E$ ist hier möglich: $A = L * R$, aber Symmetrie würde zerstört.
 Stattdessen Zerlegung:

$$A = S^t * S \text{ mit } S = \begin{pmatrix} S_{11} & \dots & S_{1n} \\ \vdots & \ddots & \vdots \\ 0 & \dots & S_{nn} \end{pmatrix}, \quad S_{ii} \ (i = 1, \dots, n) \quad (2.9)$$

angestrebt.

2 Lineare Gleichungssysteme, direkte Verfahren

(2.9) heißt CHOLESKY-Zerlegung.

1. Gauß -Schritt:

$$C_1^{-1}A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & * & * & * \\ \vdots & * & * & * \\ 0 & * & * & * \end{pmatrix}$$

von rechts mit $(C_1^{-1})^t$ multiplizieren.

$$a_{ik} = a_{ki} \Rightarrow \underbrace{C_1^{-1}A(C_1^{-1})^t}_B = \begin{pmatrix} a_{11} & 0 & \cdots & a_{1n} \\ 0 & * & * & * \\ \vdots & & A^{(i)} & \\ 0 & & & \end{pmatrix} \quad B = B^t, \text{ da } A = A^t.$$

B und $A^{(i)}$ sind wieder positiv definit:

$\vec{x} \neq \vec{0}$, $\vec{x}^t B \vec{x} = \vec{y}^t A \vec{y}$ mit $\vec{y} = (C_1^{-1})^t * \vec{x} \neq \vec{0}$, da $\vec{x} \neq \vec{0}$ gilt und $\det(C_1^{-1})^t = 1$.

A positiv definit $\Rightarrow \vec{y}^t A \vec{y} > 0$.

Entsprechend muß man das Verfahren für $A^{(i)}$ (1. Komponente weglassen) fortsetzen

$$\Rightarrow \cdots C_2^{-1} C_1^{-1} A (C_1^{-1})^t (C_2^{-1})^t \cdots = D = \begin{pmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_n \end{pmatrix} \quad d_i > 0$$

$$\Rightarrow C := C_1 * C_2 * \cdots * C_{n-1} \quad C^{-1} A (C^{-1})^t = D$$

$$\Rightarrow A = C * D * C^t$$

$$\text{Setzt man } D^{\frac{1}{2}} := \begin{pmatrix} \sqrt{d_1} & & 0 \\ & \ddots & \\ 0 & & \sqrt{d_n} \end{pmatrix}, \text{ so ist}$$

$$A = C * D^{\frac{1}{2}} * D^{\frac{1}{2}} * C^t = (C D^{\frac{1}{2}}) \underbrace{(C D^{\frac{1}{2}})^t}_{=: S} \quad (2.10)$$

CHOLESKY-Zerlegung von A

Ist umgekehrt $A = S^t * S$ Cholesky-Zerlegung von $A (= A^t)$, so ist A positiv definit:

$$\begin{aligned} \vec{x}^t A \vec{x} = \vec{x}^t * S^t \underbrace{S \vec{x}}_{=: \vec{y}} &= \vec{y}^t \vec{y} \geq 0 \quad \Rightarrow \vec{x}^t A \vec{x} > 0 \forall \vec{x} \neq \vec{0} \\ &= 0 \quad \Leftrightarrow \vec{y} = \vec{0} \Rightarrow \vec{x} = S^{-1} \vec{y} = \vec{0} \end{aligned}$$

Satz 2.2 Die symmetrische Matrix A ist genau dann positiv definit, wenn sie eine Cholesky-Zerlegung hat.

Formeln der Cholesky-Zerlegung ergeben sich über (2.10) oder direkt:

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} \stackrel{!}{=} \begin{pmatrix} s_{11} & 0 & 0 & 0 \\ s_{12} & s_{22} & 0 & 0 \\ \vdots & & \ddots & 0 \\ s_{1n} & \cdots & \cdots & s_{nn} \end{pmatrix} \begin{pmatrix} s_{11} & \cdots & s_{1n} \\ & \ddots & \vdots \\ 0 & & s_{nn} \end{pmatrix}$$

$$(0 <) a_{11} = s_{11}^2 \Rightarrow s_{11} = +\sqrt{a_{11}} > 0$$

$$a_{12} = s_{12} * s_{11} \Rightarrow s_{12} = \frac{a_{12}}{s_{11}} \text{ u.s.f.}$$

$$s_{ii}^2 = a_{ii} - s_{1i}^2 - \cdots - s_{i-1,i}^2 \quad (i = 1, \cdots, n)$$

$$s_{ik} = (a_{ik} - s_{1i}s_{1k} - \cdots - s_{i-1,i}s_{i-1,k})|_{s_{ii}} \quad (k > i, i = 1, \cdots, n-1)$$

2 Lineare Gleichungssysteme, direkte Verfahren

Anmerkungen des Autors:

Hier entstand in der Vorlesung ein kleines Indizee-Wirrwarr. Sämtliche Indizees in den S-Matrizen sind verdreht und die Formeln beinahe darauf abgestimmt.

Vermutlich sind die Formeln richtig, wenn man in der a-Matrix beim Benutzen der Formeln alle Indizees vertauscht ($a_{n1} \rightarrow a_{1n}$, etc.),

Übertragung auf rechte Seite des LGS:

$A\vec{x} = \vec{r}$ soll äquivalent zu $S\vec{x} = \vec{\rho} = \begin{pmatrix} \rho_1 \\ \vdots \\ \rho_n \end{pmatrix}$ sein.

$$S^t|_{S\vec{x}=\vec{\rho}} \Rightarrow \underbrace{S^t S}_A \vec{x} = S^t \vec{\rho} = \vec{r}$$

$$\Rightarrow \rho_i = (r_i - s_{1i}\rho_1 - \cdots - s_{i-1,i}\rho_{i-1})|_{s_{ii}} \quad (2.11b)$$

(wie s_{ik} ($k > i$) ohne Index k)

dann x_i aus $S\vec{x} = \vec{\rho}$ rekursiv berechnen wie in b).

Bemerkung: zu Cholesky:

1. i.a. gute numerische Stabilität des Verfahrens, unter anderem auch wegen $\sqrt{\cdots}$ bei s_{ii}
2. Pivotwahl i.a. nicht nötig, könnte (wegen Symmetrieerhaltung) auch nur auf Diagonale durchgeführt werden (Zeilen und gleiche Spalten vertauschen).
3. Arbeitsaufwand \approx halb so groß wie bei Gauß .

3 Vektor- und Matrixnormen, Fehlerabschätzungen für LGS, iterative Verfahren

3.1 Vektor- und Matrixnormen

\mathbb{R} ist mit der Euklidischen Vektor-Norm

$$\|\vec{x}\|_2 := \sqrt{\sum_{i=1}^n x_i^2} \text{ für } \vec{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$$

normiert (oft auch $|\vec{x}|$), das heißt dadurch wird ein Abstandsbegriff $\|\vec{x} - \vec{y}\|_2$ gegeben. Allgemeine Definition der Norm auf \mathbb{R}^n :

Die Funktion $\|\cdot\| : \mathbb{R}^n \rightarrow \mathbb{R}$ heißt Norm auf \mathbb{R}^n , wenn gilt:

- (N1) $\forall \vec{x} \in \mathbb{R}^n: \|\vec{x}\| \geq 0$ und $\|\vec{x}\| = 0 \Leftrightarrow \vec{x} = \vec{0}$ (Definitheit)
- (N2) $\forall \vec{x} \in \mathbb{R}^n \forall \alpha \in \mathbb{R}: \|\alpha \vec{x}\| = |\alpha| * \|\vec{x}\|$ (positive Homogenität)
- (N3) $\forall \vec{x}, \vec{y} \in \mathbb{R}^n: \|\vec{x} + \vec{y}\| \leq \|\vec{x}\| + \|\vec{y}\|$ (Dreiecksungleichung)

Beispiel:

$$\|\vec{x}\|_p = (\sum_{i=1}^n |x_i|^p)^{\frac{1}{p}} \quad (p \geq 1 \text{ reel}) \tag{3.1}$$

wichtigste Fälle:

- $p = 1$: $\|\vec{x}\|_1 = \sum_{i=1}^n |x_i|$
- $p = 2$: Euklidische Norm, s.o.

$$\|\vec{x}\|_\infty := \max_{i=1}^n |x_i| \tag{3.2}$$

($p \rightarrow \infty$ in (3.1) ergibt (3.2))

Für die numerische Mathematik ist (3.2) wichtiger als $\|\cdot\|_2$, aber z.B. für Konvergenzaussagen werden auch andere Normen betrachtet.

Für (3.1) erfolgt der Nachweis von (N1),(N2),(N3) über die Höldersche Ungleichung. Diese entspricht im Fall $p = 2$ der Cauchy-Schwarzschen Ungleichung.

$$\|\vec{x} \top \vec{y}\|_2 \leq \|\vec{x}\|_2 * \|\vec{y}\|_2.$$

Für (3.2) ist (N1)-(N3) trivial.

Beispiel: LGS $A\vec{x} = \vec{r}$

Lösung näherungsweise (Rundungsfehler) $\tilde{\vec{x}}$.

“ $\tilde{\vec{x}}$ ist mit einem Fehler $\leq \epsilon$ ($\epsilon \in \mathbb{R}, > 0$) behaftet”:

$\|\vec{x} - \tilde{\vec{x}}\|_\infty \leq \epsilon$ heißt jede Komponente hat einen Fehler vom

Betrag $\leq \epsilon$ ($\forall i : (x_i - \tilde{x}_i) \leq \epsilon$)

Bei $\|\vec{x} - \tilde{\vec{x}}\|_p \leq \epsilon$ “Mischen” der Einzelfehler.

$\|\vec{x} - \tilde{\vec{x}}\|_\infty \leq \|\vec{x} - \tilde{\vec{x}}\|_p$, aber dadurch im Allgemeinen Überschätzung der Einzelfehler.

3 Vektor- und Matrixnormen, Fehlerabschätzungen für LGS, iterative Verfahren

Konvergenz von Folgen in \mathbb{R}^n

Eine Folge $(\vec{x}^k)_{k \in \mathbb{N}}$ konvergiert Koordinatenweise gegen $\vec{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$, wenn gilt:

$$\forall x_i^{(k)} \rightarrow x_i$$

Konvergenz bezüglich $\|\cdot\|$

Zu jedem $\epsilon > 0$ existiert ein $M \in \mathbb{N}$ so, daß für alle $k \geq M$ $\|\vec{x} - \vec{x}^{(k)}\| < \epsilon$

Es gilt: Die Konvergenz bezüglich Vektor-Norm und koordinatenweise Konvergenz sind (im \mathbb{R}^n !) gleichwertig, d.h. es wird keine neue Geometrie definiert.

Der Allgemeine Beweis ist etwas aufwendig, für die Gleichwertigkeit der Konvergenz bezüglich (3.1), (3.2) folgt aus

$$\|\vec{x}\|_\infty \leq \|\vec{x}\|_p \leq n^{\frac{1}{p}} \|\vec{x}\|_\infty \quad \forall \vec{x} \in \mathbb{R}^n \quad (*)$$

$$\vec{x} = \vec{0} \quad \checkmark$$

$$\vec{x} \neq \vec{0}: \|\vec{x}\|_\infty = |x_i| \neq 0 \text{ ifest.}$$

$$\Rightarrow \|\vec{x}\|_p = |x_i| \left(\underbrace{\sum_{j=1}^n \left(\frac{|x_j|}{|x_i|} \right)^p}_{\leq 1} \right)^{\frac{1}{p}} \leq n^{\frac{1}{p}} \underbrace{|x_i|}_{\|\vec{x}\|_\infty}$$

$$\Rightarrow (*)$$

Bemerkung:

Für die V-Norm gilt das Entsprechende in \mathbb{C}^n ((N2) dann mit $\alpha \in \mathbb{C}$ (aber auch $\|\underbrace{\vec{x}}_{\in \mathbb{C}}\| \in \mathbb{R}^+$))

Matrix-Normen

$A \in \mathbb{R}^{n \times n}$ (auch entsprechend auf $\mathbb{C}^{n \times n}$, insbesondere im Zusammenhang mit Eigenwertaufgaben auch bei reellen Problemen).

Die Matrix-Norm ist Vektornorm auf \mathbb{R}^{n^2} , aber zusätzlich gilt:

$$\forall A, B \in \mathbb{R}^{n \times n} (\in \mathbb{C}^{n \times n}): \|A * B\| \leq \|A\| * \|B\|$$

Beispiel:

1. $\|\cdot\|_\infty$ auf $\mathbb{R}^{n \times n}$ übertragen ist keine M -Norm, denn

$$\underbrace{\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}}_{\|A\|_\infty=1} \underbrace{\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}}_{\|A^2\|_\infty=2} = \underbrace{\begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix}}_{\|A^2\|_\infty=2} \Rightarrow (3.3) \text{ gilt nicht.}$$

2. $\|\cdot\|_F$ auf $\mathbb{R}^{n \times n}$ ist sogenannte Frobenius-Norm

$$\|A\|_F = \sqrt{\sum_{i,k} a_{ik}^2} \text{ ist zwar } M\text{-Norm, aber ungünstig wegen } \|E\|_F = \sqrt{n}$$

($\|E\| = 1$ soll möglichst erreicht werden !)

3 Vektor- und Matrixnormen, Fehlerabschätzungen für LGS, iterative Verfahren

Beispiele von M-Normen:

$$\|A\|_Z := \max_{i=1}^n \sum_{k=1}^n \|a_{ik}\| \quad \text{Zeilensummennorm (3.4)}$$

$$\|A\|_S := \max_{k=1}^n \sum_{i=1}^n \|a_{ik}\| \quad \text{Spaltensummennorm (3.5)}$$

Dabei sind (N1),(N2),(N3) und (3.3) über die Δ -Ungleichung nachzuweisen.

Einführung der Spektralnorm

Ist $A \in \mathbb{R}^{n \times n}$ (bzw. $\mathbb{C}^{n \times n}$) und sind $\lambda_1, \dots, \lambda_n \in \mathbb{C}$ die Eigenwerte von A (entsprechend ihrer Vielfachheit), so heißt

$$\rho(A) := \max_{i=1}^n |\lambda_i| \in \mathbb{R} \quad (3.6)$$

der Spektralradius von A.

Ist $\|\bullet\|$ eine beliebige M-Norm, so gilt für jede Matrix $A \in \mathbb{R}^{n \times n}$ ($\mathbb{C}^{n \times n}$)

$$\rho(A) \leq \|A\| \quad (3.7)$$

Beweis: $A\vec{x} = \lambda\vec{x}$, $\vec{x} \neq 0$ Eigenvektor zu λ

$$\text{Sei (z.B.) } \lambda \in \mathbb{R}, \vec{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \leftrightarrow \begin{pmatrix} x_1 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ x_n & 0 & \cdots & 0 \end{pmatrix} = \mathbf{X} \in \mathbb{R}^{n \times n} \neq 0$$

$$A\vec{x} = \lambda\vec{x} \Leftrightarrow AX = \lambda X$$

$$\Rightarrow \|\lambda X\| = |\lambda| \|X\| = \|AX\| \leq \|A\| * \underbrace{\|X\|}_{\neq 0}$$

$$\Rightarrow \|\lambda X\| \stackrel{(N2)}{=} |\lambda| \|X\| = \|AX\| \stackrel{(3.3)}{\leq} \|A\| * \underbrace{\|X\|}_{\neq 0} \Rightarrow \frac{1}{\|X\|} * |\lambda| * \|X\| = |\lambda| \leq \|A\| \text{ für}$$

betragsmaximales λ anschreiben $\Rightarrow \delta(A) \leq \|A\| \quad \square$

Läßt sich in (3.7) auch "=" schreiben ?

$A \in \mathbb{R}^{n \times n}$ beliebig, $B := A^T A$ ist positiv definit,

denn $\vec{x}^T B \vec{x} = \vec{x}^T A^T A \vec{x} = (A\vec{x})^T (A\vec{x}) \geq 0$

\Rightarrow alle Eigenwerte von B sind $\in \mathbb{R}$ und ≥ 0

$$\Rightarrow \|A\|_{\text{spek}} := \sqrt{\delta(B)} = \max_{\lambda \in \text{EW von } B} \sqrt{\lambda} \quad (3.8)$$

Bemerkung: Für $A \in \mathbb{C}^{n \times n}$: $B := \overline{A}^T A$

Hinweis: N1, N2, N3 und (3.3) gelten \Rightarrow also $\|A\|_{\text{spek}}$ ist Norm (Spektralnorm).

Speziell: $A = A^T \in \mathbb{R}^{n \times n} \Rightarrow \delta(A) = \|A\|_{\text{spek}}$

\Rightarrow Für symmetrische Matrizen hat $\|A\|_{\text{spek}}$ Minimaleigenschaft, d.h. $\|A\|_{\text{spek}} \leq \|A\|$ für alle M-Normen $\|\bullet\|$

Bemerkung: $\|\bullet\|_{\text{spek}}$ ist i.a. nur aufwendig zu bestimmen und mehr von theoretischem Interesse für Konvergenzüberlegungen.

Beziehungen zwischen V- und M- Normen:

LGS $A\vec{x} = \vec{r}$, A^{-1} existiert, Näherungen $\tilde{\vec{x}}$ für \vec{x}

$\vec{d} = A\tilde{\vec{x}} - \vec{r}$ (i.a. $\neq \vec{0}$) Defekt der Näherung $\tilde{\vec{x}}$

3 Vektor- und Matrixnormen, Fehlerabschätzungen für LGS, iterative Verfahren

$$\frac{\begin{array}{l} A\vec{x} = \vec{r} \\ A\vec{\tilde{x}} = \vec{r} + \vec{d} \\ \vec{\tilde{x}} - \vec{x} = A^{-1}\vec{d} \end{array}}{\left| \begin{array}{l} - \\ + \\ \end{array} \right.} \Rightarrow A(\vec{\tilde{x}} - \vec{x}) = \vec{d} \Rightarrow \|\vec{\tilde{x}} - \vec{x}\| = \|A^{-1}\vec{d}\|$$

daraus soll gelten: $\Rightarrow \|\vec{\tilde{x}} - \vec{x}\| \leq \|A^{-1}\|_M \|\vec{d}\|$

$\|A^{-1}\|_M$ ist eventuell abschätzbar, wobei eine möglichst gute Abschätzung bevorzugt wird.

Definition: $\|\cdot\|$ sei V - Norm auf \mathbb{R}^n , $\|\cdot\|_M$ sei M - Norm auf $\mathbb{R}^{n \times n}$. $\|\cdot\|$ und $\|\cdot\|_M$ heißen miteinander verträglich (passend), wenn gilt $\forall \vec{x} \forall A : \|A\vec{x}\| \leq \|A\|_M \|\vec{x}\|$

Darüber hinaus heißt $\|\cdot\|_M$ der V - Norm $\|\cdot\|$ zugeordnet, wenn es ein $\vec{x} \neq \vec{0}$ mit $\|A\vec{x}\| = \|A\|_M * \|\vec{x}\|$ gibt. ($\|\cdot\|_M$ dann eindeutig bestimmt)

Beispiel: $E = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix} \in \mathbb{R}^{n \times n}$, $\| \underbrace{E * E}_E \|_M \stackrel{(3.3)}{\leq} \underbrace{\|E\|_M * \|E\|_M}_{\neq 0} * \frac{1}{\|E\|_M}$

$\Rightarrow 1 \leq \|E\|_M$ z.B. $\|E\|_{Frob} = \sqrt{\sum_{i,k=1}^n a_{ik}^2} = \sqrt{n}$

$\|\cdot\|_M$ sei $\|\cdot\|$ zugeordnet: $\exists \vec{x} \neq \vec{0}$

$\| \underbrace{E\vec{x}}_{\text{vecx}} \| = \|E\|_M \underbrace{\|\vec{x}\|}_{\neq 0} \Rightarrow \|E\|_M = 1$, daher für $n > 1$ ist $\|E\|_{Frob}$ keiner V - Norm zugeordnet.

zugeordnet.

Paare zugeordneter V - und M - Normen:

$\|\cdot\|_\infty \leftrightarrow \|\cdot\|_Z$ (Zeilensumme) (3.9)

$\|\cdot\|_1 \leftrightarrow \|\cdot\|_S$ (Spaltensumme)

$\|\cdot\|_2 \leftrightarrow \|\cdot\|_{spek}$

Die Kondition

LGS $A\vec{x} = \vec{r}$, $A \in \mathbb{R}^{n \times n}$, $\det A \neq 0$, $\vec{r} \neq \vec{0}$ ($\Rightarrow \vec{x} \neq \vec{0}$)

$\|\cdot\|$ und $\|\cdot\|_M$ seien einander zugeordnete V - bzw. M -Normen.

\vec{A} , \vec{r} seien mit "Eingangsfehlern" $\rho_A \in \mathbb{R}^{n \times n}$, $\vec{\rho}_r \in \mathbb{R}^n$ behaftet ($\Rightarrow \vec{x} \rightarrow \vec{x} + \vec{\rho}_x$)

LGS: $(A + \rho_A)(\vec{x} + \vec{\rho}_x) = \vec{r} + \vec{\rho}_r$

Der Fehler ρ_A sei so klein, daß gilt:

$\|A^{-1}\|_M \|\rho_A\|_M \stackrel{!}{<} 1$ ♡

$\|A^{-1}\|_M$ ist dabei i.a. nur abschätzbar.

Die Größe $\kappa(A)$ ($= \text{cond}(A)$) $= \|A\|_M \|A^{-1}\|_M$ (3.11)

heißt Kondition von A .

Es gilt (durch geeignete Abschätzungen)

$$\underbrace{\frac{\|\vec{\rho}_x\|}{\|\vec{x}\|}}_{\text{rel. Fehler von } \vec{x}} \leq \frac{\kappa(A)}{1 - \kappa(A)} \left(\underbrace{\frac{\|\rho_A\|_M}{\|A\|_M}}_{\text{rel. Fehler von } A} + \underbrace{\frac{\|\vec{\rho}_r\|}{\|\vec{r}\|}}_{\text{rel. Fehler von } \vec{r}} \right) \text{ wegen } \heartsuit$$

$\|A^{-1}\|_M \stackrel{!}{<} 1$

Bemerkung:

$\kappa(A)$ kann i.a. nur abgeschätzt werden.

3 Vektor- und Matrixnormen, Fehlerabschätzungen für LGS, iterative Verfahren

$1 \stackrel{s.o.}{=} \|E\|_M = \|AA^{-1}\|_M \stackrel{(3.3)}{\leq} \|A\|_M \|A^{-1}\|_M =: \kappa(A)$
schlechte Fehlerfortpflanzung für $\kappa(A) \gg 1$.
Dann heißt A schlecht konditioniert.

3.2 Iterative Verfahren zur Lösung von LGS

(Gauß 1833 (Brief), Seidel, Jacobi)

Vorteile gegenüber direkten Verfahren: Selbstkorrektur von Rundungsfehlern

Nachteil: nur für Spezialfälle, schwach besetzte Matrizen (aber in Anwendungen häufig)

$A\vec{x} = \vec{r}$, $A = (a_{ik})_{i,k} \in \mathbb{R}^{n \times n}$, $\vec{x}, \vec{r} \in \mathbb{R}^n$

$$\text{Zerlegung } A = \left(\begin{array}{cccc} & & & A_k \\ & \backslash & \backslash & \\ & & D & \\ A_L & & & \end{array} \right) = A_L + D + A_R \text{ mit } A_L = \begin{pmatrix} 0 & \cdots & \cdots & 0 \\ a_{21} & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ a_{nn} & \cdots & a_{nn-1} & 0 \end{pmatrix}$$

$A\vec{r} = \vec{r} \Rightarrow D\vec{x} = \vec{r} - A_L\vec{x} - A_R\vec{x}$, daraus Iteration in Gesamtschritten (Jacobi) GSV

$$D\vec{x}^{(k+1)} = \vec{r} - A_L\vec{x}^{(k)} - A_R\vec{x}^{(k)} \quad (D = \text{Diagonalmatrix}) \quad (3.13a)$$

$\vec{x}^{(0)}$ vorgegeben, z.B. $\vec{x}^{(0)} = \vec{0}$ oder Näherung.

Mindestvoraussetzung: $\forall_i a_{ii} \neq 0$ ($\Rightarrow D^{-1}$ existiert) \Rightarrow

$$x_i^{(k+1)} = \frac{1}{a_{ii}} \left(r_i - \sum_{j=1, j \neq i}^n a_{ij} x_j^{(k)} \right), \quad i = 1, \dots, n, \quad k = 0, 1, 2, \dots \quad (3.13b)$$

daneben Iteration in Einzelschritten (Gauß-Seidel) ESV

Bei Berechnung von $x_i^{(k+1)}$ schon $x_1^{(k+1)} \dots x_{i-1}^{(k+1)}$ einsetzen (meist besser),

d.h. $(A_L + D)\vec{x} = \vec{r} - A_R\vec{x} \Rightarrow$

$$(A_L + D)\vec{x}^{(k+1)} = \vec{r} - A_R\vec{x}^{(k)}, \quad \vec{x}^{(0)} \text{ vorgegeben, } k = 0, 1, \dots \quad (3.14a)$$

$$x_i^{(k+1)} = \frac{1}{a_{ii}} \left(r_i - \sum_{j=1}^{i-1} a_{ij} x_j^{(k+1)} - \sum_{j=i+1}^n a_{ij} x_j^{(k)} \right) \quad (i = 1, \dots, n, \quad k = 0, 1, \dots) \quad (3.14b)$$

Verallgemeinerung von (3.14):

Das SOR-Verfahren (Jung):

(successive overrelaxatio)

Relaxationsverfahren

$$A\vec{x} = \vec{r} \Rightarrow A_L\vec{x} = -D\vec{x} - A_R\vec{x} + \vec{r} \quad | * \omega \in \mathbb{R} \quad | + D\vec{x}$$

$$\Rightarrow (D + \omega A_L)\vec{x} = [(1 - \omega)D - \omega A_R]\vec{x} + \omega \vec{r}$$

$$\Rightarrow (D + \omega A_L)\vec{x}^{(k+1)} = [(1 - \omega)D - \omega A_R] * \vec{x}^{(k)} + \omega \vec{r}, \quad (3.15a)$$

$\vec{x}^{(0)}$ gegeben, $k = 0, 1, \dots$

$$x_i^{(k+1)} = (1 - \omega)x_i^k + \frac{\omega}{a_{ii}} \left(r_i - \sum_{j=1}^{i-1} a_{ij} x_j^{(k+1)} - \sum_{j=i+1}^n a_{ij} x_j^{(k)} \right) \quad \omega = 1 \Rightarrow \text{ESV} \quad (3.15b)$$

Anwendung:

$A\vec{x} = \vec{b}$. Finde Näherung zur Lösung \vec{x} ohne A komplett zu invertieren.

Bemerkung über Matrixpotenzen:

$$A \in \mathbb{R}^{x \times x}, \quad m \in \mathbb{N}$$

Wann gilt $A^m \xrightarrow{(m \rightarrow \infty)} 0$ komponentenweise ?

Satz 3.1 $A^m \rightarrow 0$ gilt genau dann, wenn $\rho(A) < 1$ ist.

In diesem Fall ist $(E - A)$ regulär.

Ist für eine Matrixnorm $\|\cdot\|_M$ mit $\|E\|_M = 1$ auch $\|A\|_M < 1$, so gilt ferner

3 Vektor- und Matrixnormen, Fehlerabschätzungen für LGS, iterative Verfahren

$$\|(E - A)^{-1}\|_M \leq \frac{1}{1 - \|A\|_M} \quad (3.18)$$

Beweis: (nur für diagonalähnliche Matrizen)

$$A := T^{-1}DT, \quad D = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_n \end{pmatrix} \quad (\lambda_i \text{ EW von } A)$$

$$A^m = \underbrace{(T^{-1}DT)}_E \underbrace{(T^{-1}DT)}_E \underbrace{(T^{-1}DT)}_E \dots$$

$$= T^{-1}D^mT \text{ mit } D^m = \begin{pmatrix} \lambda_1^m & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_n^m \end{pmatrix} \text{ und } T \text{ unabhängig von } m.$$

$$\Rightarrow A^m \rightarrow 0 \Leftrightarrow D^m \rightarrow 0 \Leftrightarrow \forall i \quad |\lambda_i| < 1 \Leftrightarrow \rho(A) < 1$$

Aus der Regularität von $E - A$ folgt:

$Ax = x$ nur für $\vec{x} = \vec{0}$ lösbar, da $\lambda = 1$ kein Eigenwert.

$$A\vec{x} = \lambda\vec{x} \quad \rho(A) = \max_{i=1}^n |\lambda_i| < 1$$

$$(E - A)\vec{x} = \vec{0} \Leftrightarrow \vec{x} = \vec{0} \quad \square$$

Konvergenzuntersuchungen:

Verfahren nach der Art $\vec{x}^{(k+1)} = M\vec{x}^{(k)} + \vec{s}$, $\vec{x}^{(0)}$ gegeben, $k = 0, 1, 2, \dots$ (3.16)

GSV: $M = -D^{-1}(A_L + A_R)$, SOR: $M = (D + \omega A_L)^{-1}[(1 - \omega)D - \omega A_R]$

Satz 3.2 Konvergenz des SOR-Verfahrens:

Das Iterationsverfahren (3.16) konvergiert genau dann für beliebige $\vec{x}^{(0)}$ gegen die Lösung des LGS $(E - M)\vec{x} = \vec{s}$ (äquivalent $A\vec{x} = \vec{r}$), wenn $\rho(M) < 1$ ist.

Beweis:

$E - M$ regulär, d.h. $(E - M)\vec{x} = \vec{s}$ eindeutig auflösbar.

$$\left. \begin{array}{l} \vec{x}^{(k+1)} = M\vec{x}^{(k)} + \vec{s} \\ \vec{x} = M\vec{x} + \vec{s} \end{array} \right\} \begin{array}{l} + \\ - \end{array} \vec{x}^{(k+1)} - \vec{x} = M(\vec{x}^{(k)} - \vec{x})$$

Mit $\vec{x}^{(k)} - \vec{x} = M(\vec{x}^{(k-1)} - \vec{x})$ ist dies

$$= M^2(\vec{x}^{(k-1)} - \vec{x}) = M^{k+1}(\vec{x}^{(0)} - \vec{x}) \xrightarrow{k \rightarrow \infty} O(\vec{x}^{(1)} - \vec{x}) = \vec{0}$$

wegen vorigem Satz. \square

Weitere Aussagen zu diesem Satz: (ohne Beweis, stets $\forall i \quad a_{ii} \neq 0$)

1. Das SOR-Verfahren kann nur dann konvergieren, wenn $0 < \omega < 2$ gilt.
2. Ist $A = A^T$ positiv definit, so konvergiert das SOR-Verfahren für alle $\omega \in (0, 2)$

3. Es sei $\omega \in (0, 2)$. $A = (a_{ik})_{i,k} \in \mathbb{R}^{n \times n}$ sowie

$$\beta_1(\omega) := |1 - \omega| + \frac{\omega}{|a_{11}|} \sum_{j=2}^n |a_{1j}|$$

$$\beta_2(\omega) := |1 - \omega| + \frac{\omega}{|a_{22}|} (|a_{21}| * \beta_1(\omega) + \sum_{j=3}^n |a_{2j}|)$$

allgemein:

$$\beta_i(\omega) := |1 - \omega| + \frac{\omega}{|a_{ii}|} (\sum_{j=1}^{i-1} |a_{ij}| \beta_j(\omega) + \sum_{j=i+1}^n |a_{ij}|) \quad (i = 2, \dots, n)$$

3 Vektor- und Matrixnormen, Fehlerabschätzungen für LGS, iterative Verfahren

ferner:

$$\beta(\omega) := \max_{i=1}^n |\beta_i(\omega)|.$$

Ist $\beta(\omega) < 1$, so gilt

$$\|M\|_Z = \|(D + \omega A_L)^{-1}[(1 - \omega)D - \omega A_R]\|_Z \leq \beta(\omega) < 1,$$

d.h. das *SOR-Verfahren* konvergiert.

$$\forall_{i=1, \dots, n} \quad \sum_{k=1; k \neq i}^n |a_{ik}| < |a_{ii}| \quad (\text{Diagonaldominant})$$

bzw. das *schwache Zeilensummenkriterium* (öfter):

$$\forall_{i=1, \dots, n} \quad \sum_{k=1; k \neq i}^n |a_{ik}| \leq |a_{ii}| \quad \text{und} \quad \exists_{i=1, \dots, n} \sum_{k=1; k \neq i}^n |a_{ik}| < |a_{ii}|$$

(Situation häufig bei Randwertaufgaben, partiellen Differentialgleichungen)

Satz 3.3

$A^m \rightarrow 0$ gilt genau dann, wenn $\rho(A) < 1$. In diesem Fall ist $(E - A)$ regulär. Ist für eine M -Norm $\|\cdot\|_M$ mit $\|E\|_M = 1$ auch $\|A\| < 1$, so gilt ferner

$$\|(E - A)^{-1}\|_M \leq \frac{1}{1 - \|A\|_M} \quad (3.18)$$

Beweis: (nur für diagonalähnliche Matrizen)

$$A := T^{-1}DT, \quad D = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} \quad (\lambda_i \text{ Eigenwerte von } A)$$

$$A^m = (T^{-1}DT)(\underbrace{T^{-1}DT}_E)(\underbrace{T^{-1}DT}_E) \dots = T^{-1}D^mT, \quad D^m = \begin{pmatrix} \lambda_1^m & & 0 \\ & \ddots & \\ 0 & & \lambda_n^m \end{pmatrix}$$

$$A^m \rightarrow 0 \Leftrightarrow D^m \rightarrow 0 \Leftrightarrow \forall_i |\lambda_i| < 1 \Leftrightarrow \rho(A) < 1$$

Regularität von $E - A$: $A\vec{x} = \vec{x}$ nur für $\vec{x} = \vec{0}$ lösbar, da $\lambda = 1$ kein Eigenwert

$$A\vec{x} = \lambda\vec{x} \quad \rho(A) = \max_{i=1}^n |\lambda_i| < 1, \quad (E - A)\vec{x} = \vec{0} \Leftrightarrow \vec{x} = \vec{0} \quad \square$$

Weitere Aussagen über Satz 3.2 (ohne Beweis)

stets $\forall_i a_{ii} \neq 0$

1. Das *SOR-Verfahren* kann nur dann konvergieren, wenn $0 < \omega < 2$ gilt.

2. Ist $A = A^T$ positiv definit, so konvergiert das *SOR-Verfahren* $\forall \omega \in (0, 2)$

3. Es sei $\omega \in (0, 2)$. $A = (a_{ik})_{i,k} \in \mathbb{R}^{n \times n}$ sowie

$$\beta_1(\omega) := |1 - \omega| + \frac{\omega}{|a_{11}|} \sum_{j=2}^n |a_{1j}|$$

$$\beta_2(\omega) := |1 - \omega| + \frac{\omega}{|a_{22}|} (|a_{21}| \beta_1(\omega) + \sum_{j=3}^n |a_{2j}|)$$

\vdots

$$\text{allgemein: } \beta_i(\omega) := |1 - \omega| + \frac{\omega}{|a_{ii}|} (\sum_{j=1}^{i-1} |a_{ij}| \beta_j(\omega) + \sum_{j=i+1}^n |a_{ij}|) \quad (i = 2; \dots, n)$$

$$\text{ferner } \beta(\omega) := \max_{i=1}^n |\beta_i(\omega)|.$$

Ist $\beta(\omega) < 1$, so gilt:

3 Vektor- und Matrixnormen, Fehlerabschätzungen für LGS, iterative Verfahren

$\|M\|_Z = \|(D + \omega A_L)^{-1}[(1 - \omega)D - \omega A_R]\|_Z \leq \beta(\omega) < 1$, d.h. das *SOR-Verfahren* konvergiert.

$A \in \mathbb{R}^{n \times n}$ heißt zerfallend (zerlegbar, reduzibel), wenn A durch Vertauschen von Zeilen und gleichnummerierten Spalten auf die Form

$$\tilde{A} = \left(\begin{array}{c|c} A_1 & A_2 \\ \hline 0 & A_3 \end{array} \right), A_1, A_3 \text{ quadratisch, gebracht werden kann.}$$

Mit anderen Worten:(vgl. Paragraph 2): Es gibt eine Permutationsmatrix P mit $\tilde{A} = PAP^T$.

Bemerkung: LGS $A\vec{x} = \vec{r}$ zerfällt $\tilde{A}\vec{x} = \vec{r}$ $\vec{x} = \begin{pmatrix} \vec{z} \\ \vec{y} \end{pmatrix}$

LGS $A_3\vec{y} = \vec{s}$ lösen mit $\vec{r} = \begin{pmatrix} \vec{t} \\ \vec{s} \end{pmatrix}$ dann $A_1\vec{z} = \vec{t} - A_2\vec{y}$

Es gilt: Genügt A dem starken Zeilensummenkriterium oder ist A nicht zerfallend und genügt dem schwachen Zeilensummenkriterium, dann konvergiert das Einzelschrittverfahren und für jedes $\omega \in (0, 1)$ das *SOR-Verfahren*.

Weitere Bemerkungen zum Iterationsverfahren für LGS:

1. Iteration mit Elimination

$A\vec{x} = \vec{r}$, $A = B + C$ Zerlegung

LGS $B\vec{y} = \vec{s}$ sei für alle \vec{s} "leicht auflösbar" (z.B. B^{-1} bekannt, A durch "Störung" C aus B hervorgegangen)

Iterationsverfahren $B\vec{x}^{(k+1)} = C\vec{x}^{(k)} + \vec{r}$, $\vec{x}^{(0)}$ gegeben (3.19)

Konvergiert für $\rho(B^{-1}C)$ (vgl. Satz 3.2).

Bemerkung: Nachiteration

LGS: $A\vec{x} = \vec{r}$

$\vec{x}^{(0)}$ Näherungslösung (z.B. Gauß, Jacobi, *SOR*, ...)

$\vec{d} = A\vec{x}^{(0)} - \vec{r}$ Defekt (Residuum)

Im Allgemeinen: $\vec{d} \neq 0$.

Ansatz:

$$\vec{x} = \vec{x}^{(0)} + \vec{z}$$

$$0 = A\vec{x} - \vec{r} = A\vec{x}^{(0)} + A\vec{z} - \vec{r} = A\vec{z} + \vec{d}$$

\Rightarrow Löse $A\vec{z} = -\vec{d}$ näherungsweise \rightsquigarrow ergibt $\vec{z}^{(0)}$

Definiere $\vec{x}^{(1)} = \begin{pmatrix} \vec{0} \\ \vec{z}^{(0)} \end{pmatrix}$

Eventuell fortfahren.

4 Nichtlineare Gleichungssysteme

4.1 Das Banach'sche Verfahren

(konvergiert linear)

(vergleiche LGS: $A\vec{x} = \vec{r} \Leftrightarrow \vec{x} = M\vec{x} + s$)

\rightsquigarrow Iteratives Verfahren $\vec{x}^{(k+1)} = M\vec{x}^{(k)} + s$ konvergiert genau dann, wenn $\rho(M) < 1$.)

Entsprechend NLGS $F(\vec{x}) = 0$, $F : D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ äquivalent umformen zu $\vec{x} = \vec{\phi}(\vec{x})$, so daß der "Einfluß" des nichtlinearen Anteils $\vec{\phi}$ möglichst klein wird,

\rightsquigarrow iteratives Verfahren $\vec{x}^{(k+1)} = \vec{\phi}(\vec{x}^{(k)})$, $\vec{x}^{(0)}$ gegeben.

Das Verfahren konvergiert, wenn es einen Bereich $B \subseteq \mathbb{R}^n$ (B abgeschlossen) gibt mit $\phi|_B \rightarrow B$ und ϕ ist kontrahierend (Lipschitz-stetig),

d.h. $\exists \alpha < 1, \forall x, y \in B : \|\vec{\phi}(\vec{x}) - \vec{\phi}(\vec{y})\| \leq \alpha \|\vec{x} - \vec{y}\|$

($\|\cdot\|$ beliebige Norm auf \mathbb{R}^n).

Beispiel:

Gesucht ist die Lösung der NLG $x = e^{-x} =: \phi(x)$, $x \in \mathbb{R}$

Nehme $B = [0, 5; 0, 69]$. Man prüft leicht nach, daß $\phi : B \rightarrow B$ (z.B. weil ϕ monoton und $\phi(0, 5) \in B$, $\phi(0, 69) \in B$).

$\alpha = \max_{x \in B} |\phi'(x)| = \max_{x \in B} | -e^{-x} | = e^{-0.5} = 0, 606531 < 1$

$\Rightarrow \phi$ ist kontrahierend $\stackrel{BFPS}{\Rightarrow} \phi$ besitzt in B einen Fixpunkt.

Banach Fixpunktsatz

k	$x^{(k)}$
0	0, 55
1	0, 5769498
2	0, 56160877
3	0, 57029086
4	0, 56536092

TI85: 0, 5671432...

Fehlerabschätzungen:

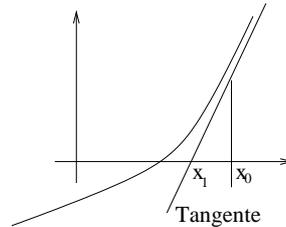
lassen sich wie bei den Sassenfeldzahlen herleiten:

$$\begin{aligned} \|\vec{x} - \vec{x}^{(k)}\| &\leq \frac{\alpha}{1-\alpha} \|\vec{x}^{(k)} - \vec{x}^{(k-1)}\| && \text{aposteriori} \\ &\leq \frac{\alpha^k}{1-\alpha} \|\vec{x}^{(1)} - \vec{x}^{(0)}\| && \text{apriori} \end{aligned}$$

Diese Abschätzung macht nur Sinn, wenn $\alpha \ll 1$.

4.2 Das Newton-Verfahren

(konvergiert quadratisch)



$$x^{(k+1)} = x^{(k)} - \frac{f(x^{(k)})}{f'(x^{(k)})} \quad (\heartsuit)$$

Herleitung auch über den Taylor'schen Satz möglich:

$$0 = f(\xi) = f(x^{(0)}) + \frac{f'(x^{(0)})}{1!}(\xi - x^{(0)}) + \frac{f''(\tau)}{2!}(\tau - x^{(0)})^2$$

für ein $\tau \in [\xi; x^{(0)}]$

“Ich kann die Indizes nicht lesen !”

“Fies, neh ?”

Für $|\xi - x^{(0)}| \ll 1$, d.h. $(\xi - x^{(0)})^2 \ll |\xi - x^{(0)}|$ vernachlässigen wir den letzten Term:

$$0 \approx f(x^{(0)}) + f'(x^{(0)})(\xi - x^{(0)}).$$

“=’ statt “ \approx ”, $x^{(1)}$ statt ξ , auflösen nach $x^{(1)}$ ”

\Rightarrow Newton-Verfahren (\heartsuit).

Der Taylorsatz gilt entsprechend für Funktionen von mehreren Veränderlichen, hier (aus übersichtlichkeitsgründen) mit 2 Variablen:

$$\text{NLGS: } \vec{F}(\vec{x}) = 0$$

$$\vec{x} = \begin{pmatrix} \xi \\ \eta \end{pmatrix}, \text{ d.h. } f(\xi, \eta) = 0, \quad g(\psi, \eta) = 0, \text{ mit } f, g : D \subseteq \mathbb{R}^2 \rightarrow \mathbb{R}$$

$$\vec{x}^{(0)} = \begin{pmatrix} \xi_0 \\ \eta_0 \end{pmatrix} \text{ sei eine Näherung, dann entsprechend wie oben}$$

$$0 = \vec{F}(\vec{x}) = \vec{F}(\vec{x}^{(0)}) + \begin{pmatrix} (\xi - \xi_0)f_\xi(\xi_0, \eta_0) + (\eta - \eta_0)f_\eta(\xi_0, \eta_0) \\ (\xi - \xi_0)g_\xi(\xi_0, \eta_0) + (\eta - \eta_0)g_\eta(\xi_0, \eta_0) \end{pmatrix} + R$$

Der Rest R enthält partielle Ableitungen 2. Ordnung an einer Zwischenstelle.

$\vec{x}^{(1)}$ statt \vec{x}, \dots (s.o.)

$$\vec{F}(\vec{x}^{(0)}) + \underbrace{\begin{pmatrix} f_{\xi_0} & f_{\eta_0} \\ g_{\xi_0} & g_{\eta_0} \end{pmatrix}}_{:=\Phi(\xi_0, \eta_0)} \Big|_{\begin{pmatrix} \xi_0 \\ \eta_0 \end{pmatrix}} * (\vec{x}^{(1)} - \vec{x}^{(0)}) = 0$$

Dabei ist $\Phi(\xi_0, \eta_0)$ die Funktionalmatrix (Jacobimatrix).

Ist $\Phi(\xi_0, \eta_0)$ regulär, so kann man obige Gleichung nach $x^{(1)}$ auflösen:

$$\vec{x}^{(k+1)} = \vec{x}^{(k)} - \Phi(\vec{x}^{(k)})^{-1} \vec{F}(\vec{x}^{(k)})$$

Entsprechende Formeln für $x \times n$ -NLGS.

4 Nichtlineare Gleichungssysteme

Bemerkungen zum Newton-Verfahren:

1. Newton konvergiert wesentlich schneller als Banach, wenn man bereits in der Nähe der Nullstelle ist, aber oft führt er auch zu einem "Hin- und Herpendeln", wenn z.B. mehrere Nullstellen vorliegen und $\vec{x}^{(0)}$ nicht sehr nahe an einer Nullstelle ist.

2. Vereinfachtes Newton-Verfahren:

Da $\Phi(\vec{x}^{(k)})^{-1}$ oft mühsam zu berechnen ist, kann man eventuell stets $\Phi(\vec{x}^{(0)})^{-1}$ verwenden:

$$\vec{F}(\vec{x}^{(k)}) + \Phi(\vec{x}^{(0)})(\vec{x}^{(k+1)} - \vec{x}^{(k)}) = 0$$

Dieses Verfahren konvergiert natürlich langsamer.

4.3 Nichtlineare GSV, ESV und SOR-Verfahren

(GSV: Gesamtschrittverfahren von Jacobi)
(ESV: Einzelschritt-Verfahren von Gauß -Seidel)

NLGS:

$$f_1(\xi_1, \dots, \xi_n) = 0$$

⋮

$$; f_i : D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$$

$$f_n(\xi_1, \dots, \xi_n) = 0$$

$$\vec{x}^{(0)} = \begin{pmatrix} \xi_1^{(0)} \\ \vdots \\ \xi_n^{(0)} \end{pmatrix} \in D \text{ gegeben.}$$

Für $i = 1, \dots, n$

$f_i(\xi_1^{(0)}, \dots, \xi_{i-1}^{(0)}, \xi_i^{(0)}, \xi_{i+1}^{(0)}, \dots, \xi_n^{(0)}) = 0$ nach ξ_i auflösen (z.B. einfaches Newton-Verfahren oder ähnliches), dann $\xi_i^{(1)} := \xi_i$ setzen.

$$\vec{x}^{(1)} = \begin{pmatrix} \xi_1^{(1)} \\ \vdots \\ \xi_n^{(1)} \end{pmatrix} \text{ und so fortiterieren.}$$

Dies ist das nichtlineare GSV (Vorteil: jeweils eine Gleichung).

Nichtlineares ESV:

Entsprechend mit ξ_i aus $f_i(\xi_1^{(1)}, \dots, \xi_{i-1}^{(1)}, \xi_i, \xi_{i+1}^{(0)}, \dots, \xi_n^{(0)}) \stackrel{!}{=} 0$ ermittelt.

Und:

$x_i^{(k+1)} = (1-\omega)x_i^{(k)} + \omega\tilde{x}_i^{(k+1)}$ ist das nichtlineare SOR-Verfahren ($\xi_i^{(0)} = (1-\omega)\xi_1^{(0)} + \omega\xi_i$).

Konjugierte Gradientenmethoden (CG)

$A \in \mathbb{R}^{n \times n}$ symmetrische positiv definite Matrix

$$f(x) = \frac{1}{2} \langle Ax, x \rangle - \langle b, x \rangle + c, \quad x \in \mathbb{R}^n, \quad b \in \mathbb{R}^n, \quad c \in \mathbb{R} \quad (0)$$

$$\langle b, x \rangle = b^T * x = b_1 x_1 + b_2 x_2 + \dots + b_n x_n$$

Das Minimum von $f : \mathbb{R}^n \leftarrow \mathbb{R}$ ist charakterisiert durch $\underbrace{\nabla f(x)}_{=Ax-b} = 0$

4 Nichtlineare Gleichungssysteme

$$\text{Gradient: } \nabla f = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}$$

\Rightarrow Minimum im Punkt $x = A^{-1}b$

Definition: $p, q \in \mathbb{R}^n$, $p \neq 0 \neq q$ sind konjugiert bezüglich $A : \Leftrightarrow \langle Ap, q \rangle = 0 \quad \forall p \neq q$
 $A \in \mathbb{R}^{n \times n}$ positiv definit, symmetrisch

$\Rightarrow A$ hat n orthogonale Eigenvektoren $\{x_i\}$ $x = (x_1, \dots, x_n)$ (d.h. $Ax_j = \lambda_j X_j$)

$$AX = X \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} \Rightarrow X^T AX = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} =: D \Rightarrow A = XDX^T$$

$$\langle Au, u \rangle = \langle XDX^T u, u \rangle = \langle DX^T u, X^T u \rangle = \underbrace{\langle \sqrt{D}X^T u, \sqrt{D}X^T u \rangle}_w$$

Annahme: Es sei p^0, p^1, \dots, p^{n-1} eine konjugierte Basis in \mathbb{R}^n bezüglich A .

Dann setze $x^{k+1} = x^k - \alpha_k p^k$ wobei $\alpha_k \in \mathbb{R}$ (1)

so gewählt, daß $f(x^k - \alpha_k p^k) \leq f(x^k - \alpha p^k) \quad \forall \alpha \in \mathbb{R}$ (Methode des steilsten Abstiegs)

Dann gilt $\alpha_k = \frac{\langle Ax^k - b, p^k \rangle}{\langle Ap^k, p^k \rangle}$ ($k = 0, \dots, n-1$) (2)

Beweis: $f(x^k - \alpha p^k) = \frac{1}{2} \langle A(x^k - \alpha p^k), x^k - \alpha p^k \rangle - \langle b, x^k - \alpha p^k \rangle + c$

$$0 \stackrel{!}{=} \frac{df}{d\alpha} = \frac{1}{2} \langle A(-p^k), x^k - p^k \rangle + \frac{1}{2} \langle A(x^k - \alpha p^k), -p^k \rangle - \langle b, -p^k \rangle$$

$$\Rightarrow -\langle Ap^k, x^k \rangle + a \langle Ap^k, p^k \rangle + \langle b, p^k \rangle = 0 \quad \Rightarrow (2)$$

Satz 4.1

Sei A symmetrisch, positiv definit, $A \in \mathbb{R}^{n \times n}$ und $\{p^k\}_{k=0}^{n-1}$ sei eine konjugierte Basis bezüglich A und $x^{k+1} = x^k - \alpha_k p^k$ mit α_k aus (2). Dann gilt $x^m = A^{-1}b$ für ein $m \leq n$.

Beweis: $\langle Ax^{k+1} - b, p^k \rangle = \langle Ax^k - b - \alpha_k Ap^k, p^k \rangle$

$$= \langle Ax^k - b, p^k \rangle - \underbrace{\langle Ap^k, p^k \rangle}_{\alpha_k} \langle Ap^k, p^k \rangle$$

$$= \begin{cases} \langle Ax^k - b, p^j \rangle & ; j \neq k \\ 0 & j = k \end{cases}$$

Folglich $\langle Ax^n - b, p^j \rangle = \langle Ax^{n-1} - b, p^j \rangle = \dots = \langle Ax^{j+1} - b, p^j \rangle = 0$ für $j = 0, \dots, n-1$

$\Rightarrow Ax^n = b$ (da $\{p^j\}$ linear unabhängig) oder $x^n = A^{-1}b$.

Falls $Ax^n = b$ für $m < n$. Nehme $\alpha_m = 0 \Rightarrow x^m = x^{m+1} = \dots = x^n$ □

Bemerkung: $x^m = x$ exakte Lösung von $Ax = b$, d.h. der Algorithmus (1),(2) stoppt nach endlich vielen Schritten, falls exakte Arithmetik benutzt wird.

Methode des steilsten Abstiegs zur iterativen Lösung von $Ax = b$:

mit

$$x^{k+1} = x^k - \alpha_k p^k \quad (\heartsuit)$$

$$\alpha^k = \frac{\langle Ax^k - b, p^k \rangle}{\langle Ap^k, p^k \rangle} \quad (\heartsuit\heartsuit)$$

4 Nichtlineare Gleichungssysteme

wobei $\{p^0, \dots, p^{k-1}\}$ eine konjugierte Basis bezüglich A ; d.h. $\langle Ap^k, p^j \rangle = 0 \quad \forall j \neq k$
 $x^k \rightarrow x = A^{-1}b$

Satz 4.2

Sei A symmetrisch positiv definit. $A \in \mathbb{R}^{n \times n}$, $\{p^0, \dots, p^{n-1}\}$ konjugiert bezüglich A und x^{k+1} aus (\heartsuit) .

Dann gilt

$$x^m = A^{-1}b$$

(für ein gewisses $m \leq n$).

konjugierte Gradienten Methode (CG):

Gegeben x^0, p^0 :

$$x^{k+1} = x^k - \alpha_k p^k \text{ mit } \alpha_k \text{ aus } (\heartsuit\heartsuit)$$

$$p^{k+1} = Ax^{k+1} - b - \beta_k p^k$$

wobei β_k so gewählt wird, daß $\langle Ap^{k+1}, p^k \rangle = 0$.

$$\text{Daraus folgt } \beta_k = \frac{\langle Ax^{k+1} - b, Ap^k \rangle}{\langle Ap^k, p^k \rangle} \tag{1}'$$

$$\text{denn } 0 \stackrel{!}{=} \langle Ap^{k+1}, p^k \rangle$$

$$= \langle p^{k+1}, Ap^k \rangle$$

$$= \langle Ax^{k+1} - b - \beta_k p^k, Ap^k \rangle$$

$$= \langle Ax^{k+1} - b, Ap^k \rangle - \beta_k \langle p^k, Ap^k \rangle$$

Satz 4.3

Voraussetzung wie oben.

Behauptung: $\langle Ap^i, p^j \rangle = 0, i \neq j$ und die Behauptung vom vorigen Satz gilt.

Beweis:

α_k, β_k sind wohldefiniert, da $p^k \neq 0$. Sei $p^0 \neq 0$ (sonst fertig).

Wir nehmen an $p^k \neq 0$ für $k \leq m-1$ und setzen $r^k := Ax^k - b$ für $k \leq m-1$.

Merke: $r^k \neq 0$, denn sonst $\beta_{k-1} = 0 \Rightarrow p_k = \underbrace{Ax^k - b}_{r^k} - \beta_{k-1} p^{k-1} = 0 \quad \#$

Als nächstes schreiben wir

$$x^{j+1} = x^j - \alpha_j p^j \text{ als } r^{j+1} = r^j - \alpha_j Ap^j$$

$$\underbrace{(-b + Ax^{j+1})}_{r^{j+1}} = Ax^j - \alpha_j Ap^j - b = \underbrace{Ax^j - b}_{r^j} - \alpha_j Ap^j$$

Dann nach Definition von $\alpha_j : \langle r^{j+1}, p^j \rangle = 0 (= \langle r^j, p^j \rangle - \alpha_j \langle Ap^j, p^j \rangle)$

Mit $r^j = Ax^j - b$ und $\alpha_j = \frac{\langle Ax^j - b, p^j \rangle}{\langle Ap^j, p^j \rangle}$ gilt somit nach Definition von β_j :

$$\langle Ap^{j+1}, p^j \rangle = 0.$$

Jetzt nehmen wir an

$$\langle Ap^k, p^j \rangle = 0 \quad (j = 0, \dots, k-1)$$

$$\langle r^k, r^j \rangle = 0 \quad (j = 0, \dots, k-1)$$

Dies ist richtig für $k=1$ (Beweis durch Induktion).

$$\text{Dann } \langle r^{k+1}, r^j \rangle = \langle r^k - \alpha_k Ap^k, r^j \rangle = \underbrace{\langle r^k, r^j \rangle}_{=0} - \alpha_k \underbrace{\langle Ap^k, p^j + \beta_{j-1} p^{j-1} \rangle}_{=0}$$

4 Nichtlineare Gleichungssysteme

$$\begin{aligned}
 &\Rightarrow \langle r^{k+1}, p^j \rangle = 0 && r^j = p^j + \beta_{j-1} p^{j-1} \\
 &\text{und } \langle r^{k+1}, r^k \rangle = \langle r^{k+1}, p^k + \beta_{k-1} p^{k-1} \rangle = 0 \\
 &\Rightarrow \langle r^{k+1}, r^j \rangle = 0 \\
 &\text{Auch } \langle Ap^{k+1}, p^j \rangle = \langle p^{k+1}, Ap^j \rangle = \langle r^{k+1} - \beta_k p^k, Ap^j \rangle = \langle r^{k+1}, Ap^j \rangle \\
 &= \hat{\alpha}_j [\underbrace{\langle r^{k+1}, r^j \rangle}_{=0} - \underbrace{\langle r^{k+1}, r^j \rangle}_{=0}] = 0 \\
 &\text{und } \langle Ap^{k+1}, p^k \rangle = 0 \text{ durch Konstruktion.} \\
 &\text{Schließlich, angenommen } p^n = 0, \quad m < n \\
 &\Rightarrow 0 = \langle p^m, p^m \rangle = \langle r^m - \beta_{m-1} p^{m-1}, r^m - \beta_{m-1} p^{m-1} \rangle \geq \langle r^m, r^m \rangle \Rightarrow r^m = 0 \\
 &\Rightarrow r^m = Ax^m - b = 0 \Rightarrow x^m = A^{-1}b \quad \square
 \end{aligned}$$

Satz 4.4

Voraussetzungen wie oben. Dann gilt die Fehlerabschätzung ($Ax = b$):

$$\|x - x^k\| \leq 2 \left(\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1} \right)^k \|x - x^0\|$$

und $\kappa = \text{Kondition der Matrix}$.

5 Approximation

Interpolation liefert zu gegebener Funktion $f : T \rightarrow R$ (hinreichend glatt) ein $p_n \in \mathbb{P}_n$ mit $\|f - p\|_{I,\infty} = \text{“klein”}$.

Frage: Kann man (unabhängig von Interpolation) ein $p \in \mathbb{P}_n$ mit $\|f - p\|_{I,\infty} = \text{“klein”}$ finden?

Solch ein p heißt Approximation von f . Falls es existiert nennt man p mit minimalem $\|f - p\|_{I,\infty}$ beste Approximation. Genauer, falls $\|\cdot\|_{I,\infty}$ genommen wird, heißt dieses Polynom die Tschebyscheff-Approximation (leider komplizierte Theorie).

Frage: Fehler in anderen Normen?

$\|p - u\|_{L^2(0,1)}^2 := \int_0^1 (p(x) - u(x))^2 dx$ mit $p = \sum_{k=0}^{n-1} a_k x^k$

Beispiel: Approximiere $u(x) = \begin{cases} 1 & , 0 \leq x \leq \frac{1}{2} \\ 0 & , \text{sonst} \end{cases}$

$$0 \stackrel{!}{=} \frac{\partial}{\partial a_j} \int_0^1 (\sum_{k=0}^{n-1} a_k x^k - u(x))^2 dx = \int_0^1 2(\sum_{k=0}^{n-1} a_k x^k - u(x)) x^j dx \text{ mit } 0 \leq j \leq n-1$$

$$V = \text{span}\left\{ \underbrace{1}_{v_1}, x, x^2, \dots, \underbrace{x^{n-1}}_{v_n} \right\} = \mathbb{P}_{n-1}$$

$$v_j := x^{j-1}$$

$$\langle v, v_j \rangle = \int_0^1 x^{i-1} x^{j-1} dx = \frac{x^{i+j-1}}{i+j-1} \Big|_0^1 = \frac{1}{i+j-1}$$

$$\langle u, v_j \rangle = \int_0^1 u x^{j-1} dx = \int_0^{\frac{1}{2}} x^{j-1} dx = \frac{x^j}{j} \Big|_0^{\frac{1}{2}} = \frac{1}{j} 2^{-j}$$

$$H_n = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \dots & \frac{1}{n} \\ \frac{1}{2} & & & & \frac{1}{n+1} \\ \frac{1}{3} & & & & \vdots \\ \vdots & & & & \vdots \\ \frac{1}{n} & \frac{1}{n+1} & \dots & \dots & \frac{1}{2n-1} \end{pmatrix} \begin{pmatrix} a_0 \\ \vdots \\ \vdots \\ \vdots \\ a_{n-1} \end{pmatrix} = \begin{pmatrix} 2^{-1} \\ \frac{1}{2} 2^{-2} \\ \vdots \\ \vdots \\ \frac{1}{n} 2^{-n} \end{pmatrix}$$

$$\Rightarrow n=2: a^T = \left(\overbrace{\frac{5}{4}}^{a_0}; \overbrace{-\frac{3}{2}}^{a_1} \right)$$

$$n=3: \Rightarrow \text{dasselbe } a$$

$$n=4: a^T = \left(\underbrace{0,8125}_{a_0}; \underbrace{3,75}_{a_1}; \underbrace{-13,125}_{a_2}; \underbrace{8,75}_{a_3} \right)$$

$$v^{(2)} = \frac{5}{4} - \frac{3}{2}x$$

$$v^{(3)} = v^{(2)} \in \mathbb{P}_1$$

$$v^{(4)} = a_0 + a_1 x + a_2 x^2 + a_3 x^3 \in \mathbb{P}_3$$

Vorteil: Mann kann viele (nämlich alle L^2 -)Funktionen approximieren

Nachteil: Die Koeffizientenmatrix H_n (sogenannte Hilbertmatrix) ist schlecht konditioniert

Definition:

Ein Vektorraum U heißt ein reeller unitärer Raum, falls ein Skalarprodukt $\langle \dots, \dots \rangle : U \times U \rightarrow \mathbb{R}$ existiert mit

5 Approximation

$$\forall f, g, h \in U; \lambda, \mu \in \mathbb{R} : 0 \neq f \quad (5.1)$$

$$\langle f, f \rangle > 0, \langle f, g \rangle = \langle g, f \rangle, \langle f, \lambda g + \mu h \rangle = \lambda \langle f, g \rangle + \mu \langle f, h \rangle$$

In einem unitären Raum können wir eine Norm definieren:

$$\forall f \in U \|f\| := \langle f, f \rangle^{\frac{1}{2}}$$

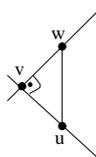
Es gilt die Cauchy-Schwarzsche Ungleichung

$$\forall f, g \in U : |\langle f, g \rangle| \leq \|f\| \cdot \|g\|$$

Beispiel:

1. $\mathbb{R}^n : \langle x, y \rangle = x^T \cdot y \quad \forall x, y \in \mathbb{R}^n$
2. $w \in C(I)$ mit $w > 0|_I$
 $L_2(I, w) := \{f : I \rightarrow \mathbb{R} \mid \int_I f^2(x)w(x)dx < \infty\}$
 $\langle f, g \rangle := \int_I f(x)g(x)w(x)dx$ (beide Räume vollständig, Hilberträume)

Sei $V \subset U$ ein linearer Unterraum von U . Zu $u \in U$ heißt $v \in V$ beste Approximation
 $:\Leftrightarrow \forall w \in V : \|u - v\| \leq \|u - w\|$



$$V = \mathbb{R}^1, \mathbb{R}^2 = U$$

Satz 5.1 (Existenz und Eindeutigkeit der besten Approximation)

Sei $v_1, \dots, v_n \in U$ linear unabhängig und $V := \text{span}\{v_1, \dots, v_n\}$.

Dann gilt:

1. Für jedes $u \in U$ existiert genau eine beste Approximation $v \in V$
2. $\forall w \in V : \langle u - v, w \rangle = 0$, d.h. $u - v \perp V$
3. Sei $A \in \mathbb{R}^{n \times n}$, $A = (\langle v_i, v_j \rangle)_{i,j}$ und $b \in \mathbb{R}^n$ mit $b_j = \langle u, v_j \rangle$
 $\Rightarrow \det A \neq 0$ wenn $a \in \mathbb{R}^n$ die sogenannten Normalgleichungen löst,
d.h. $Aa = b \Rightarrow v = \sum_{j=1}^n a_j v_j$

□

Satz 5.2

Sei $v_1, \dots, v_n \in U$ linear unabhängig und $V = \text{span}\{v_1, \dots, v_n\}$
mit U Hilbertraum. Dann gilt:

1. $\forall u \in U \exists^1$ beste Approximation $v \in V$.
2. $\forall w \in U : \langle u - v, w \rangle = 0$, d.h. $u - v \perp V$.

5 Approximation

3. Sei $A \in \mathbb{R}^{n \times n}$, $A = (\langle v_i, v_j \rangle)_{i,j}$ und $b \in \mathbb{R}^n$ mit $b_j = \langle u, v_j \rangle$ ($1 \leq j \leq n$)
 $\Rightarrow \det A \neq 0$ (Sei $x \in \mathbb{R}^n$ Lösung der Normalgleichungen $Ax = b$) (5.4)

$$\Rightarrow v = \sum_{j=1}^n x_j v_j \quad (5.5)$$

Definition: (5.1)

$$\forall f \in U : \|f\| := \langle f, f \rangle^{\frac{1}{2}} \quad (5.2)$$

$$1. 0 \neq f \in U \Rightarrow \langle f, f \rangle > 0$$

$$2. \forall f, g, h \in U : \langle f, g \rangle = \langle g, f \rangle$$

$$3. \lambda, \mu \in \mathbb{R} : \langle f, \lambda g + \mu h \rangle = \lambda \langle f, g \rangle + \mu \langle f, h \rangle$$

$$\text{Cauchy-Schwarz-Ungleichung: } |\langle f, g \rangle| \leq \|f\| \|g\| \quad \forall f, g \in U \quad (5.3)$$

Notation: A heißt Gramsche Matrix, $\det A$ Gram-determinante.

Beweis: c) Zeige zuerst $\det A \neq 0$ (folgt aus $\{v_1, \dots, v_n\}$ linear unabhängig)

Annahme: Spaltenvektoren von A sind linear abhängig.

$$\Rightarrow \exists a \in \mathbb{R}^n, a_j \neq 0 \quad (1 \leq j \leq n) : \sum_{j=1, i \text{ fest}}^n a_j \langle v_i, v_j \rangle = 0$$

$$\Rightarrow \langle v_i, \sum a_j v_j \rangle \quad \forall i \text{ fest}$$

$$\Rightarrow \langle \sum_{j=1}^n a_j v_j, \sum_{j=1}^n a_j v_j \rangle = 0 \stackrel{(5.1)}{\Rightarrow} \sum a_j v_j = 0 \text{ Widerspruch zu } \{v_j\} \text{ linear unabhängig}$$

$\Rightarrow \det A \neq 0$. Folglich $x = A^{-1}b$ ist wohldefiniert. Lösung von (5.4):

$$\text{Sei } y := \sum_{j=1}^n x_j v_j \in V$$

$$\stackrel{(v_i \in V)}{\Rightarrow} \langle u - y, v_i \rangle = \langle u, v_i \rangle - \langle y, v_i \rangle = b_i - \underbrace{\sum_{j=1}^n x_j \langle v_j, v_i \rangle}_{(AX)_i} = 0, \text{ (d.h. } u - y \perp V)$$

Sei $w \in V$

$$\stackrel{(y \in V)}{\Rightarrow} \left\| \underbrace{u}_{\in U} - w \right\|^2 = \langle u - w, u - w \rangle = \langle u - y - \underbrace{(w - y)}_{\in V}, u - y - (w - y) \rangle$$

$$\stackrel{(5.1)}{=} \langle u - y, u - y \rangle - 2 \langle u - y, \underbrace{w - y}_{\in V} \rangle + \langle w - y, w - y \rangle$$

$$\underbrace{\hspace{10em}}_{=0 \text{ (5.6)}}$$

$$= \|u - y\|^2 + \|w - y\|^2 > \|u - y\|^2 \text{ für } w \neq y.$$

d.h. $\|u - w\|$ minimal für genau ein $w \in V \Rightarrow a) w = y = v \Rightarrow v = \sum x_j v_j$ (wg. (5.5))
 und es gilt (5.6) $\Rightarrow b)$ □

Mit einem Orthonormalsystem (ONS) $\{l_1, \dots, l_n\}$ als Basis von V wird die Approximationsaufgabe (Finde beste Approximation zu $u \in U$) sehr einfach, nämlich:

Normalgleichungen (5.4):

$$\text{I Einheitsmatrix (A=I) } \langle l_j, l_k \rangle = \begin{cases} 0, & j \neq k \\ 1, & j = k \end{cases}$$

5 Approximation

$Ix = b \Rightarrow x = b \quad v = \sum_{j=1}^n \langle u, l_j \rangle l_j$ beste Approximierende aus ONS

Approximationsfehler? $u \in U, v \in V$

$$\begin{aligned} \|u - v\|^2 &\geq \|u\|^2 - \|v\|^2 \quad v = \sum_{j=1}^n \langle u, l_j \rangle l_j \\ \|v\|^2 &= \langle v, v \rangle = \langle \sum_{k=1}^n \langle u, l_k \rangle l_k, \sum_{k=1}^n \langle u, l_k \rangle l_k \rangle \\ &= \sum_{k=1}^n \langle u, l_k \rangle \langle l_k, \underbrace{\sum_{j=1}^n \langle u, l_j \rangle l_j}_{\langle u, l_k \rangle \langle l_k, l_k \rangle} \rangle \\ &\geq \|u\|^2 - \sum_{j=1}^n \langle u, l_j \rangle^2 = \sum_{k=1}^n \langle u, l_k \rangle^2 \end{aligned}$$

Bemerkung: In $L_2[-1, 1]$ erhalten wir durch Orthonormalisierung von $\{1, x, x^2, \dots\}$

die Legendre-Polynome $l_1(x) = \frac{1}{\sqrt{2}}, l_2(x) = \sqrt{\frac{3}{2}}x, l_3(x) = \frac{1}{2}\sqrt{\frac{5}{2}}(3x^2 - 1), \dots$

($u \in L_2[-1, 1] : \Leftrightarrow \int_{-1}^1 u^2(x) dx < \infty$)

Behauptung: $\int_{-1}^1 l_j(x) l_k(x) dx = \begin{cases} 0, & j \neq k \\ 1, & j = k \end{cases}$

Approximationsfehler

$u \in U, v \in V \subset U$

$$\|u - v\|^2 = \langle u - v, u - v \rangle = \langle u, u - v \rangle - \underbrace{\langle v, u - v \rangle}_{=0 \text{ nach Satz 5.1}} = \langle u, u \rangle - \langle u, v \rangle$$

ONS $\{e_1, \dots, e_n\}$ Basis von V

$v = \sum_{j=1}^n \langle u, e_j \rangle e_j$ Beste Approximierende zu n in V

$$\langle u, v \rangle = \langle u, \sum_{j=1}^n \langle u, e_j \rangle e_j \rangle = \sum_{j=1}^n \langle u, e_j \rangle \langle u, e_j \rangle = \sum_{j=1}^n \langle u, e_j \rangle^2$$

$$\Rightarrow \|u - v\|^2 = \|u\|^2 - \sum_{j=1}^n \langle u, e_j \rangle^2 \Rightarrow \|u\|^2 \geq \sum_{j=1}^n \langle u, e_j \rangle^2 \quad \text{Bessel-Ungleichung}$$

$$\|v\|^2 = \langle v, v \rangle = \langle \sum_{k=1}^n \langle u, e_k \rangle e_k, \sum_{j=1}^n \langle u, e_j \rangle e_j \rangle$$

$$= \sum_{k=1}^n \langle u, e_k \rangle \langle e_k, \sum_{j=1}^n \langle u, e_j \rangle e_j \rangle = \sum_{k=1}^n \langle u, e_k \rangle \langle u, e_k \rangle$$

$$\Rightarrow \|u - v\|^2 = \|u\|^2 - \|v\|^2 \text{ wenn } v \text{ Bestapproximierende}$$

Satz 5.3 (Gram-Schmidt-Orthonormalisierung)

Zu $u_1, \dots, u_n \in U$ linear unabhängig definiere rekursiv $e_i = \frac{v_i}{\|v_i\|}$. Angenommen: k orthonormale Vektoren $e_1, \dots, e_k, k < n$ sind schon bekannt.

$$d_{k+1} = v_{k+1} - \sum_{\nu=1}^k \langle v_{k+1}, e_\nu \rangle e_\nu, \quad e_{k+1} := \frac{d_{k+1}}{\|d_{k+1}\|} \quad (5.7)$$

Diese Rekursion kann immer durchgeführt werden und liefert n orthonormale Vektoren $e_\nu, \nu = 1, \dots, n$

Beweis: (Induktion über n) $m = 1$ trivial

Angenommen: Behauptung gilt für $m < n \Rightarrow e_1, \dots, e_m$ orthonormal

Aus (5.7) folgt: $e_\nu = \sum_{\mu=1}^\nu a_{\nu\mu} e_\mu$

Angenommen: $d_{m+1} = 0 \Rightarrow \nu_1, \dots, \nu_{m+1}$ linear abhängig (ν hat Koeffizienten 1) Widerspruch

$\Rightarrow d_{m+1} \neq 0 \Rightarrow e_{m+1} \neq 0$ wohldefiniert. Sei $\nu \leq m$:

5 Approximation

$$\langle d_{m+1}, e_\nu \rangle \langle \nu_{m+1}, e_\nu \rangle = \sum_{\mu=1}^m \underbrace{\langle \nu_{m+1}, e_\mu \rangle \langle e_\mu, e_\nu \rangle}_{\langle \nu_{m+1}, e_\nu \rangle} = 0$$

$\Rightarrow \langle e_{m+1}, e_\nu \rangle = 0 \quad \forall \nu \leq m \Rightarrow \{e_1, \dots, e_{m+1}\}$ orthonormal \Rightarrow Behauptung. \square

Beispiel: $v_1 = 1 = e_1, \quad v_2 = x = e_2, \quad v_3 = x^2 \quad \langle u, v \rangle = \int_0^1 u(x)v(x)dx \quad U = L^2(0, 1)$

$$e_1(x) = \frac{1}{\int_0^1 (1dx)^{\frac{1}{2}}} = 1 \quad d_2(x) = v_2 - \langle v_2, e_1 \rangle e_1 = x - \langle x, 1 \rangle \cdot 1 = -x - \int_0^1 x dx = -\frac{1}{2}$$

$$\|d_2\|^2 = \langle d_2, d_2 \rangle = \int_0^1 (x - v_2)^2 dx = \int_0^1 x^2 - x + \frac{1}{4} = \frac{1}{12}$$

$$e_2(x) = \frac{d_2}{\|d_2\|} = \sqrt{12}(x - \frac{1}{2}) \text{ usw.}$$

$\{\underbrace{1}_{v_1}, x, x^2, x^3, \dots\} \rightarrow$ orthonormieren bzgl. $L^2(-1, 1), \quad \langle v, u \rangle = \int_{-1}^1 v(x)u(x)dx$

$$l_1 = \frac{1}{\sqrt{2}} = \frac{v_1}{\|v_1\|} \quad \|v_1\| = \int_{-1}^1 1 dx = 2$$

$$d_2(x) = x - \langle x, \frac{1}{\sqrt{2}} \rangle \frac{1}{\sqrt{2}} = x - \int_{-1}^1 \frac{x}{\sqrt{2}} dx * \frac{1}{\sqrt{2}} = x$$

$$\|d_2(x)\| = \langle d_2(x), d_2(x) \rangle = \int_{-1}^1 x^2 dx = \frac{2}{3}$$

$$l_2(x) = \frac{d_2(x)}{\|d_2(x)\|} = \sqrt{\frac{3}{2}}x$$

$$\Rightarrow l_3(x) = \frac{1}{2}\sqrt{\frac{5}{2}}(3x^2 - 1)$$

$l_{j+1}(x)$ Legendre-Polynom vom Grad j.

spezieller Fall : $x \in U = \mathbb{R}^n \quad \|x\|_{Euklid} = \sqrt{\sum_{i=1}^n |x_i|^2}$

$I \in \mathbb{R}$

Standard-Situation: Gegeben $x_1, x_2, \dots, x_m \in I$ (Knotenpunkte)

und $v_1, v_2, \dots, v_n, I \rightarrow \mathbb{R}$

Finde $n: I \rightarrow \mathbb{R}$, wobei nur Funktionswerte $u(x_1), \dots, u(x_n)$ bekannt, d.h. u soll approximiert werden durch $v \in \text{span}\{v_1, \dots, v_m\}$ bzgl. der Methode der kleinsten Fehlerquadrate,

d.h. mit $u^T = \{u(x_1), \dots, u(x_n)\}$

$v_\nu^T = \{v_\nu(x_1), \dots, v_\nu(x_n)\} \quad (1 \leq \nu \leq n)$ finde $v \in \text{span}\{v_1, \dots, v_n\}$ mit

$$\sum_{i=1}^m (u(x_i) - v(x_i))^2 = \|u - v\|_{Eukl.}^2 \leq \|u - w\|_{Eukl.}^2 \quad \forall w \in \text{span}\{v_1, \dots, v_n\}$$

d.h. $\sum_{i=1}^n (u(x_i) - v(x_i))^2 \stackrel{!}{=} \min$

Beispiel: (Kurvenfunktional mit least squares)

Gegeben $n = 7$ Punkte (x_i, y_i) . Passe eine Gerade $y = k_0 + k_1x$ so an, daß

$$S = \sum_{i=1}^n (Y_i - y_i)^2 \stackrel{!}{=} \min \quad Y_i - y_i \text{ heißen Residuen } (Y_i = Y(x_i))$$

S wird minimal, wenn $\frac{\partial S}{\partial k_0} = 0, \quad \frac{\partial S}{\partial k_1} = 0$

$$\frac{\partial S}{\partial k_0} = \sum_{i=1}^n \frac{\partial}{\partial k_0} (k_0 + k_1x_i - y_i)^2 = \sum_{i=1}^n 2(k_0 + k_1x_i - y_i) * 1 \stackrel{!}{=} 0$$

$$\frac{\partial S}{\partial k_1} = \sum_{i=1}^n \frac{\partial}{\partial k_1} (k_0 + k_1x_i - y_i)^2 = \sum_{i=1}^n 2(k_0 + k_1x_i - y_i) * x_i \stackrel{!}{=} 0$$

$$\Rightarrow \sum_{i=1}^n y_i = \sum_{i=1}^n k_0 + k_1x_i = k_0n + k_1 \sum_{i=1}^n x_i$$

$$\sum_{i=1}^n x_i y_i = k_0 \sum_{i=1}^n x_i + k_1 \sum_{i=1}^n x_i^2$$

5 Approximation

Beispiel:

i	x_i	y_i	$x_i y_i$	$(x_i)^2$	
1	0	2	0	0	
2	1	3	3	1	
3	2	5	10	4	$42 = 7k_0 + 21k_1$
4	3	5	15	9	$164 = 21k_0 + 91k_1$
5	4	9	36	16	$\Rightarrow k_0 = 1,928571 = \frac{27}{14}$
6	5	8	40	25	$k_1 = 1,357143 = \frac{19}{14}$
7	6	10	60	36	
Σ	21	42	164	91	

Fitten einer Datenmenge mit einem Polynom vom Grad m , n Punkte (x_i, y_i)

$$S = \sum_{i=1}^n (Y - i - y_i)^2 = \sum_{i=1}^n (k_0 + k_1 x_i + k_2 x_i^2 + \dots + k_m x_i^m - y_i)^2$$

Forderung: S soll maximal werden $\Rightarrow \frac{\partial S}{\partial k_j} \stackrel{!}{=} 0$ $(0 \leq j \leq m)$

$$\frac{\partial S}{\partial k_0} = \sum_{i=1}^n 2(k_0 + k_1 x_i + k_2 x_i^2 + \dots + k_m x_i^m - y_i) * 1 \stackrel{!}{=} 0$$

$$\frac{\partial S}{\partial k_1} = \sum_{i=1}^n 2(k_0 + k_1 x_i + k_2 x_i^2 + \dots + k_m x_i^m - y_i) * x_i \stackrel{!}{=} 0$$

\vdots

$$\frac{\partial S}{\partial k_n} = \sum_{i=1}^n 2(k_0 + k_1 x_i + k_2 x_i^2 + \dots + k_m x_i^m - y_i) * x_i^m \stackrel{!}{=} 0$$

$$\begin{aligned} k_0 n + k_1 \sum x_i + k_2 \sum x_i^2 + \dots + k_m \sum x_i^m &= \sum y_i \\ k_0 \sum x_i + k_1 \sum x_i^2 + k_2 \sum x_i^3 + \dots + k_m \sum x_i^{m+1} &= \sum x_i y_i \\ &\vdots \end{aligned} \quad \Sigma := \sum_{i=1}^n$$

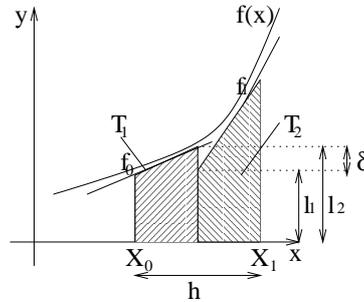
$$k_0 \sum x_i^m + k_1 \sum x_i^{m+1} + k_2 \sum x_i^{m+2} + \dots + k_m \sum x_i^{2m} = \sum x_i^m y_i$$

$\Rightarrow m + 1$ Gleichungen für k_1, \dots, k_m Normalgleichungen

$$A \vec{k} = \vec{b} \quad A = \begin{pmatrix} n & \sum x_i & \dots & \sum x_i^m \\ \sum x_i & \sum x_i^2 & & \sum x_i^{m+1} \\ \vdots & & & \vdots \\ \sum x_i^m & \sum x_i^{m+1} & \dots & \sum x_i^{2m} \end{pmatrix} \quad \vec{k} = \begin{pmatrix} k_0 \\ k_1 \\ \vdots \\ k_m \end{pmatrix} \quad \vec{b} = \begin{pmatrix} \sum y_i \\ \sum x_i y_i \\ \vdots \\ \sum x_i^m y_i \end{pmatrix}$$

6 Numerische Quadratur

Trapez-Regel:



$f(x)$ konvex in $(x_0, x_1) \Rightarrow f''(x) > 0 \forall x \in (x_0, x_1)$

$$f'_0 = \frac{\partial}{\partial x}$$

$$l_1 = f_0 + \frac{h}{2} f'_0$$

$$l_2 = f_1 - \frac{h}{2} f'_1$$

$$\text{Fläche}(T_1) = f_0 \frac{h}{2} + \frac{\partial}{\partial x} \frac{h}{2} = f_0 \frac{h}{2} + \frac{h}{4} f'_0 \frac{h}{2} = \frac{h}{2} f_0 + \frac{h^2}{8} f'_0$$

$$\text{Fläche}(T_2) = \frac{l_2 + f_1}{2} \frac{h}{2} = (2f_1 - \frac{h}{2} f'_1) \frac{h}{4} = \frac{h}{2} f_1 - \frac{h^2}{8} f'_1$$

$$\int_{x_0}^{x_1} f(x) dx \approx \text{Trapezregel: } \frac{(f_1 + f_0)}{2} h$$

$$\text{Fläche}(T_1) + \text{Fläche}(T_2) = \frac{h}{2} (f_0 + f_1) + \frac{h^2}{8} (f'_0 - f'_1)$$

Diskretisierungsfehler:

$$|E(f)| := \left| \int_{x_0}^{x_1} f(x) dx - \frac{h}{2} (f_1 + f_0) \right|$$

$$\Rightarrow |E(f)| \leq \frac{h^2}{8} |f'_1 - f'_0|$$

(Ebenso gültig für $f''(x) < 0 \forall x \in (x_0, x_1)$)

$\int_a^b f(x) dx$ kann im Allgemeinen nur näherungsweise berechnet werden,

z.B. für $f(x) = e^{-x^2}$, $f(x) = \sin(x^2)$

$$\text{Quadraturfehler: } E(f) := \int_a^b f(x) dx - \underbrace{\sum_{i=0}^n \omega_i f_i}_{\text{Quadraturformel}}$$

Dabei sind die ω_i die Gewichte.

Aufgabe: Bestimme x_0, \dots, x_n Knoten und Gewichte $\omega_0, \dots, \omega_n$ so, daß $|E(f)|$ "klein" wird.

2. Methode: Substituiere das Lagrange-Interpolationspolynom $p_n(x) = \sum_{\nu=0}^n f_{\nu} l_{\nu}(x)$ für den Integranden $f(x)$:

$$\Rightarrow \underline{n=1}: \int_a^b f(x) dx \approx \int_a^b \{f_0 l_0(x) + f_1 l_1(x)\} dx$$

$$= \int_a^b \left[\underbrace{\frac{x-x_1}{x_0-x_1}}_{l_0(x)} f_0 + \underbrace{\frac{x-x_0}{x_1-x_0}}_{l_1(x)} f_1 \right] dx \text{ mit } x_0 = a, x_1 = b, h = x_1 - x_0$$

6 Numerische Quadratur

$$\begin{aligned}
 &= \frac{1}{2h} [-(x-x_1)^2 f_0 + (x-x_0)^2 f_1] \Big|_{a=x_0}^{b=x_1} && \text{Trapezregel} \\
 &= \frac{h}{2} (f_0 + f_1) \\
 \Rightarrow \underline{n=2}: \int_a^b f(x) dx &\approx \int_a^b [l_0(x)f_0 + l_1(x)f_1 + l_2(x)f_2] dx \\
 &= \int_a^b \underbrace{\frac{(x-x_1)(x-x_2)}{(x_0-x_1)(x_0-x_2)}}_{l_0(x)} f_0 + \frac{(x-x_0)(x-x_2)}{(x_1-x_0)(x_1-x_2)} f_1 + \frac{(x-x_0)(x-x_1)}{(x_2-x_0)(x_2-x_1)} f_2 dx \\
 &\text{interpolieren in } x_0, x_1, x_2 \text{ mit Funktionswerten } f_0, f_1, f_2: \\
 &= \frac{1}{2h^2} \left[\frac{2h^3}{3} f_0 + \frac{8h^3}{3} f_1 + \frac{2h^3}{3} f_2 \right] \\
 &= \frac{h}{3} [f_0 + 4f_1 + f_2] && \text{Simson's } \frac{1}{3}\text{-Regel (1710-1761)}
 \end{aligned}$$

$E(f) = ?$

Benutze "Interpolationsfehler"-Term:

$$f \in C^{(n+1)}[a, b] : f(x) - p_n(x) = \frac{\omega(x)}{(n+1)!} f^{(n+1)}(\xi) \text{ mit } \xi \in (a, b) = (x_0, x_n) \text{ und}$$

$$\omega(x) := \prod_{\nu=0}^n (x - x_\nu)$$

$$E(f) = \int_a^b f(x) dx - \int_a^b p_n(x) dx = \int_a^b f^{(n+1)}(\xi) \frac{\omega(x)}{(n+1)!} dx$$

$n=1$: (Trapezregel):

$$E(f) = \int_{a=x_0}^{b=x_1} f^{(2)}(\xi) \frac{(x-x_0)(x-x_1)}{2!} dx = f^{(2)}(\eta) \int_a^b \frac{(x-x_0)(x-x_1)}{2} dx = -f^{(2)}(\eta) \frac{h^3}{12}$$

nach dem 2. Mittelwertsatz.

2ter Mittelwertsatz (Differential-/Integralrechnung):

$$\begin{aligned}
 &f(x), g(x) \text{ stetig auf } a \leq x \leq b \\
 &g(x) \text{ wechselt das Vorzeichen nicht} \Rightarrow \int_a^b f(x)g(x) dx = f(\eta) \int_a^b g(x) dx, \eta \in (a, b)
 \end{aligned}$$

$$\frac{(x-x_0)(x-x_1)}{2} \text{ hat konstantes Vorzeichen in } x \in (a, b) = (x_0, x_1)$$

Mit partieller Ableitung gilt:

$$\int_a^b \frac{(x-a)(x-b)}{2} dx = \underbrace{\frac{(x-a)^2}{2} \frac{(x-b)}{2}}_{=0} \Big|_{x=a}^b - \int_a^b \frac{(x-a)^2}{2 \cdot 2} dx = -\frac{(x-a)^3}{3 \cdot 4} \Big|_{x=a}^b = -\frac{h^3}{12}$$

Newton-Cotes-Formeln:

$$\begin{array}{ccccccc}
 & & | & | & | & | & \\
 & & | & | & | & | & \\
 a=x_0+ph & & x_1 & & x_2 & & x_3 & & b=x_0+qh
 \end{array}$$

Knoten $x_i, i = 0, \dots, n$, äquidistant

$$a = x_0 + ph$$

$$b = x_0 + qh, 0 \leq p < q \leq n, p, q \in \mathbb{R}$$

$$I_{n+1}(f) = \int_a^b \underbrace{\sum_{j=0}^n l_j(x) f_j}_{\text{Lagrangeinterpolationpolynom}} dx$$

$$= \sum_{j=0}^n \int_p^q \prod_{\substack{k=0 \\ k \neq j}}^n \frac{(x_0+sh) - (x_0+kh)}{(x_0+jh) - (x_0+kh)} f_j h ds$$

6 Numerische Quadratur

$$\begin{aligned}
 &= h \sum_{j=0}^n \int_p^q \underbrace{\prod_{\substack{k=0 \\ k \neq j}}^n \frac{s-k}{j-k}}_{=: \alpha_j \frac{1}{h}} f_j ds \\
 &= \sum_{j=0}^n \alpha_j f_j, \quad \alpha_j = h \int_p^q \prod_{\substack{k=0 \\ k \neq j}}^n \frac{s-k}{j-k} ds
 \end{aligned}$$

1. geschlossene Newton-Cotes-Formeln:

$$b - a = n * h, \quad x_0 = a, \quad x_n = b, \quad p = 0, \quad q = n$$

$$\underline{n=2}: \alpha_j = h \int_0^{n=2} \prod_{\substack{k=0 \\ k \neq j}}^2 \frac{s-k}{j-k} ds$$

$$\begin{aligned}
 \alpha_0 &= h \int_0^2 \frac{(s-1)(s-2)}{(-1)(-2)} ds \\
 &= \frac{h}{2} \int_0^2 (s^2 - 3s + 2) ds \\
 &= \frac{h}{2} \left(\frac{s^3}{3} - \frac{3}{2}s^2 + 2s \right) \Big|_0^2 \\
 &= \frac{h}{3}
 \end{aligned}$$

$$\begin{aligned}
 \alpha_1 &= h \int_0^2 \frac{s(s-2)}{1(1-2)} ds \\
 &= h \int_0^2 (-s^2 + 2s) ds \\
 &= h \left(-\frac{s^3}{3} + s^2 \right) \Big|_0^2 \\
 &= \frac{4}{3} h
 \end{aligned}$$

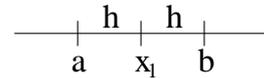
$$\alpha_2 = \frac{h}{3}$$

2. offene Newton-Cotes-Formeln:

Nehme Knoten x_1, \dots, x_{n-1}

$$I(f) = \sum_{j=1}^{n-1} \beta_j f_j, \quad \beta_i := h \int_p^1 \prod_{\substack{k=1 \\ k \neq j}}^{n-1} \frac{s-k}{j-k} ds$$

$$\underline{n=2}: \beta_1 = h \int_0^2 ds = 2h$$



$$E(f) := \int_a^b f(x) dx - \underbrace{\sum_{i=0}^n \omega_i f_i}_{I(f)}$$

$$\text{Definiere } a_+ = \begin{cases} |a| & , a \geq 0 \\ 0 & , a \leq 0 \end{cases} = \frac{1}{2}(|a| + a), \quad a \in \mathbb{R}$$

$$2a_+ = |a| + a$$

Satz 6.1 Peano

Sei $f \in C^{m+1}(I)$, $I = [a, b]$, $\forall p \in \mathbb{P}_m : E(p) = 0$

$$\Rightarrow E(f) = \int_a^b f^{(m+1)}(z) K(z) dz$$

$$K(z) := \frac{1}{m!} E_x([(x-t)_+]^m), t \in \mathbb{R}$$

$$(6.3) \\ \text{(Peano - Kern)}$$

Beweis:

Taylor-Reihe: $x \in I = (a, b]$

6 Numerische Quadratur

$$\Rightarrow f(x) = f(a) + (x-a)f'(a) + \dots + \frac{(x-a)^m}{m!} f^{(m)}(a) + \frac{1}{m!} \underbrace{\int_a^x f^{(m+1)}(t)(x-t)^m dt}_{= \int_a^b f^{(m+1)}(t) [(x-t)_+]^m dt}$$

$= 0 \Leftrightarrow x-t \leq 0$
 $\Leftrightarrow t \geq x$

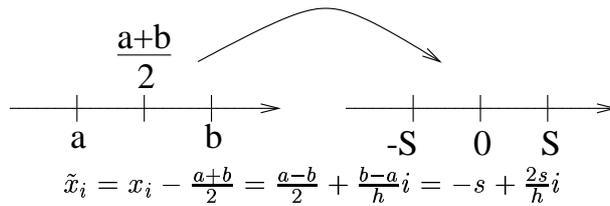
$$\begin{aligned} E(p) &= 0 \quad \forall p \leq \mathbb{P}_m \\ \Rightarrow E(f) &= \frac{1}{m!} E_x \int_a^b f^{(m+1)}(t) [(x-t)_+]^m dt \\ &= \frac{1}{m!} \left\{ \int_a^b \int_a^b f^{(m+1)}(t) [(x-t)_+]^m dt dx - \sum_{i=0}^n w_i \left(\int_a^b f^{(m+1)}(t) [(x_i-t)_+]^m dt \right) \right\} \\ &= \frac{1}{m!} \int_a^b f^{(m+1)}(t) \left\{ \underbrace{\int_a^b [(x-t)_+]^m dx - \sum_{i=0}^n w_i [(x_i-t)_+]^m}_{= E_x([(x-t)_+]^m) =: K(t)} \right\} dt \end{aligned}$$

⇒ □

Ab jetzt, zur Vereinfachung:

$$J = [-s, s], \quad s = \frac{b-a}{2}$$

$$g(\tilde{x}) = f\left(\tilde{x} + \frac{a+b}{2}\right) \Rightarrow \int_a^b f(x) dx \stackrel{\tilde{x} = x - \frac{a+b}{2}}{=} \int_{-s}^s g(\tilde{x}) d\tilde{x}$$



Ab jetzt: $x_i = -s + \frac{2s}{h} * i$

Berechnung der geschlossenen Newton-Cotes-Formel auch wie folgt:

$v_i(x) = x^i, i = 0, \dots, m, \{v_0, \dots, v_m\}$ eine Basis von \mathbb{P}_m .

Wir wollen: $E(p) = 0 \quad \forall p \in \mathbb{P}_m$,

d.h. $E(v_i) = 0 \quad \forall i = 0, \dots, m \Leftrightarrow$

$$\int_{-s}^s x^i dx = \sum_{\nu=0}^n w_\nu x_\nu^i, \quad \forall i = 0, \dots, m$$

d.h.

$$\begin{aligned} w_0 + w_1 + \dots + w_n &= 2s & (i=0) \\ x_0 w_0 + x_1 w_1 + \dots + x_n w_n &= 0 & (i=1) \\ x_0^2 w_0 + x_1^2 w_1 + \dots + x_n^2 w_n &= \frac{2s^2}{3} & (i=2) \\ \vdots & & \\ x_0^m w_0 + x_1^m w_1 + \dots + x_n^m w_n &= (1 - (-1)^{m+1}) \frac{s^{m+1}}{m+1} & (i=m) \end{aligned}$$

⇒ LGS $A\vec{x} = \vec{s}$, für $m = n$ ist A die Vandermondsche Matrix.

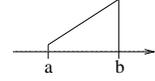
$$A = \begin{pmatrix} 1 & \dots & 1 \\ x_0 & \dots & x_n \\ \vdots & & \vdots \\ x_0^n & \dots & x_n^n \end{pmatrix} \Rightarrow \text{eindeutige Gewichte } \omega_i !$$

Beispiel: (Trapezregel)

6 Numerische Quadratur

(=geschlossene Newton-Cotes-Formel für $n = 1$, hier: $x_0 = -s$, $x_n = s$)

$$\text{LGS: } \left. \begin{array}{l} w_0 + w_1 = 2s \\ -sw_0 + sw_1 = 0 \end{array} \right\} \Rightarrow w_0 = w_1 = s = \frac{b-a}{2}$$



$$E^T(f) = \int_a^b f(x)dx - \frac{b-a}{2}(f(a) + f(b))$$

Frage: $E(x^2) = 0$? Nein, siehe Übung.

Beispiel: (Simpson-Regel)

(=abgeschlossene NCF für $n = 2$, $x_0 = -s$, $x_1 = 0$, $x_2 = s$)

$$\text{LGS: } \left. \begin{array}{l} w_0 + w_1 + w_2 = 2s \\ -sw_0 + sw_2 = 0 \\ s^2w_0 + s^2w_2 = \frac{2}{3}s^2 \end{array} \right\} \Rightarrow w_0 = w_2 = \frac{s}{3}, w_1 = \frac{4}{3}s$$

$$E^S(f) = \int_a^b f(x)dx - \frac{b-a}{6}(f(a) + 4f(\frac{a+b}{2}) + f(b)) = \int_{-s}^s g(x)dx - \frac{1}{3}(g(-s) + g(0) + g(s))$$

Klar: $E(x^2) = 0$

Frage: $E(x^3) = 0$? Ja! Siehe Übung.

Fehler der Simpson-Regel ? $\underbrace{[a, b]}_I \rightarrow \underbrace{[-s, s]}_J$

Beispiel: (Mittelpunktregel)

(=offene NCF, $n = 2$, $m = 0$, $w_0 = 2s$)

Quadraturregel mit Peano ausrechnen:

Trapezregel ($m = 1$)

$$E_x^T[(x-t)_+]^1 = \int_{-s}^s (x-t)_+ dx - s[(-s-t)_+ + \underbrace{(s-t)_+}_0], t \in (-s, s)$$

$$= \underbrace{\int_t^s (x-t)dx}_{= \frac{s^2}{2} - ts + \frac{t^2}{2}} - s(s-t) = -\frac{1}{2}(s-t)(s+t) = -\frac{s^2-t^2}{2} = k(t)$$

$$\Rightarrow E^T(g) = -\frac{1}{2} \int_{-s}^s s g''(t) \underbrace{(s^2 - t^2)}_{\geq 0} dt = -\frac{1}{2} g''(\eta) \int_{-s}^s (s^2 - t^2) dt$$

nach dem 2. Mittelwertsatz für Integrale

$$= -g''(\eta) \frac{2}{3} s^3, \eta \in (-s, s)$$

$$\Rightarrow E^T(f) = -\frac{1}{12} f''(\xi) (b-a)^3, \xi \in (a, b)$$

Mit der Simpson-Regel folgt: $E^S(f) = -\frac{1}{90} f^{(4)}(\xi) (\frac{b-a}{2})^5 = -\frac{1}{2880} f^{(4)}(\xi) (b-a)^5$

Mittelpunktregel: $E^M(f) = f^{(2)}(\frac{b-a}{3})$

(siehe auch Übung 10, dort durch 24 !!!!)

Simpson:

6 Numerische Quadratur

$$\begin{aligned}
 E_x^s[(x-t)_+]^3 &= \int_{-s}^s [(x-t)_+]^3 dx - \frac{s}{3} \{ \overbrace{[(-s-t)_+]^3}^{=0} + 4(-t)_+^3 + (s-t)_+^3 \} \\
 &= \int_t^s (x-t)^3 dx - \frac{s}{3} [2(|t|^3 - t^3) + (s-t)^3] \\
 &= \underbrace{\frac{(s-t)^4}{4}}_{\leq 0, \text{ da } s-t \leq s} - \frac{s}{3}(s-t)^3 - \frac{2s}{3} \underbrace{(|t|^3 - t^3)}_{\geq 0} \\
 &\Rightarrow k/t \leq 0 \text{ f\"ur } t \in J
 \end{aligned}$$

$$[g(\tilde{x}) = f(\hat{x} + \frac{a+b}{2})] \Rightarrow \int_a^b f(x) dx = \int_{-s}^s g(\tilde{x}) d\tilde{x}$$

$$\begin{aligned}
 E^s(g) &= \int_{-s}^s g^{(4)}(t) k(t) dt && (6.11) \\
 &= g^{(4)}(\eta) \int_{-s}^s k(t) dt \text{ nach dem 2. Mittelwertsatz.}
 \end{aligned}$$

2. Mittelwertsatz:

r, s stetig in $a \leq x \leq b$

$s(x)$ wechselt Vorzeichen nicht in $[a, b]$

$$\Rightarrow \int_a^b r(x) dx = r(\xi) \int_a^b s(x) dx$$

mit $\xi \in (a, b)$

$$g(t) := t^4, \quad E^s(t^4) = \int_{-s}^s \underbrace{g^{(4)}(t)}_{4!} k(t) dt = \frac{-4}{15} s^5$$

$$E(t^4) = [\int_{-s}^s t^4 dt - \frac{s}{3}((-s)^4 + 0 + s^4)] = [\frac{2}{5}s^5 - \frac{2}{3}s^5] = \frac{-4}{15}s^5$$

$$\int_{-s}^s k(t) dt = \frac{1}{4!} * \frac{-4}{15} s^5$$

$$E^s(g) = -g^{(4)}(\eta) \frac{1}{90} s^5$$

$$E^s(f) = -\frac{1}{90} f^{(4)}(\xi) (\frac{b-a}{2})^5, \quad \xi \in [a, b] \tag{6.12}$$

Quadraturfehler der Simpson-Regel

Entsprechend gilt f\"ur die Mittelpunktsformel:

$$E^M(f) = f^{(2)}(\xi) \frac{(b-a)^3}{2}$$

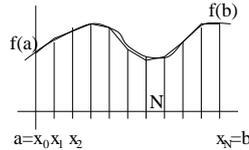
(siehe auch \u00dcbung 10, dort durch 24 !!!)

Entsprechend gilt f\"ur die Trapezregel:

$$E^T(f) = -\frac{1}{12} f''(\xi) (b-a)^3 \tag{6.10}$$

7 Zusammengesetzte Quadraturformeln

Bemerkung: Größere Genauigkeit mit numerischer Intergration (bei äquidistanten Knoten) kann nicht durch höhere Quadraturformeln erreicht werden (vgl. Interpolation)
Stattdessen: Partition des Integrationsintervalls



$x_\nu := a + \nu h$, $0 \leq \nu \leq N$ mit $h := \frac{b-a}{N}$ (äquidistant)

Mit der Trapezregel T angewandt auf jedes Intervall $I_\nu := [x_{\nu-1}, x_\nu]$, $1 \leq \nu \leq N$, erhalten wir als Approximation von $\int_a^b f(x) dx$:

$$T(h) = \sum_{\nu=1}^N T_{I_\nu}(h) = \sum_{\nu=0}^N \frac{h}{2} [f(x_{\nu-1}) + f(x_\nu)] \\ = h \left[\frac{1}{2} f_0 + f_1 + f_2 + \dots + f_{N-1} + \frac{1}{2} f_N \right] \quad (7.1)$$

Zusammengesetzte Trapezregel
(composite trapezoidal rule)

Entsprechend für die Simpson-Regel:

$$S(h) := \sum_{\nu=1}^N S_{I_\nu}(h) = \sum_{\nu=1}^N \frac{h}{6} \left(\underbrace{f(x_{\nu-1})}_{=: f_{\nu-1}} + 4 \underbrace{f\left(x_{\nu-1} + \frac{h}{2}\right)}_{=: f_{\nu-\frac{1}{2}}} + \underbrace{f(x_\nu)}_{f_\nu} \right) \\ = \frac{h}{3} \left[\frac{1}{2} f_0 + 2f_{\frac{1}{2}} + f_1 + 2f_{\frac{3}{2}} + \dots + f_{N-1} + 2f_{N-\frac{1}{2}} + \frac{1}{2} f_N \right] \quad (7.2)$$

zusammengesetzte Simpson-Regel

Aus (6.10) folgt

$$\left| \int_a^b f(x) dx - T(h) \right| = \left| \sum_{\nu=1}^N \frac{h^3}{12} f''(\xi_\nu) \right| \quad (7.3) \\ \leq (Nh) \frac{h^2}{12} \max_{x \in I=(a,b)} |f''(x)| \\ = \frac{b-a}{12} \|f''\|_{I,\infty} h^2 \\ = O(h^2)$$

mit $\xi_\nu \in I_\nu = (x_{\nu-1}, x_\nu)$ und falls $f \in C^2$

$$\left| \int_a^b f(x) dx - S(h) \right| = \left| \sum_{\nu=1}^N \frac{h^5}{25 \cdot 90} f^{(4)}(\xi) \right| \quad (7.4) \\ \leq (Nh) \frac{h^4}{25 \cdot 90} \max_{x \in I} |f^{(4)}(x)| \\ = \frac{b-a}{180} \|f^{(4)}\|_{I,\infty} \left(\frac{h}{2}\right)^4 \\ = O(h^4)$$

falls $f \in C^4$

Falls f glatter ist, nehme asymptotische Entwicklung des Fehlers bezüglich $h = \frac{b-a}{N}$

Satz 7.1

Sei $f \in C^{2k+2}(I)$. Dann existieren falls k klein genug (d.h. N groß genug) Konstanten $a_1, a_2, \dots, a_{k+1} \in \mathbb{R}$ unabhängig von h und $b(h) \in \mathbb{R}$ mit

$$ET(h) := T(h) - \int_a^b f(x) dx = a_1 h^2 + a_2 h^4 + \dots + a_k h^{2k} + b(h) h, \quad a_k \in \mathbb{R} \\ \text{mit } |b(h)| \leq |a_{k+1}| \|f^{(2k+2)}\|_{I,\infty}$$

Merke:

Diese Entwicklung wird benutzt zur genaueren numerischen Berechnung. Man berechnet $T(h)$ für verschiedene Schrittweiten h und extrapoliert dann für $h = 0$. (Richardson-Extrapolation)

Oft gebrauchte Version: In der Romberg Integration beginnt man mit einer festen Schrittweite h_0 und nimmt nacheinander die Hälfte davon, d.h. $h_\nu = 2^{-\nu}h$ (7.6)

und berechnet das zugehörige $T(h_\nu) =: T_{\nu 0}$.

nach (7.5) folgt die $T_{\nu 0}$ also einer asymptotischen Entwicklung und mit $I := \int_a^b f(x)dx$ gilt:

$$I_{\nu 0} = I + a_1 h^2 2^{-2\nu} + a_2 h^4 2^{-4\nu} + \dots + a_k h^{2k} 2^{-2k} + O(h^{2k+2}) \quad (\heartsuit)$$

Wir sehen, daß eine Linearkombination

$$T_{\nu 1} := \frac{1}{3}(4T_{\nu 0} - T_{\nu-1,0})$$

einen Nullterm vor h^2 bei \heartsuit produziert.

$$T_{\nu 1} = I + a_1 h^2 \underbrace{\left(\frac{4 \cdot 2^{-2\nu} - 2^{-2(\nu-1)}}{3} \right)}_{=0} + \dots$$

$$T_{\nu 1} = I + a_2^{(1)} h_\nu^4 + a_3^{(1)} h_\nu^6 + \dots + a_k^{(1)} h_\nu^{2k} + O(h^{2k+2})$$

Jetzt produziere man eine Null für $a_2^{(1)}$, u.s.w..

Im Allgemeinen:

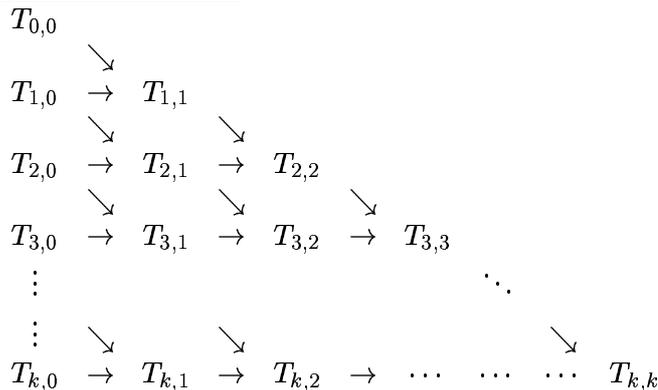
$$T_{\nu, \mu} := \frac{2^{2\mu} T_{\nu, \mu-1} - T_{\nu-1, \mu-1}}{2^{2\mu} - 1}, \quad \nu \geq \mu \quad (7.7)$$

Jedesmal erhalten wir eine Entwicklung

$$I_{\nu, \mu} = I + a_{\mu+1}^{(\mu)} h_\nu^{2(\mu+1)} + \dots + O(h^{2k+2})$$

bis zu $T_{k,k} = O(h^{2k+2})$ einschließlich.

Schema für (7.7)



Gauß-Quadratur

Newton-Cotes Formeln: äquidistante Knoten

Frage: Können wir Polynome selbst höheren Grades durch geeignete Wahl der Knoten exakt integrieren?

Optimale Wahl der Knoten wird in den Gauß-Quadratur-Formeln gewonnen.

7 Zusammengesetzte Quadraturformeln

$$\int_a^b f(x) dx \approx \sum_{i=1}^n \omega_i f_i \quad (1)$$

G.Q.F. sind exakt für Polynome P_{2n-1} vom Grad $2n - 1$.

Lemma: Seien p_1, p_2, \dots, p_m Polynome derart, daß $\text{span}\{p_1, p_2, \dots, p_m\} = \mathbb{P}_{m-1}$. Für paarweise verschiedene $x_\nu \in \mathbb{R}$ ($\nu = 1, \dots, m$) sei

$$A := \begin{pmatrix} p_1(x_1) & p_1(x_2) & \cdots & p_1(x_m) \\ p_2(x_1) & p_2(x_2) & & p_2(x_m) \\ \vdots & & & \vdots \\ p_m(x_1) & p_m(x_2) & \cdots & p_m(x_m) \end{pmatrix} \quad (2)$$

$$\Rightarrow \det A \neq 0 \quad (3)$$

(d.h. $\{p_1, p_2, \dots, p_m\}$ bilden ein Haar-System)

Nehme (von jetzt an) $I := [-1, 1]$, Reduktion auf allgemeinen Fall. Später durch lineare Transformation. Als eine Basis des Raumes \mathbb{P}_{m-1} benutze jetzt (Grund wird klar durch folgenden Satz) die normierten Legendre Polynome $L_1(x) = \frac{1}{\sqrt{2}}$, $L_2(x) = \sqrt{\frac{3}{2}}x, \dots, L_m$. Falls $p_k \in \mathbb{P}_k$ k Nullstellen $y_1, y_2, \dots, y_k \in \mathbb{R}$ hat (gezählt nach Vielfachheit), so hat $p_k^{(\nu)}$ $k - \nu$ Nullstellen $\in \text{conv}(y_1, \dots, y_k)$ $\nu \leq k$.

Satz 7.2

$x_1, x_2, \dots, x_n \in I = [-1, 1]$ seien Nullstellen von L_{n+1} und $\omega \in \mathbb{R}^n$ mit

$$A\omega = \begin{pmatrix} \sqrt{2} \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad (4)$$

mit A aus (2) und mit L_ν statt p_ν ($\nu = 1, \dots, n$) Dann gilt: $\omega_i > 0$ für $i = 1, \dots, n$ (5)

und $\forall p \in \mathbb{P}_{2n-1} \quad \int_{-1}^1 p(x) dx = \sum_{i=1}^n \omega_i p(x_i)$ (6)

Beweis: Nach Lemma gibt es ein eindeutiges ω aus (4), da $\det A \neq 0$.

Nach dem Euklidischen Algorithmus (für Polynome) kann ein beliebiges

Polynom $p \in \mathbb{P}_{2n-1}$ dargestellt werden als:

$$p = L_{n+1}q + r \text{ mit } q, r \in \mathbb{P}_{n-1} = \text{span}\{L_1, \dots, L_n\} \quad (7)$$

$$\Rightarrow q = \sum_{k=1}^n \alpha_k L_k, \quad r = \sum_{k=1}^n \beta_k L_k \quad (8)$$

$$\Rightarrow \int_I p(x) dx = \int_I L_{n+1}(x)q(x) dx + \underbrace{\sqrt{2} \int_I L_1(x) r(x) dx}_{(da \ L_1(x) = \frac{1}{\sqrt{2}})}$$

$$=: \underbrace{\langle L_{n+1}, q \rangle}_{=0, \text{ da } L_{n+1} \perp \text{span}\{L_1, \dots, L_n\} \in q} + \sqrt{2} \langle L_1, r \rangle \quad (9)$$

x_i Nullstellen von L_{n+1}

$$\begin{aligned} \sum_{i=1}^n \omega_i p(x_i) &\stackrel{(7)}{=} \sum_{i=1}^n \omega_i r(x_i) \stackrel{(8)}{=} \sum_{k=1}^n \beta_k \sum_{i=1}^n \omega_i L_k(x_i) \stackrel{(4)}{=} \sum_{k=1}^n \beta_k \sqrt{2} \delta_{k1} \\ &= \sqrt{2} \beta_1 \stackrel{(8)}{=} \sqrt{2} \langle L_1, r \rangle \end{aligned} \quad (10)$$

7 Zusammengesetzte Quadraturformeln

Kronecker Delta $\delta_{jk} = \begin{cases} 1, j = k \\ 0, \text{sonst} \end{cases}$

$\stackrel{(10)}{\Rightarrow} \stackrel{(9)}{(6)}$

Einsetzen in (6) für p die Polynome $q_i(x) = \prod_{k=1, k \neq i}^n (x - x_k)^2 \in \mathbb{P}_{2n-2}$
 $\Rightarrow q_i(x_i) > 0 \Rightarrow \omega_i > 0$ □

n=1 x_i Nullstelle von $L_2(x) = \sqrt{\frac{3}{2}}x \Rightarrow x_1 = 0$

(4) $\frac{1}{\sqrt{2}}\omega_1 = \sqrt{2} \Rightarrow \omega_1 = 2$ d.h. $\sum \dots = 2 * p(0)$ Mittelpunktsregel

n=2 x_1, x_2 Nullstellen von $L_3(x) = \frac{\sqrt{3}}{8}(3x^2 - 1) \Rightarrow x_1 = \frac{-1}{\sqrt{3}}, x_2 = \frac{1}{\sqrt{3}}$

$$(4) \left. \begin{aligned} &\frac{1}{\sqrt{2}}\omega_1 + \frac{1}{\sqrt{2}}\omega_2 = \sqrt{2} \\ &\underbrace{\sqrt{\frac{3}{2}} \left(-\frac{1}{\sqrt{3}}\right)}_{L_2(x_1)} \omega_1 + \underbrace{\sqrt{\frac{3}{2}} \left(\frac{1}{\sqrt{3}}\right)}_{L_2(x_2)} \omega_2 = 0 \end{aligned} \right\} \Rightarrow \omega_1 = \omega_2 = 1$$

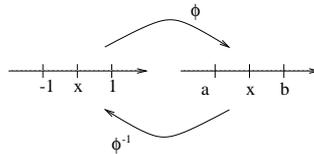
d.h. $\sum \dots = 1p\left(-\frac{1}{\sqrt{3}}\right) + 1p\left(\frac{1}{\sqrt{3}}\right)$ exakt für $p \in \mathbb{P}_3$.

Für das Intervall $[a, b]$ nehme man eine lineare Transformation vor:

$$t = \phi(x) = \frac{b+a}{2} + \frac{b-a}{2}x \in [a, b] \text{ für } x \in I := [-1, 1]$$

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

Rücktransformation



Fehler der Gauß -Quadratur:

$f : [a, b] \rightarrow \mathbb{R}$

$$E^G f(\phi(\dots)) = \frac{b-a}{2} \int_{-1}^1 f(\phi(x)) dx - \frac{b-a}{2} \sum_{i=1}^n \omega_i f(\phi(x_i)) = \int_a^b f(t) dt - \frac{b-a}{2} \sum_{i=1}^n \omega_i f(t_i)$$

Gauß -Quadratur für beliebiges Intervall (exakt für Polynome vom Grad $< 2n$)

Lemma: Sei $f \in C^{2n}(I)$. Dann existiert eine Konstante $c_n > 0$, unabhängig von a, b, f so, daß

$$|E^G f(\phi(\cdot))| \leq c_n (b-a)^{2n+1} \underbrace{\|f^{(2n)}\|_{I, \infty}}_{\max_{x \in [-1, 1]} |f^{(2n)}(x)|}$$

Beweis:

Setze $E^G(f(\phi(\dots))) := \frac{2}{b-a} E^G(f(\phi(\dots))) \Rightarrow F^G(p) = 0$, denn $E^G(p) = 0 \forall p \in \mathbb{R}_{2n-1}$

Nach dem Satz von Peano gilt:

$E^G f(\phi(\dots)) = \int_{-1}^1 \left\{ \frac{d^{2n}}{ds^{2n}} f(\phi(s)) \right\} K(s) ds$ mit $\|K\|_{[-1, 1], \infty} =: K_0 = \text{const}$ unabhängig von a, b, f .

$$|F^G f(\phi(\dots))| \leq K_0 \int_{-1}^1 |f^{(2n)}(\phi(s))| * \underbrace{\left| \frac{d\phi}{ds} \right|^{(2n)}}_{\left(\frac{b-a}{2}\right)^{2n}} ds$$

7 Zusammengesetzte Quadraturformeln

$$\leq K_0 \|f^{(2n)}\|_{I,\infty} \left(\frac{b-a}{2}\right)^{(2n)} * 2$$

⇒ Behauptung mit $C_n = \frac{K_0}{2^{2n}}$. □

Beispiel:

Approximiere $\int_{-1}^1 x \sin(x) dx$ durch Gauß -Quadratur mit den Nullstellen von L_4 .

(L_4 hat die Nullstellen $x_1 = 0, x_{2,3} = \pm \sqrt{\frac{3}{5}}$).

Nehme $L_1 = \frac{1}{\sqrt{2}}, L_2 = \sqrt{\frac{3}{2}}, L_3 = \frac{1}{2} * \sqrt{\frac{5}{2}}(3x^2 - 1)$

$$A\omega = \begin{pmatrix} \sqrt{2} \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad A = \begin{pmatrix} L_1(x_1) & L_1(x_2) & L_1(x_3) \\ L_2(x_1) & L_2(x_2) & L_2(x_3) \\ L_3(x_1) & L_3(x_2) & L_3(x_3) \end{pmatrix}$$

$$\frac{1}{\sqrt{2}}\omega_1 + \frac{1}{\sqrt{2}}\omega_2 + \frac{1}{\sqrt{2}}\omega_3 = \sqrt{2}$$

$$0 + \sqrt{\frac{3}{2}}\sqrt{35}\omega_2 + \sqrt{35}\sqrt{3}2\omega_3 = 0$$

$$-\sqrt{58}\omega_1 + \sqrt{58}(3 * \frac{3}{5} - 1)\omega_2 + \sqrt{\frac{5}{8}}(3 * \frac{3}{5} - 1)\omega_3 = 0$$

$$\Leftrightarrow \left. \begin{matrix} \omega_1 + \omega_2 + \omega_3 = 2 \\ \omega_2 - \omega_3 = 0 \\ -\omega_1 + \frac{4}{5}\omega_2 + \frac{4}{5}\omega_3 = 0 \end{matrix} \right\} \Rightarrow \begin{matrix} \omega_2 = \frac{5}{9} = \omega_3 \\ \omega_1 = \frac{5}{9} \end{matrix}$$

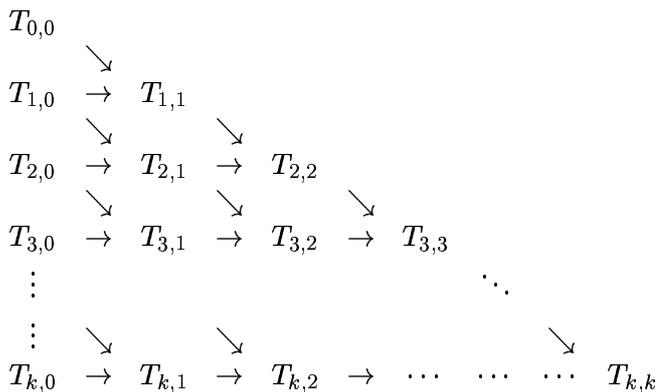
$$\int_{-1}^1 x \sin(x) dx \approx \sum_{i=1}^3 \omega_i f_i = \frac{8}{9} * 0 * \sin(0) + \frac{5}{9} * \sqrt{\frac{3}{5}} \sin \sqrt{\frac{3}{5}} + \frac{5}{9} (-\sqrt{\frac{3}{5}}) (-\sin \sqrt{\frac{3}{5}})$$

$$= \frac{10}{9} \sqrt{\frac{3}{5}} \sin \sqrt{\frac{3}{5}} = 0,60199.$$

Der exakte Wert liegt bei 0,60234.

Romberg benutzt die zusammengesetzte Trapezregel: $T_{\nu 0} := T(h_\nu)$

$$T_{\nu \mu} = \frac{2^{2\mu} T_{\nu, \mu-1} - T_{\nu-1, \mu-1}}{2^{2\mu} - 1}, \nu \geq \mu$$



Beispiel: $h_0 = 1, k = 2$

$$\int_1^2 \frac{dx}{x}$$

$$T_{0,0} = \frac{1}{2} \left(1 + \frac{1}{2}\right) = 0,75$$

$$T_{1,0} = \frac{1}{4} \left(1 + 2 * \frac{2}{3} + \frac{1}{2}\right) = 0,7083\bar{3}$$

7 Zusammengesetzte Quadraturformeln

$$T_{2,0} = \frac{1}{8} \left(1 + 2 * \frac{4}{5} + 2 * \frac{4}{7} + \frac{1}{2} \right) = 0,697024$$

$$T_{1,1} = \frac{2^2 T_{1,0} - T_{0,0}}{2^2 - 1} = 0,699\bar{9}$$

$$T_{2,1} = \frac{2^2 T_{2,0} - T_{1,0}}{3} = 0,693254$$

$$T_{2,2} = \frac{2^4 T_{2,1} - T_{1,1}}{15} = 0,69317604 \quad F : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^n, \quad F \in C(D) \text{ gegeben.}$$

Aufgabe: Finde alle $x \in D$ mit $F(x) = 0$.

Beispiel: ($n = 2$)

$$F(x) = \begin{pmatrix} 10x_1 + x_2 + \sin(x_1 + x_2) \\ x_1 + 10x_2 + \sin(x_1 - x_2) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad D = \mathbb{R}^2$$

Hier existiert genau ein $x^* \in \mathbb{R}^2$ mit $F(x^*) = 0$, nämlich $x^* = 0$.

Im allgemeinen sind alle Sorten von Lösungsmengen möglich (keine, genau ein, endlich viele, abzählbar viele, unendlich viele Lösungen).

Üblicherweise existiert genau eine Lösung $x^* \in D$, falls nötig macht man D entsprechend kleiner.

Numerische Methoden: Iterationsverfahren

Man benötigt einen Startwert $x^{(0)}$, eine Iterationsfunktion $\Phi : \mathbb{R}^n \rightarrow \mathbb{R}^n$ und es gilt

$$x^{(\nu+1)} = \Phi(x^{(\nu)}) \quad \text{Einschrittverfahren}$$

Frage: Implizieren $x^{(0)}$ und Φ eine (gegen x^*) konvergierende Iterationsfolge ?

7.1 Picard Iteration

Transformiere $F(x) = 0$ in eine äquivalente Fixpunktgleichung:

$$f(x) = x, \quad f : D \rightarrow \mathbb{R}^n. \quad \text{Nehme } \Phi := f \Rightarrow x^{(\nu+1)} = f(x^{(\nu)}), \quad \nu = 0, 1, 2, \dots \quad (1)$$

Picard-Iteration

Beispiel:

$$f(x) = \frac{1}{10} \begin{pmatrix} x_2 + \sin(x_1 + x_2) \\ x_1 + \sin(x_1 - x_2) \end{pmatrix} \Rightarrow F(x) = 0 \Leftrightarrow f(x) = x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

Definition:

$$D \subset \mathbb{R}^n, \quad f : D \rightarrow \mathbb{R}^n,$$

$$f \text{ heißt } \underline{\text{Kontraktion}} : \Leftrightarrow \exists L \in \mathbb{R}, \quad L < 1 : \forall x, y \in D \quad \|f(x) - f(y)\| \leq L \|x - y\| \quad (2)$$

d.h. f ist Lipschitz-stetig in D mit $L < 1$.

Fundamental für die Theorie von Iterationsverfahren ist der folgende Satz.

Banach-Fixpunktsatz:

$D \subset \mathbb{R}^n$ abgeschlossen, $D \neq \emptyset$.

$f : D \rightarrow \mathbb{R}^n$ habe folgende Eigenschaften

1. $f(D) \subset D$ (Abbildung auf sich selbst)
2. f ist Kontraktion

7 Zusammengesetzte Quadraturformeln

⇒ Dann hat $f(x) = x$ genau eine Lösung $x^* \in D$ und die Picard-Iteration (1) konvergiert gegen x^* für beliebige Wahl de Startwertes $x^{(0)} \in D$.

Beweis:

7 Zusammengesetzte Quadraturformeln

$x^{(0)} \in D$ beliebig, $x^{(\nu+1)} = f(x^{(\nu)}) \stackrel{(1.)}{\Rightarrow} x^{(\nu)} \in D$ für alle $\nu \in \mathbb{N}$.
 $\|x^{(\nu+1)} - x^{(\nu)}\| = \|f(x^{(\nu)}) - f(x^{(\nu-1)})\| \leq L\|x^{(\nu)} - x^{(\nu-1)}\| \leq L^\nu \|x^{(1)} - x^{(0)}\|$
 wegen (2), (2.). (3)

Sei $p \in \mathbb{N}$. Mit der Dreieckungleichung folgt
 $\|x^{(\nu+p)} - x^{(\nu)}\| \leq \sum_{i=1}^n \|x^{(\nu+i)} - x^{(\nu+i-1)}\|$
 $\stackrel{(3)}{\leq} \|x^{(1)} - x^{(0)}\| \sum_{i=1}^p L^{\nu+i-1}$
 $= \|x^{(1)} - x^{(0)}\| L^\nu \sum_{i=1}^p L^{i-1}$ (4)
 $= \|x^{(1)} - x^{(0)}\| \underbrace{\frac{1-L^p}{1-L}}_{\sum_{i=1}^p L^{i-1}} L^\nu$

$\lim_{\nu \rightarrow \infty} \|x^{(\nu+p)} - x^{(\nu)}\| \leq \lim_{\nu \rightarrow \infty} \|x^{(1)} - x^{(0)}\| * L^\nu * \left(\frac{1-L^p}{1-L}\right) \rightarrow 0$

$\Rightarrow \{x_{\nu=0,1,2}^{(\nu)}\}$ bilden eine Fundamentalfolge (Cauchy-Folge) $\nu = 0, 1, 2$

D abgeschlossen \Rightarrow Limes $x^* \in D$ existiert, $x^* = \lim_{\nu \rightarrow \infty} x^{(\nu)}$

x^* ist Fixpunkt von f , da f stetig ist nach (2.), d.h.

$x^* = \lim_{\nu \rightarrow \infty} x^{(\nu)} = \lim_{\nu+1 \rightarrow \infty} f(x^{(\nu+1)}) \stackrel{\text{stetig}}{=} f(\lim_{\nu \rightarrow \infty} x^{(\nu+1)}) = f(x^*) \Rightarrow$ Existenz.

Seien x^*, y^* zwei Fixpunkte, dann gilt nach (2.):

$$\|x^* - y^*\| = \|f(x^*) - f(y^*)\| \leq \underbrace{L}_{<1} \|x^* - y^*\| = 0 \Rightarrow \text{Eindeutigkeit} \quad \square$$

Korollar:

Unter den Voraussetzungen aus dem Banach-Fixpunktsatz gilt für $\nu \in \mathbb{N}$:

$$\|x^{(\nu)} - x^*\| \leq \frac{L^\nu}{1-L} \|x^{(1)} - x^{(0)}\| \quad (5)$$

“a-priori”-Fehlerabschätzung

$$\|x^{(\nu)} - x^*\| \leq \frac{L}{1-L} \|x^{(\nu)} - x^{(\nu-1)}\| \quad (6)$$

“a-posteriore”-Fehlerabschätzung

Beweis:

(5) folgt aus (4) mit $\underbrace{\lim_{p \rightarrow \infty} \|x^{(\nu+p)} - x^{(\nu)}\|}_{\|x^{(\nu)} - x^*\|} \leq \frac{L^\nu}{1-L} \|x^{(1)} - x^{(0)}\|$

(6) folgt aus (4) durch dortiges Einsetzen von $\nu = 1$

und $x^{(\mu-1)}, x^{(\mu)}, x^{(\mu+p)}$ für $x^{(0)}, x^{(1)}, x^{(p+1)}$.

$$\|x^{(1+p)} - x^{(1)}\| \leq L \left(\frac{1-L^p}{1-L}\right) \|x^{(1)} - x^{(0)}\| \Rightarrow \|x^{(\mu+p)} - x^{(\mu)}\| \leq L \frac{1-L^p}{1-L} \|x^{(\mu)} - x^{(\mu-1)}\|$$

dann $\lim_{p \rightarrow \infty}$ und $\mu := \nu$ □

Definition: $D \in \mathbb{R}^n, f : D \rightarrow \mathbb{R}^n$

f Kontraktion $:\Leftrightarrow \exists L < 1 \forall x, y \in D : \|f(x) - f(y)\| \leq L\|x - y\|$

Satz 7.3

$D \in \mathbb{R}^n, D \neq \emptyset$ konvex, $f \in C^1(D)$ mit $\forall x \in D : \|f'(x)\| \leq L$ mit $L < 1$.

(in einer Matrixnorm induziert durch durch die Vektornorm in \mathbb{R}^n .)

\Rightarrow Dann ist f eine Kontraktion in D .

7 Zusammengesetzte Quadraturformeln

Beweis: (f' := Funktionalmatrix von f)

D konvex, d.h. $x, y \in D \Rightarrow x + t(y - x) \in D$ für $t \in [0, 1]$, $g : [0, 1] \rightarrow \mathbb{R}^n$ mit $g(t) := f(x + t(y - x))$, $t \in [0, 1] \Rightarrow g \in C^1[0, 1]$.

$$f(y) - f(x) = g(1) - g(0) = \int_0^1 g'(s) ds \stackrel{\text{Kettenregel}}{=} \int_0^1 f'(x + s(y - x))(y - s) ds$$

$$\|f(y) - f(x)\| \leq \int_0^1 \|f'(x + s(y - x))(y - s)\| ds$$

$$\leq \int_0^1 \|f'(\underbrace{x + s(y - x)}_{\in D})\| \|y - x\| ds$$

$$\underbrace{\hspace{10em}}_{\leq L}$$

$$\leq L \|y - x\| \quad \square$$

Beispiel: $f(x) = -\frac{1}{10} \begin{pmatrix} x_2 + \sin(x_1 + x_2) \\ x_1 + \sin(x_1 - x_2) \end{pmatrix}$, $f(x) = x$ $f(x_1, x_2) = \begin{pmatrix} f_1(x_1, x_2) \\ f_2(x_1, x_2) \end{pmatrix}$

$$f'(x) := \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} \end{pmatrix} = -\frac{1}{10} \begin{pmatrix} \cos(x_1 + x_2) & 1 + \cos(x_1 + x_2) \\ 1 + \cos(x_1 - x_2) & -\cos(x_1 - x_2) \end{pmatrix}$$

$\|f'(x)\|_Z \leq \frac{1}{10}(1 + 2) = 0.3 = L < 1$ Kontraktion in Zeilensummennorm (Banach-Fixpunktsatz)

$$\left. \begin{array}{l} D \in \mathbb{R}^n \text{ abgeschlossen} \\ f(D) \subset D, f \text{ Kontraktion} \end{array} \right\} \Rightarrow x^{(\nu+1)} = f(x^{(\nu)})$$

konvergiert gegen $x^* = f(x^*) \in D \quad \forall x^{(0)} \in D$

Hinweis: In Problemen im \mathbb{R}^n ist speziell $f(D) \subset D$ oft schwierig zu überprüfen, aber sie ist -wie gesagt- nötig für globale Konvergenz,

z.B. in Kugel $K_r := \{x \in \mathbb{R}^n : \|x - x^*\| < r\}$.

Definition: (Konvergenzordnung)

Die Einschmittmethode $x^{(\nu+1)} = \emptyset(x^{(\nu)})$, Startvektor $x^{(0)}$, heißt konvergent von der Ordnung $p \in \mathbb{N} : \Leftrightarrow \exists m \in \mathbb{N}, M \in \mathbb{R} : \|x^{(\nu+1)} - x^*\| \leq M \|x^{(\nu)} - x^*\|^p, \nu \geq m$.

Satz 7.4

$f \in C^p(K_\delta(x^*))$ für ein $p \in \mathbb{N}, p > 0$ und

$\forall |\alpha| \leq p - 1 : D^\alpha f|_{x=x^*} = 0 \quad \heartsuit$

Dann konvergiert die Picard-Iteration $x^{(\nu+1)} = f(x^{(\nu)})$ von der Ordnung p .

Beweis: ($p = 2$) lokale Konvergenz (gesichert) \Leftrightarrow

$\exists \tau > 0$ mit globaler Konvergenz in $\overline{K_\tau(x^*)}$. Sei $y \in \overline{K_\tau(x^*)}$, $g(t) := f(x^* + t(y - x^*))$.

$$f(y) - x^* = f(y) - f(x^*) = \int_0^1 g'(t) dt = \int_0^1 [g'(t) - \underbrace{g'(0)}_{=0, \text{ wg. } \heartsuit}] dt = \int_0^1 \int_0^1 g''(s) ds dt$$

$$\leq \int_0^1 \int_0^1 \underbrace{\|f''\|_Z}_{\leq M \text{ wg. } f \in C^2(K_\delta(x^*))} \|y - x^*\|^2 ds dt \leq M \|y - x^*\|^2$$

$$\Rightarrow \|f(y) - x^*\|_Z \leq M \|y - x^*\|_Z^2 \quad (y := x^\nu, f(y) := x^{\nu+1})$$

7.2 Newton-Verfahren

$F : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^n, F \in C^2(D), F(x^*) = 0$

Finde x^* . Hauptidee: Linearisiere $F(x)$ im Punkt $x^{(\nu)}$:

$$F(x^*) = F(x^{(\nu)}) + F'(x^{(\nu)})(x^* - x^{(\nu)}) + \underbrace{0(\|x^* - x^{(\nu)}\|^2)}_{2\text{-te Abl. von } F \text{ (Hesse-Matrix)}} \stackrel{(!)}{=} 0$$

Term $0(\dots)$ vernachlässigen.

\Rightarrow Nehme $F(x^{(\nu)}) + F'(x^{(\nu)})(x^{(\nu+1)} - x^{(\nu)}) = 0$ statt volle Gleichungen $F(x) = 0$.

$\Rightarrow F'(x^{(\nu)})(x^{(\nu+1)} - x^{(\nu)}) = -F(x^{(\nu)}), \nu = 0, 1, 2, \dots$ (1) *Newton - Verfahren*

Lineares Gleichungssystem für $x^{(\nu+1)}$, gegeben: $x^{(\nu)}$.

(1) benötigt $\det F'(x^{(\nu)}) \neq 0 \forall \nu$

Explizit: $x^{(\nu+1)} = x^{(\nu)} - F'(x^{(\nu)})^{-1} F(x^{(\nu)})$ (2)

$n = 1 : x^{(\nu+1)} = x^{(\nu)} - \frac{F(x^{(\nu)})}{F'(x^{(\nu)})}$

$F : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^n, F \in C^2(D), F(x^*) = 0$

$F'(x^{(\nu)})(x^{(\nu+1)} - x^{(\nu)}) = -F(x^{(\nu)}) \nu = 0, 1, 2, 3, \dots$

Satz 7.5 (lokale Konvergenz des Newton-Verfahrens)

$D \subset \mathbb{R}^n$ offen, $F : \overline{D} \rightarrow \mathbb{R}^n, F \in C^2(\overline{D}),$

$\det F'(x) \neq 0 \forall x \in \overline{D}, F(x^*) = 0$ für ein $x^* \in D$

$\Rightarrow \exists \tau > 0$ derart, daß das Newton-Verfahren konvergiert für jedes $x^{(0)} \in K_\tau(x^*)$ und falls $F \in C^3(\overline{D})$ dann gilt:

$$\exists m \in \mathbb{N}, M \in \mathbb{R} : \|x^{(\nu+1)} - x^*\| \leq M \|x^{(\nu)} - x^*\|^2 \text{ für } \nu > m$$

\Rightarrow Das Newton-Verfahren konvergiert quadratisch

Beweis: (n=1, nur)

$$x^{(\nu+1)} = x^{(\nu)} - \frac{f(x^{(\nu)})}{f'(x^{(\nu)})}, (\nu = 0, 1, 2, \dots), f := \begin{cases} x^{(\nu)} \rightarrow x_n \\ x^{(\nu+1)} \rightarrow x_{n+1} \end{cases}$$

Taylor: $f(x) = f(x_n) + f'(x_n)(x - x_n) + \frac{1}{2}f''(\eta_n)(x - x_n)^2, \eta_n \in (x_n, x)$

Sei $f(\xi) = 0 \Rightarrow 0 = f(x_n) + f'(x_n)(\xi - x_n) + \frac{1}{2}f''(\eta_n)(\xi - x_n)^2$

$$\Rightarrow f'(x_n)(\xi - x_n) = -f(x_n) - \frac{1}{2}f''(\eta_n)(\xi - x_n)^2$$

$$\Rightarrow \xi = \underbrace{x_n - \frac{f(x_n)}{f'(x_n)}}_{=: x_{n+1}} - \frac{\frac{1}{2}f''(\eta_n)}{f'(x_n)}(\xi - x_n)^2$$

mit $\xi - x_n := e_n, \xi - x_{n+1} := e_{n+1}$

$$\Rightarrow e_{n+1} = -\frac{1}{2} \frac{f''(\eta_n)}{f'(x_n)} e_n^2$$

$$\Rightarrow \frac{|e_{n+1}|}{|\xi - x_{n+1}|} \leq \frac{1}{2} \frac{|f''(\eta_n)|}{|f'(x_n)|} |e_n|^2 \leq M$$

mit $|e_n|^2 = |\xi - x_n|^2$ □

$n = 2 : \begin{cases} F(\vec{x}) = 0 \\ \vec{x} = (x, y) \end{cases} \Leftrightarrow \begin{cases} f_1(x, y) = 0 \\ f_2(x, y) = 0 \end{cases} ; 0 = F(\vec{x}^{(\nu)}) + f'(\vec{x}^{(\nu)})(\vec{x}^{(\nu+1)} - \vec{x}^{(\nu)})$

$$\Leftrightarrow 0 = f_1(x^{(\nu)}, y^{(\nu)}) + \frac{\partial f_1}{\partial x}(x^{(\nu+1)} - x^{(\nu)}) + \frac{\partial f_1}{\partial y}(y^{(\nu+1)} - y^{(\nu)})$$

$$\Leftrightarrow 0 = f_2(x^{(\nu)}, y^{(\nu)}) + \frac{\partial f_2}{\partial x}(x^{(\nu+1)} - x^{(\nu)}) + \frac{\partial f_2}{\partial y}(y^{(\nu+1)} - y^{(\nu)})$$

$h := (x^{(\nu+1)} - x^{(\nu)})$

$k := (y^{(\nu+1)} - y^{(\nu)})$

7 Zusammengesetzte Quadraturformeln

Aufgabe: Finde h, k für jede Iteration !

$$\frac{\partial f_1}{\partial x} h + \frac{\partial f_1}{\partial y} k = -f_1(x^{(\nu)}, y^{(\nu)})$$

(siehe oben)

$$\frac{\partial f_2}{\partial x} h + \frac{\partial f_2}{\partial y} k = -f_2(x^{(\nu)}, y^{(\nu)})$$

Cramersche Regel:

$$J := \begin{vmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} \end{vmatrix} \neq 0 \text{ Jacobi-Determinante}$$

$$h = \frac{1}{J} \begin{vmatrix} -f_1(x^{(\nu)}, y^{(\nu)}) & \frac{\partial f_1}{\partial y} \\ -f_2(x^{(\nu)}, y^{(\nu)}) & \frac{\partial f_2}{\partial y} \end{vmatrix}, \quad k = \frac{1}{J} \begin{vmatrix} \frac{\partial f_1}{\partial x} & -f_1(x^{(\nu)}, y^{(\nu)}) \\ \frac{\partial f_2}{\partial x} & -f_2(x^{(\nu)}, y^{(\nu)}) \end{vmatrix}$$

$$\Rightarrow x^{(\nu+1)} = x^{(\nu)} + h, \quad y^{(\nu+1)} = y^{(\nu)} + k$$

\Rightarrow nächste Approximation existiert.

Vorteil: Konvergenz zweiter Ordnung.

Nachteil: $J \neq 0$ schwierig zu überprüfen im laufenden Programm.

\Rightarrow im voraus wissen wir nicht, ob Methode funktioniert.

Beispiel:

$$F(x) = \begin{pmatrix} 10x_1 + x_2 + \sin(x_1 + x_2) \\ x_1 + 10x_2 + \sin(x_1 - x_2) \end{pmatrix} \stackrel{!}{=} 0$$

$$F'(x) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} \end{pmatrix} = \begin{pmatrix} 10 + \cos(x_1 + x_2) & 1 + \cos(x_1 + x_2) \\ 1 + \cos(x_1 - x_2) & 10 - \cos(x_1 - x_2) \end{pmatrix}$$

$$x^{(0)} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

ν	0	1	2	3	∞
$x^{(n,y)}$	$\begin{pmatrix} 1 \\ 1 \end{pmatrix}$	$\begin{pmatrix} -0.1842 \\ 0.0409 \end{pmatrix}$	$\begin{pmatrix} 1.363 * 10^{-5} \\ 4.164 * 10^{-4} \end{pmatrix}$	$\begin{pmatrix} -2.89 * 10^{-12} \\ 9.13 * 10^{-13} \end{pmatrix}$	$\begin{pmatrix} 0 \\ 0 \end{pmatrix}$

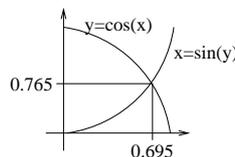
Newton konvergiert also viel schneller als Picard !

Beispiel:

$$f_1(x, y) = \cos x - y = 0$$

$$f_2(x, y) = x - \sin y = 0$$

$$J = \begin{vmatrix} -\sin x & -1 \\ 1 & -\cos y \end{vmatrix} = \sin x \cos y + 1 \neq 0 \Leftrightarrow \sin x \cos y \neq -1 \Rightarrow \text{falls } \begin{matrix} x \geq 0 \\ y \geq 0 \end{matrix} \Rightarrow J \neq 0$$



Newton: im allgemeinen n gleichzeitige Gleichungen mit n Unbekannten

$\Rightarrow n^2$ partielle Ableitungen

\Rightarrow löse für n Inkremente (Zuwächse) h, k

\Rightarrow suche einfachere Methode, die weniger schnell konvergiert, aber leichter zu anzuwenden ist

7 Zusammengesetzte Quadraturformeln

⇒ modifiziere Newton-Verfahren: besteht aus Anwendung von n -mal einwertiges Newton-Verfahren, einmal für jede Variable unter Annahme, daß die anderen Variablen festgehalten werden.

Beispiel:

$$f_1(x, y) = 0$$

$$f_2(x, y) = 0$$

$$(x_0, y_0) = \vec{x}^{(0)}$$

$$\Rightarrow x_1 = x_0 - \frac{f_1(x_0, y_0)}{\frac{\partial f_1}{\partial x}|_{(x_0, y_0)}}, \quad y_1 = y_0 - \frac{f_2(x_1, y_0)}{\frac{\partial f_2}{\partial y}|_{(x_1, y_0)}}$$

$$\Rightarrow x_2 = x_1 - \frac{f_1(x_1, y_1)}{\frac{\partial f_1}{\partial x}|_{(x_1, y_1)}}, \quad y_2 = y_1 - \frac{f_2(x_2, y_1)}{\frac{\partial f_2}{\partial y}|_{(x_2, y_1)}}, \text{ etc...}$$

Im Allgemeinen:

Man kann entweder f_1 oder f_2 benutzen um einen neuen x -Wert zu berechnen und die andere Funktion benutzen um y zu berechnen.

Aber üblicherweise wird eine Wahl konvergieren, die andere (je nach Problem) divergieren.

Beispiel:

$$\left. \begin{array}{l} f_1(x, y) = \cos x - y \\ f_2(x, y) = x - \sin y \end{array} \right\} \begin{array}{l} \frac{\partial f_1}{\partial x} = -\sin x, \quad \frac{\partial f_2}{\partial x} = 1 \\ \frac{\partial f_1}{\partial y} = -1, \quad \frac{\partial f_2}{\partial y} = -\cos y \end{array}$$

Nehme f_1 um nächstes x zu berechnen. Nehme f_2 um nächstes y zu berechnen.

$$x_{i+1} = x_i - \frac{f_1(x_i, y_i)}{\frac{\partial f_1}{\partial x}} = x_i - \frac{\cos x_i - y_i}{-\sin x_i}$$

$$y_{i+1} = y_i - \frac{f_2(x_{i+1}, y_i)}{\frac{\partial f_2}{\partial y}} = y_i - \frac{x_{i+1} - \sin y_i}{-\cos y_i}$$

i	x	y
0	1	1
1	0.4532	0.2823
2	1.86036	1.92929
3	0.45068	5.88268
⋮	⋮	⋮
8	-1336	-19634

Nehme nun f_2 statt f_1 , f_2 berechnet nächstes x und f_1 das nächste y .

$$x_{i+1} = x_i - \frac{f_2(x_i, y_i)}{\frac{\partial f_2}{\partial x}} = x_i - \frac{x_i - \sin y_i}{1}$$

$$y_{i+1} = y_i - \frac{f_1(x_{i+1}, y_i)}{\frac{\partial f_1}{\partial y}} = y_i - \frac{\cos x_{i+1} - y_i}{-1}$$

i	x	y
0	1	1
1	0.84147	0.66637
2	0.618134	0.81496
3	0.727699	0.746706
⋮	⋮	⋮
8	0.694896	0.768169

7.3 Bisektions-Methode

Annahmen: $F \in C[a, b]$, $F(a) \cdot F(b) < 0$

$a_0 := a$, $b_0 := b$ (oBdA: $F(a) < 0$)

$$[a_{k+1}, b_{k+1}] := \begin{cases} [\frac{a_k+b_k}{2}, b_k] & , \text{ falls } F(\frac{a_k+b_k}{2}) < 0 \\ [a_k, \frac{a_k+b_k}{2}] & , \text{ falls } F(\frac{a_k+b_k}{2}) > 0 \end{cases}$$

Dies ist eine numerische sehr stabile Methode. Sie funktioniert selbst dann, wenn $F \in C^0$ (nur Konvergenz) ist allerdings sehr langsam.

Beispiel:

Löse $F(x) := x^3 - x - 1 = 0$, $F(1) = -1 < 0 < 5 = F(2)$

$F \in C^0 \Rightarrow \exists^1 F(\xi) = 0$, $\xi \in [1, 2]$ ($F'(x) = 3x^2 - 1 > 0 \forall x \in [1, 2]$)

$F(1.5) = 0.875 > 0 > -1 = F(1) \Rightarrow \xi \in [1, 1.5]$

$F(1.25) = -0.296 < 0 < 0.875 = F(1.5) \Rightarrow \xi \in [1.25, 1.5]$

Fehler: $|x^{(k)} - x^*|$ in Bisektionsmethode $x^{(k)} = \frac{1}{2}(a_k + b_k)$, $F(x^*) = 0$

$x^* \in [a_k, b_k]$

$$\Rightarrow x^{(k)} - x^* \in [\frac{1}{2}(a_k + b_k) - b_k, \frac{1}{2}(a_k + b_k) - a_k] = [-\frac{1}{2}(b_k - a_k), \frac{1}{2}(b_k - a_k)]$$

$$\Rightarrow |x^{(k)} - x^*| \leq \frac{1}{2}|b_k - a_k|$$

Ferner:

$$|b_k - a_k| = \left\{ \begin{array}{l} |b_{k-1} - \frac{a_{k-1} + b_{k-1}}{2}| \\ \text{(oder)} \\ |\frac{a_{k-1} + b_{k-1}}{2} - a_{k-1}| \end{array} \right\} = \frac{1}{2}|b_{k-1} - a_{k-1}| = \dots = 2^{-k}|b_0 - a_0| = 2^{-k}|b - a|$$

$$\Rightarrow |x^{(k)} - x^*| \leq 2^{-(k+1)}|b - a|$$

Beispiel:

Finde $[a_0, b_0]$ geeignet um $xe^x = 1$ zu lösen.

Wieviele Iterationen (der Bisektion-Methode) sind nötig für 2,4,6,8,10 Dezimalen Genauigkeit ?

$F(x) = xe^x - 1$, $F(0) = -1$, $F(1) = e - 1 > 0 \Rightarrow [a_0, b_0] = [0, 1]$ geeignet.

Genauigkeit bis auf p Digits $\Leftrightarrow |x^{(k)} - x^*| \leq 5 \cdot 10^{-(p+1)}$

$$|x^{(k)} - x^*| \leq 2^{-(k+1)} \underbrace{|b - a|}_1 \stackrel{!}{\leq} 5 \cdot 10^{-(p+1)}$$

$$\Rightarrow k \geq -\frac{\log 5 - p - 1}{\log 2} - 1 = 3.32193 * p$$

p	k
2	7
4	14
6	20
8	27
10	34

8 Matrix-Eigenwertaufgaben

8.1 Vorbemerkungen, Eigenwertabschätzungen

$A \in \mathbb{R}^{n \times n}$ (oder $\in \mathbb{C}^{n \times n}$; in der Praxis selten)

$x (\neq 0) \in \mathbb{R}^n$ (oder \mathbb{C}^n zugelassen, auch für $A \in \mathbb{R}^{n \times n}$) ist Eigenvektor (EV) zum Eigenwert (EW) $\lambda \in \mathbb{C}$ der Matrix A, wenn gilt

$$Ax = \lambda x \tag{8.1}$$

EW-Aufgaben z.B. bei Schwingungsproblemen, DGL-EWA

numerische Lösungsansätze \Rightarrow Matrizen-EWA

(8.1) $\Rightarrow charDet = 0$ lösen.

$$det(a - \lambda E) = \begin{vmatrix} a_{11} - \lambda & a_{12} & \cdots & a_{1n} \\ \vdots & & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} - \lambda \end{vmatrix} = 0 \tag{8.2}$$

Spektralradius $\rho(A) := \max_{i=1}^n |\lambda_i|$, wenn $\lambda_1, \dots, \lambda_n$ die Eigenwerte von A sind (gezählt entsprechend ihrer Vielfachheit)

Definition: Zwei Matrizen $A, B \in \mathbb{R}^{n \times n}$ sind ähnlich, wenn es eine Matrix $T \in \mathbb{R}^{n \times n}$ (bzw. $\in \mathbb{C}^{n \times n}$) gibt mit $A = T^{-1}BT$

Ähnliche Matrizen haben dieselben Eigenwerte:

$$det(A - \lambda E) = det(T^{-1}BT - \lambda E) = det(T^{-1}BT - \lambda T^{-1}T) = det[T^{-1}(B - \lambda E)T] = \underbrace{detT^{-1}}_{\frac{1}{detT}} * det(B - \lambda E) detT = det(B - \lambda E) \quad \square$$

Sei $AX = \lambda x$, d.h. $(T^{-1}BT)x = \lambda x \Rightarrow B \underbrace{Tx}_y = \lambda \underbrace{Tx}_y \Rightarrow y$ ist EV von B zum EW λ

(und umgekehrt).

Ist $\lambda \in \mathbb{C}$ EW von $A \in \mathbb{R}^{n \times n}$, so ist auch $\bar{\lambda}$ EW von A:

$$AX = \lambda x \Rightarrow A\bar{x} = \bar{\lambda}\bar{x} \quad (da A \in \mathbb{R}^{n \times n})$$

$\Rightarrow \bar{\lambda}$ EW zum EV \bar{x}

Ist $A = A^T$, so sind alle EW von A reell:

$$\left. \begin{array}{l} \bar{x}^T Ax = \lambda \bar{x}^T x \\ \left. \begin{array}{l} \bar{x}^{T*} \\ x^{T*} \end{array} \right\} \left. \begin{array}{l} Ax = \lambda x \\ A\bar{x} = \bar{\lambda}\bar{x} \end{array} \right\} \Rightarrow \underbrace{x^T A \bar{x}}_{\substack{(A^T x)^T \bar{x} \\ = A}} = \underbrace{\lambda \bar{x}^T x}_{= \lambda x^T \bar{x}} \Rightarrow 0 = (\lambda - \bar{\lambda}) \underbrace{\bar{x}^T x}_{>0} \Rightarrow \lambda = \bar{\lambda} \Leftrightarrow \lambda \in \mathbb{R} \quad \square \end{array}$$

$A = A^T$ besitzt n linear unabhängige EV, die als ONB gewählt werden können.

$$\exists B = (\vec{b}_1, \dots, \vec{b}_n) \quad B^{-1}AB = D, \text{ so daß } D = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}, \lambda_i \text{ EW zu } \vec{b}_i$$

Ist $A = A^T$ positiv definit, so gilt: alle EW von A sind positiv:

$$Ax = \lambda x, \quad x \in \mathbb{R}^n$$

$$\underbrace{x^T Ax}_{>0, da x \neq 0} = \lambda \underbrace{x^T x}_{>0} \Rightarrow \lambda > 0 \text{ (Entsprechend gilt Umkehrung)}$$

8 Matrix-Eigenwertaufgaben

Eigenwert-Abschätzungen:

(i) $\|\cdot\|_M$ Matrix-Norm $\Rightarrow |\lambda| \leq \|A\|_M$ für alle EW von A .

(ii) Abschätzung mit den Gerschgorin-Kreisen: Sei $A = (a_{ik})_{i,k=1,\dots,n}$, die Mengen
 $K_\mu := \{z \in \mathbb{C} : |z - a_{\mu\mu}| \leq \sum_{\nu=1, \nu \neq \mu}^n |a_{\mu\nu}|\}$ ($\mu = 1, \dots, n$) (8.3)

heißen Gerschgorin-Zeilenkreise.

Entsprechend

$K'_\nu := \{z \in \mathbb{C} : |z - a_{\nu\nu}| \leq \sum_{\mu=1, \mu \neq \nu}^n |a_{\mu\nu}|\}$ (8.4)

Gerschgorin-Spaltenkreise.

Satz 8.1

1. Seien $\lambda_1, \dots, \lambda_n \in \mathbb{C}$ die Eigenwerte von A (gezählt entsprechend der Vielfachheit), so gilt für alle k
 $\lambda_k \in \bigcup_{\mu=1}^n K_\mu, \lambda_k \in \bigcup_{\nu=1}^n K'_\nu \Rightarrow \lambda_k \in (K_\mu \cap K'_\nu)$
2. Ist die Vereinigung von $m (\leq n)$ Gerschgorin-Zeilenkreisen bzw. Spaltenkreisen disjunkt zu den übrigen (d.h. Durchschnitt leer), so enthält sie genau m EW (entsprechend ihrer Vielfachheit).

Bemerkung: Numerisch brauchbare Werte i.a. nur bei diagonaldominanten Matrizen zu erwarten.

Beispiel: $A = \begin{pmatrix} 1 & 2 & -1 \\ -2 & 3 & 1 \\ -3 & 8 & 1 \end{pmatrix}$

EW: $\lambda_{1,2} = 0, \lambda_3 = 5$

Gerschgorin-Zeilenkreise:

$$|z - 1| \leq 3 = |2| + |-1|$$

$$|z - 3| \leq 3 = |-2| + |1|$$

$$|z - 1| \leq 11 = |-3| + |8|$$

eventuell Verbesserung durch Ähnlichkeitstransformation mit Diagonalmatrizen; für obiges Beispiel:

$$T = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}, T^{-1}AT = \begin{pmatrix} 1 & 2 & -2 \\ -2 & 3 & 2 \\ -\frac{3}{2} & 4 & 1 \end{pmatrix} \text{ hat gleiche EW.}$$

Gerschgorin-Zeilenkreise:

$$|z - 1| \leq 4$$

$$|z - 3| \leq 4$$

$$|z - 1| \leq 5, 5$$

8.2 Die Verfahren von Wilkinson und Householder

Ziel: A durch Ähnlichkeitstransformation auf eine einfachere Form zu bringen, i.a. auf sogenannte Hessenberg-Form

8 Matrix-Eigenwertaufgaben

Definition: (Hessenberg-Matrix)

Eine Matrix $H = (h_{i,k})_{i,k=1,\dots,n}$ heißt Hessenberg-Matrix, wenn gilt $h_{i,k} = 0$ für $k + 1 < i$ ($i, k = 1, \dots, n$)

$$H = \begin{pmatrix} h_{1,1} & h_{1,2} & \cdots & \cdots & \cdots & h_{1,n} \\ h_{2,1} & h_{2,2} & & & & h_{2,n} \\ 0 & h_{3,2} & & & & h_{3,n} \\ \vdots & 0 & & & & \vdots \\ \vdots & \vdots & \ddots & & & \vdots \\ 0 & 0 & \cdots & 0 & h_{n,n-1} & h_{n,n} \end{pmatrix}$$

Wilkinson

Schrittweise Überführung auf Hessenberg-Form mittels Transformationsmatrizen $E_{p,q}$ und C_p (vergl. LR-Zerlegung).

$$(A =) A^{(0)} \rightarrow A^{(1)} \rightarrow A^{(2)} \rightarrow \cdots \rightarrow A^{(n-2)} = H$$

→ ist Ähnlichkeitstransformation

Die Matrizen $A^{(j-1)}$ sind von der Form:

$$A^{(j-1)} = \begin{pmatrix} * & \cdots & \cdots & \cdots & \cdots & \cdots & * \\ * & \ddots & & & & & \vdots \\ 0 & \ddots & \ddots & & & & \vdots \\ \vdots & \ddots & \ddots & \ddots & & & \vdots \\ 0 & \cdots & 0 & * & \ddots & & \vdots \\ \vdots & & \vdots & \vdots & \ddots & * & \vdots \\ 0 & \cdots & 0 & * & \cdots & * & * \end{pmatrix} = (a_{ik}^{(j-1)})_{i,k=1,\dots,n}$$

(wobei die UnterNullDeterminante in der $(j + 1)$ ten Zeile, bzw. $(j - 1)$ -ten Spalte beginnt).

8.2.1 Wilkinson-Algorithmus

$A^{(j-1)}$ sei bestimmt.

1. q ($j + 1 \leq q \leq n$) so bestimmen, daß gilt:

$$|a_{q,j}^{(j-1)}| = \max_{j+1 \leq \mu \leq n} |a_{\mu,j}^{(j-1)}|$$

(vgl. partielle Pivotwahl bei Gauß, d.h. es ist in der j -ten Spalte unterhalb der Hauptdiagonalen das Betragsgrößte Element zu suchen (bei mehreren, nehme z.B. das mit dem kleinsten Zeilenindex))

2. Für $q \neq j + 1$ vertausche die $(j + 1)$ -te Zeile und Spalte mit der q -ten Zeile und Spalte.

$$A^{(j-1)} \rightarrow E_{j+1,q} A^{(j-1)} E_{j+1,q} =: B^{(j-1)} = (b_{i,k}^{(j-1)})_{j,k}$$

Wegen $E_{j+1,q}^{-1} = E_{j+1,q}$ sind $A^{(j-1)}$ und $B^{(j-1)}$ ähnlich.

8 Matrix-Eigenwertaufgaben

$$A^{(1)} = C_2^{-1}B^{(0)}C_2 = \begin{pmatrix} 1 & 0 & -2 & -2 \\ 2 & 1.5 & 3 & -2 \\ 0 & -0.25 & 3.5 & 0 \\ 0 & 0.5 & -5 & -1 \end{pmatrix}$$

Pivot ist die 0.5 in der 2. Spalte, 4. Zeile.

Vertausche also die 3. und 4. Spalte bzw. Zeile.

$$\Rightarrow B^{(1)} = \begin{pmatrix} 1 & 0 & -2 & -2 \\ 2 & 1.5 & -2 & 3 \\ 0 & 0.5 & -1 & -5 \\ 0 & -0.25 & 0 & 3.5 \end{pmatrix}$$

$$C_3 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -0.5 & 1 \end{pmatrix}$$

$$A^{(2)} = H = C_3^{-1}B^{(1)}C_3 = \begin{pmatrix} 1 & 0 & -1 & -2 \\ 2 & 1.5 & -3.5 & 3 \\ 0 & 0.5 & 1.5 & -5 \\ 0 & 0 & -1 & 1 \end{pmatrix}$$

8.2.2 Verfahren von Householder

entsprechend wie bei Wilkinson:

$$A \rightarrow A^{(1)} \rightarrow \dots \rightarrow H$$

(Hessenbergmatrix)

Transformationsmatrizen $P = E - \frac{1}{C}WW^T$, $W \in \mathbb{R}^n$, $C \in \mathbb{R}$

Dabei gilt wegen des dyadischen Produktes $WW^T \in \mathbb{R}^{n \times n}$

$$(W * W^T)^T = W^{TT}W^T = WW^T \Rightarrow P = P^T$$

Es sei $C = \frac{1}{2}W^TW$, dann ist

$$P^2 = (E - \frac{1}{C}WW^T)(E - \frac{1}{C}WW^T) = E - \frac{2}{C}WW^T + \frac{1}{C^2}W \underbrace{W^TW}_{2C} W^T = E - \frac{2}{C}WW^T +$$

$$\frac{2}{C}WW^T = E$$

$\Rightarrow P^{-1} = P$ (P symmetrisch und orthogonal $\Rightarrow P$ involutorisch)

Householder für 1. Schritt:

$$W = \begin{pmatrix} 0 \\ W_2 \\ \dots \\ W_n \end{pmatrix}$$

8 Matrix-Eigenwertaufgaben

$$A^{(1)} = PAP \stackrel{!}{=} \begin{pmatrix} * & \cdots & \cdots & \cdots & \cdots & \cdots & * \\ * & \ddots & & & & & \vdots \\ 0 & \ddots & \ddots & & & & \vdots \\ \vdots & \ddots & \ddots & \ddots & & & \vdots \\ 0 & \cdots & 0 & * & \ddots & & \vdots \\ \vdots & & \vdots & \vdots & \ddots & * & \vdots \\ 0 & \cdots & 0 & * & \cdots & * & * \end{pmatrix} = (a_{ik}^{(-1)})_{i,k}$$

Es gilt $a_{11}^{(1)} = a_{11}$

Also: zu bestimmen $a_{21}, W_2, \dots, W_n \rightarrow$ führt auf NLGS, speziell gilt:

$$\frac{W_2^2}{2\|W\|_2^2} = 1 - \frac{a_{21}}{\pm\sqrt{\sum_{i=2}^n a_{i1}^2}}$$

1. $\sum_{j=2}^n a_{j1}^2 = 0$, dann $A = \begin{pmatrix} * \\ 0 \\ \vdots \\ 0 \end{pmatrix} \begin{matrix} \\ * \\ \\ \end{matrix}$, d.h. A hat schon gewünschte Form

\Rightarrow 1. Schritt überspringen.

2. $\sum_{j=2}^n a_{j1}^2 > 0$. Das Vorzeichen ($\pm\sqrt{\sum \dots}$) ist eigentlich beliebig, aber wegen Rechenstabilität so zu wählen, daß $|W_2|$ möglichst groß wird.
(Householder ist dann i.a. sehr rechenstabil)

Verfahren fortsetzen $\rightsquigarrow A^{(1)}, A^{(2)}, \dots$

\Rightarrow Algorithmus:

$$A^{(0)} = A$$

für $k = 1, \dots, n-2$:

$$s = \sqrt{\sum_{j=k+1}^n a_{jk}^{(k-1)2}}$$

ist $s = 0$, dann nächster Schritt ($k \rightsquigarrow k+1$).

$$t := a_{k+1,k}^{(k-1)}, \quad C := s(s + |t|)$$

$$W_j^{(k)} = 0 \quad (j = 1, \dots, k)$$

$$W_{k+1}^{(k)} = t + s * \operatorname{sgn}(t)$$

$$W_j^{(k)} = a_{jk}^{(k-1)} \quad (j = k+2, \dots, n)$$

$$P^{(k)} := E - \frac{1}{C} W^{(k)} W^{(k)T}$$

$$A^{(k)} = P^{(k)} A^{(k-1)} P^{(k)}$$

Vergleich der Rechenoperationen (Division/Plutimikation)

Wilkinson: $\frac{5}{6}n^3 + O(n^2)$

Householder: $\frac{5}{3}n^3 + O(n^2)$ ($+(n-2)$ *Wurzelziehen)

Also:

Für allgemeine Matrizen ist Wilkinson günstiger, Householder allerdings Rechenstabiler,

8 Matrix-Eigenwertaufgaben

aber: $A = A^T \Rightarrow A^{(1)T} = (PAP)^T = \underbrace{P^T}_{=P} \underbrace{A^T}_{=A} P = PAP = A^{(1)}$,

d.h. auch alle "Zwischenmatrizen" $A^{(k)}$ sowie $H = A^{(n-2)}$ sind symmetrisch

$$\Rightarrow H = \begin{pmatrix} * & * & & 0 \\ * & \ddots & \ddots & \\ & \ddots & \ddots & * \\ 0 & & * & * \end{pmatrix} \text{ Tridiagonalmatrix}$$

Für $A = A^T$ müssen die $P^{(k)}$ nicht explizit berechnet werden:

$C, W^{(k)}$ wie bisher,

$$u^{(k)} = \frac{1}{C} A^{(k-1)} W^{(k)}$$

$$v^{(k)} = u^{(k)} - \frac{1}{2C} W^{(k)} * W^{(k)T} * u^{(k)}$$

$$\Rightarrow A^{(k)} = A^{(k-1)} = v^{(k)} W^{(k)T} - W^{(k)} v^{(k)T}$$

In diesem Fall werden nur $\frac{2}{3}n^3 + O(n^2)$ Divisionen/Plutimikationen ($+(n-2)*$ Wurzelziehen) benötigt, d.h. schneller als Wilkinson.

8.3 Berechnung von Eigenwerten und Eigenvektoren von Hessenberg-Matrizen

$$H = (h_{ik}) \in \mathbb{R}^{n \times n} \quad (H - \lambda E)x = 0$$

Charakteristisches LGS:

$$(h_{11} - \lambda)x_1 + h_{12}x_2 + \dots + h_{1n}x_n = 0$$

$$h_{21}x_1 + (h_{22} - \lambda)x_2 + \dots + h_{2n}x_n = 0$$

$$\vdots \quad \vdots$$

$$h_{n,n-1}x_{n,n-1} + (h_{nn} - \lambda)x_n = 0$$

(8.7)

Charakteristische Gleichung:

$$\begin{vmatrix} h_{11} - \lambda & \dots & \dots & h_{1n} \\ h_{21} & \ddots & & \\ & & \ddots & \\ 0 & & h_{n,n-1} & h_{nn} - \lambda \end{vmatrix} \stackrel{!}{=} 0$$

- Es gibt ein k ($1 \leq k \leq n-1$) mit $h_{k+1,k} = 0 \Rightarrow$ LGS (8.7) zerfällt in 2 getrennt LGS

$$\Rightarrow \det(H - \lambda E_n) = \det(H - \lambda E_\rho) * \det(J - \lambda E_{n-\rho})$$

d.h. weiter mit 2) bei geringerer Reihenzahl.

- für alle k ($1 \leq k \leq n-1$) $h_{k+1,k} \neq 0$ (im folgenden stets erfüllt)

- Entwicklung der $k \times k$ -Unterdeterminanten in der linken oberen Ecke der charakteristischen Determinante nach jeweils letzter Spalte:

$$(f_0(\lambda) := 1) \quad f_1(\lambda) = (h_{11} - \lambda)f_0(\lambda)$$

$$f_2(\lambda) = \begin{vmatrix} h_{11} - \lambda & h_{12} \\ h_{21} & h_{22} - \lambda \end{vmatrix} = (h_{22} - \lambda)f_1(\lambda) - h_{12}h_{21}$$

8 Matrix-Eigenwertaufgaben

$$\begin{aligned}
 f_3(\lambda) &= \begin{vmatrix} h_{11} - \lambda & h_{12} & h_{13} \\ h_{21} & h_{22} - \lambda & h_{23} \\ 0 & h_{32} & h_{33} - \lambda \end{vmatrix} \\
 &= (h_{33} - \lambda)f_2(\lambda) - h_{23}h_{32}f_1(\lambda) + h_{13}h_{32}h_{21} \overbrace{f_0(\lambda)}^1 \\
 &\text{und so fort (vollständige Induktion).} \\
 f_0(\lambda) &= 1 \\
 f_j(\lambda) &= (\alpha_{jj} - \lambda)f_{j-1}(\lambda) + \alpha_{j-1,j}f_{j-2}(\lambda) + \dots + \alpha_{2j}f_1(\lambda) + \alpha_{1j}f_0(\lambda) \\
 &\quad (j = 1, \dots, n) \\
 \alpha_{jj} &= h_{jj} \\
 \alpha_{kj} &= (-1)^{j+k} h_{kj} * h_{j,j-1} * h_{j-1,j-2} * \dots * h_{k+1,k} \quad (k = 1, \dots, j-1) \\
 f_n(\lambda) &= \det(H - \lambda E)
 \end{aligned} \tag{8.8}$$

- b) Bestimmung von Eigenvektoren von H (zum berechneten Eigenwert λ)
 Das charakteristische LGS (8.7) hat eine nicht-triviale Lösung.
- i. $x_n = 0$ setzen in letzte Gleichung. $\Rightarrow x_{n-1} = 0$
 in vorletzte Gleichung $\Rightarrow x_{n-2} = 0$
 \vdots
 in 2. Gleichung $\Rightarrow x_1 = 0$
 Insgesamt also: kein Eigenvektor mit $x_n=0$.
 Daher:
 - ii. $x_n \neq 0$ (o.B.d.A.) $x_n = 1$ setzen.
 $\Rightarrow x_{n-1}, \dots, x_1$ sind eindeutig bestimmt.
 (d.h. jede nicht zerfallende Hessenbergmatrix hat genauso viel linear unabhängige Eigenvektoren, wie sie verschiedene Eigenwerte hat)

Beispiel:

$$H = \begin{pmatrix} 1 & 0 & -1 & -2 \\ 2 & 1.5 & -3.5 & 3 \\ 0 & 0.5 & 1.5 & -5 \\ 0 & 0 & -1 & 1 \end{pmatrix}$$

$$f_0(\lambda) = 1 \text{ (gesetzt !)}$$

$$f_1(\lambda) = 1 - \lambda$$

$$f_2(\lambda) = (1.5 - \lambda)f_1(\lambda) - 0 * 2 = \lambda^2 - 2.5\lambda + 1.5$$

$$f_3(\lambda) = (1.5 - \lambda)f_2(\lambda) - (-3.5) * 0.5f_1(\lambda) + (-1) * 0.5 * 2 * 1 = -\lambda^3 + 4\lambda^2 - 7\lambda + 3$$

$$\begin{aligned}
 f_4(\lambda) &= (1 - \lambda)f_3(\lambda) - (-5)(-1)f_2(\lambda) + 3 * (-1) * 0.5f_1(\lambda) - (-2)(-1) * 0.5 * 2f_0(\lambda) \\
 &= \lambda^4 - 5\lambda^3 + 6\lambda^2 + 4\lambda - 8
 \end{aligned}$$

$$f_4(\lambda) \stackrel{!}{=} 0$$

$$\text{Lösungen: } \lambda_{1,2,3} = 2, \lambda_4 = -1$$

Beispiel:

obige Matrix H , (8.7) für $\lambda = 2$, $x_4 = 1$ gewählt,

1. Gleichung wegen linearer Unabhängigkeit weglassen.

8 Matrix-Eigenwertaufgaben

$$\left. \begin{array}{l} 2x_1 - 0.5x_2 - 3.5x_3 = -3 \\ 0.5x_2 - 0.5x_3 = 5 \\ -x_3 = 1 \end{array} \right\} \Rightarrow x = \begin{pmatrix} -1 \\ 9 \\ -1 \\ 1 \end{pmatrix}$$

Dies ist der einzige Eigenvektor von H zu $\lambda = 2$

$$(H = \overbrace{\lambda}^{=2} E)x = 0$$

entsprechend: $\lambda = -1, x_4 = 1 \Rightarrow \tilde{x} = \begin{pmatrix} 2 \\ 0 \\ 2 \\ 1 \end{pmatrix}$

$$A = \begin{pmatrix} 1 & -2 & 3 & -2 \\ 1 & 5 & -1 & -1 \\ 2 & 3 & 2 & -2 \\ 2 & -2 & 6 & -3 \end{pmatrix}$$

Wilkinson-Algorithmus:

$$H = \begin{pmatrix} 1 & 0 & -1 & -2 \\ 2 & 1.5 & -3.5 & 3 \\ 0 & 0.5 & 1.5 & -5 \\ 0 & 0 & -1 & 1 \end{pmatrix}$$

Transformationsmatrix:

$$T = E_{23} * C_2 * E_{34} * C_3 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0.5 & -0.5 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{pmatrix}$$

$$C_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0.5 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}$$

$$C_3 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -0.5 & 1 \end{pmatrix}$$

$$\Rightarrow T * \begin{pmatrix} -1 \\ 9 \\ -1 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 6 \\ 0 \\ 8 \end{pmatrix} \text{ ist einziger Eigenvektor von } A \text{ zum Eigenwert } \lambda =$$

$$\text{Entsprechend: } T * \begin{pmatrix} 2 \\ 0 \\ 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \\ 0 \\ 2 \end{pmatrix} \text{ Eigenvektor von } A \text{ zum Eigenwert } \lambda = -1$$

8.4 von Mises'sche Iterationsverfahren (1929)

Ziel: Bestimmung einzelner Eigenwerte, z.B. dem betragsgrößten
 A symmetrisch $\Leftrightarrow A = A^T$

\Rightarrow

1. Es existiert eine Orthonormalbasis aus Eigenvektoren $x^{(1)}, \dots, x^{(n)}$
2. Die Eigenwerte sind reell

$|\lambda_1| \geq |\lambda_2| \geq \dots \geq |\lambda_n|$ Ordnung der Eigenwerte
 entsprechend der Vielfachheit auffüllen.

$Z^{(0)} \in \mathbb{R}^n$ Startvektor

$$Z^{(0)} = c_1 x^{(1)} + c_2 x^{(2)} + \dots + c_n x^{(n)}$$

$$Z^{(1)} = AZ^{(0)} = c_1 \lambda_1 x^{(1)} + c_2 \lambda_2 x^{(2)} + \dots + c_n \lambda_n x^{(n)}$$

$$Z^{(2)} = AZ^{(1)} = c_1 \lambda_1^2 x^{(1)} + c_2 \lambda_2^2 x^{(2)} + \dots + c_n \lambda_n^2 x^{(n)}$$

$$z^{(\nu)} = AZ^{(\nu-1)} = \underbrace{c_1 \lambda_1^\nu x^{(1)}}_{\text{dominiert für } \nu \rightarrow \infty} + c_2 \lambda_2^\nu x^{(2)} + \dots + c_n \lambda_n^\nu x^{(n)} \tag{8.11}$$

♡

$$z^{(\nu)} = Az^{(\nu-1)} \quad (\nu = 1, 2, \dots)$$

für $\nu \rightarrow \infty$:

$$z^{(\nu)} \rightsquigarrow c_1 \lambda_1^\nu x^{(1)}$$

$$z^{(\nu+1)} \rightsquigarrow c_1 \lambda_1^{(\nu+1)} x^{(1)}$$

Definition:

$$q_i^{(\nu+1)} := \frac{z_i^{(\nu+1)}}{z_i^{(\nu)}} \text{ } i\text{-te Komponente, } z_i^{(\nu)} \neq 0, \text{ } i = 1, \dots, n$$

$$q_i^{(\nu+1)} \rightarrow \lambda_1 \text{ für } \nu \rightarrow \infty$$

Sinnvoll für Rechnung: Normierung der $z^{(\nu)}$:

$$\hat{z}^{(\nu)} = Az^{(\nu-1)}$$

$$z^{(\nu)} = \frac{\hat{z}^{(\nu)}}{\|\hat{z}^{(\nu)}\|_\infty}$$

(Konvergenz statt asymptotisch wegen Normierung)

Beispiel:

$A(=A^T)$	$Z^{(0)}$	$\hat{Z}^{(1)}$	$Z^{(1)}$	$\hat{Z}^{(2)}$	$Z^{(2)}$	$\hat{Z}^{(3)}$
5 -2 -4	1	5	1	9	1	
-2 2 2	0	-2	-0.4	-4.4	-0.489	
-4 2 5	0	-4	-0.8	-8.8	-0.978	

$$z^{(\nu)} \rightarrow \begin{pmatrix} 1 \\ -0.5 \\ -1 \end{pmatrix} \quad (\nu \rightarrow \infty)$$

$$q_1^{(4)} = 9.9888$$

$$q_2^{(4)} = 10.0086$$

8 Matrix-Eigenwertaufgaben

$$q_3^{(4)} = 10.0085$$

$$q_j^{(\nu+1)} \rightarrow 10 = \lambda_1 \quad (\nu \rightarrow \infty) \quad (\lambda_2 = \lambda_3 = 1 \ll 10, \text{ daher hier gute Konvergenz})$$

Aus (♥) folgt

$$|\lambda_1 - q_j^{(\nu+1)}| = \left| \frac{\lambda_2}{\lambda_1} \right|^\nu * O(1) \tag{8.12}$$

(d.h. Konvergenz $(\nu \rightarrow \infty)$ für $|\lambda_2| \approx |\lambda_1|$ mäßig gut)

$$|\lambda_1 - a_j^{(\nu+1)}| = \left| \lambda_1 - \frac{z_j^{(\nu+1)}}{z_j^{(\nu)}} \right| = \left| \lambda_1 - \frac{c_1 \lambda_1^{\nu+1} x_j^{(1)} + c_2 \lambda_2^{\nu+1} x_j + \dots + c_n \lambda_n^{\nu+1} x_j^{(n)}}{c_1 \lambda_1^\nu x_j^{(1)} + c_2 \lambda_2^\nu x_j^{(2)} + \dots + c_n \lambda_n^\nu x_j^{(n)}} \right| \leq c \left| \frac{\lambda_2}{\lambda_1} \right|^\nu$$

da $|\lambda_1| > |\lambda_2| \geq \dots$

Bemerkung: von Mises auch für $A \neq A^T$ möglich, entscheidend ist nur, daß A n linear unabhängige Eigenvektoren besitzt.

Verbesserung für $A = A^T$ mit Rayleighquotient:

$X \in \mathbb{R}^n$ beliebig, $x \neq 0$.

Die Reelle Größe

$$R[x] = \frac{x^T A x}{x^T x} \text{ heißt Rayleighquotient.} \tag{8.13}$$

Satz 8.2

Für $A = A^T$ nimmt der Rayleighquotient sein Maximum bzw. Minimum für einen zum größten bzw. kleinsten Eigenwert gehörenden Eigenvektor an.

Beweis:

$0 \neq x \in \mathbb{R}^n$, $x = c_1 x^{(1)} + \dots + c_n x^{(n)}$ Entwicklung nach Eigenvektoren von $x^{(j)}$ von A .

$$x^T A x = \lambda_1 c_1^2 + \dots + \lambda_n c_n^2 \text{ wegen } x^{(j)T} x^{(k)} = \begin{cases} 1 & , j = k \\ 0 & , \text{sonst} \end{cases}$$

$$x^T x = c_1^2 + \dots + c_n^2$$

$$\lambda_{\min} \leq R[x] := \frac{\lambda_1 c_1^2 + \dots + \lambda_n c_n^2}{c_1^2 + \dots + c_n^2} \leq \lambda_{\max} \quad \square$$

Berechnung des $R[z^{(\nu)}]$ für von Mises ohne Mehraufwand

$$R[z^{(\nu)}] := \frac{z^{(\nu)T} A z^{(\nu)}}{z^{(\nu)T} z^{(\nu)}} = \frac{z^{(\nu)T} \hat{z}^{(\nu+1)}}{z^{(\nu)T} z^{(\nu)}}$$

$$\text{aus } \heartsuit \text{ folgt } |\lambda_1 - R[z^{(\nu)}]| = \left(\frac{\lambda_2}{\lambda_1} \right)^{2\nu} * O(1) \tag{8.14}$$

also Verbesserung gegenüber (8.12).

Siehe obiges Beispiel $R[z^{(\nu)}] = \frac{z^{(3)T} \hat{z}^{(4)}}{z^{(3)T} z^{(3)}} = 9.9997$ ($< \lambda_1 = 10$, da $\lambda_{\max} = \lambda_1$)

Weitere Bemerkungen zu von Mises:

$$1. \text{ Sei } \lambda_1 = \dots = \lambda_p, |\lambda_p| > |\lambda_{p+1}| \geq \dots$$

$$z^{(0)} = \underbrace{c_1 x^{(1)} + \dots + c_p x^{(p)}}_{=y \text{ EV zum EW } \lambda_1 (= \dots = \lambda_p)} + \sum_{j=p+1}^n c_j x^{(j)}$$

$$z^{(\nu)} = \lambda_1^\nu y + \sum_{j=p+1}^n c_j \lambda_j^\nu x^{(j)}$$

$$q_i^{(\nu)} \rightarrow \lambda_1 \text{ asymptotisch} \rightarrow \lambda_1^\nu y$$

d.h. praktisch keine Veränderung des Verfahrens.

8 Matrix-Eigenwertaufgaben

z.B. (Jacobi): das Beträgsgrößte Element in A außerhalb der Hauptdiagonalen suchen:

$$a_{jk} = a_{kj} \quad (j \neq k)$$

“Drehwinkel ϕ ” in Ω_{jk} so gestrichen, daß $a_{jk}^{(1)} = 0 = a_{kj}^{(1)}$ gilt, Verfahren fortsetzen. Eine erzeugte “0” kann dabei wieder durch eine Zahl $\neq 0$ ersetzt werden, aber die Konvergenz läßt sich nachweisen.

(**Formel** für $A = (a_{jk}) \rightarrow A^{(1)} = (a'_{jk})$:

(numerisch stabil; bei (j, k) werden 0 erzeugt.)

$$\theta := \frac{a_{jj} - a_{kk}}{2a_{jk}} \quad (= \cot 2\phi) \tag{8.15}$$

$$t = \frac{s(\theta)}{|\theta| + \sqrt{1 + \theta^2}} \quad (= \tan \phi) \quad \text{mit } s(\theta) = \begin{cases} 1 & , \theta \geq 0 \\ -1 & , \theta < 0 \end{cases}$$

(für sehr große θ , d.h. Diagonalmatrix schon näherungsweise erreicht)

$$c = \frac{1}{\sqrt{1 + t^2}} \quad (= \cos \phi)$$

$$s = t * c \quad (= \sin \phi)$$

$$\tau = \frac{s}{1 + c} \quad (= \tan \frac{\phi}{2})$$

$$\left. \begin{aligned} a'_{rj} &= a'_{jr} = a_{rj} + s(a_{rk} - \tau a_{rj}) \\ a'_{rk} &= a'_{kr} = a_{rk} - s(a_{rj} + \tau a_{rk}) \end{aligned} \right\} r \neq j, k$$

$$a'_{jj} = a_{jj} + ta_{jk}$$

$$a'_{jk} = a'_{kj} = 0$$

$$a'_{kk} = a_{kk} - ta_{jk}$$

Bemerkung:

Wegen A ähnlich zu D liefert das Produkt der Transformationsmatrizen Ω_{jk} eine Orthonormalbasis von Eigenvektoren von A .

Allgemein gilt: Jacobi ziemlich rechenaufwendig, aber sehr stabil.

8.6 QR-Verfahren (Francis, 1961)

$A \in \mathbb{R}^{n \times n}$, zunächst beliebig.

$$\det A \neq 0 \Rightarrow P * A = L * R$$

LR-Zerlegung von A

Hier: Matrix Q mit $Q = Q^T$ (orthogonale Matrix) bestimmen mit $A = Q * R$, $R =$

$$\begin{pmatrix} \heartsuit & \heartsuit & \heartsuit \\ & \ddots & \heartsuit \\ & & \heartsuit \end{pmatrix}$$

QR-Zerlegung von A läßt sich für jede beliebige Matrix durchführen (z.B. Q als Produkt von Householder-Matrizen).

Iterationsverfahren:

$$A_i = Q_i R_i \quad (A_1 = A, Q_i, R_i \text{ wie oben})$$

$$A_{i+1} = R_i Q_i = \underbrace{Q_i^T A_i}_{R_i} \underbrace{Q_i}_{=Q_i^{-1}} \quad (Q_i^T Q_i = E)$$

$$\text{d.h. } A_{i+1} = Q_i^T \cdots Q_1^T A \underbrace{Q_1 \cdots Q_i}_{:=U_i} \Rightarrow \boxed{A_{i+1} = U_i^T A U_i} \quad (U_i^T * U_i = E)$$

8 Matrix-Eigenwertaufgaben

(alle Iterationsmatrizen sind "orthogonal"-ähnlich)

Durchführung der QR-Zerlegung mit Jacobi-Gievens Drehmatrizen (8.14). Der Aufwand für allgemeine Matrizen ist sehr groß, daher meistens nur für Hessenbergmatrizen

$\begin{pmatrix} \ddots & \heartsuit & \heartsuit \\ \ddots & \ddots & \heartsuit \\ & \ddots & \ddots \end{pmatrix}$ bzw. im Fall $A = A^T$ bei Tridiagonalmatrizen $\begin{pmatrix} \ddots & & & \\ & \ddots & & \\ & & \ddots & \\ & & & \ddots \end{pmatrix}$
sinnvoll.

(Frau sollte Wilkinson bzw. Householder zuerst anwenden, dann das QR-Verfahren, da diese Methoden für große Matrizen numerisch oft instabil sind.)

Semester ENDE !

Index

- Ökonomiebedingung, 5
- ähnlich, 73
- a-posteriore, 67
- a-priori, 67
- Abschätzung, 19
 - kubischer Spline, 20
- Abschneiden, 8
- Algorithmus, 5
 - Implementierung, 5
- Approximation, 48
 - beste, 48
 - SOR-, 38
 - Tschebyscheff, 48
- Banach, 65
- Bedingung
 - Ökonomie, 5
- Bessel, 51
- Choleski
 - Zerlegung, 31
- Cramer, 24
- Defekt, 35, 41
- Definit, 33
- Dezimalen
 - gültige, 7
- diagonaldominant, 40
- Differenz
 - dividierte, 11
 - vorwärtsgerichtet, 13
- Diskretisierungsfehler, 54
- dividierte
 - Differenz, 11
- Dreiecksungleichung, 33
- Einzelrittverfahren, 38
- ESV
 - nichtlinear, 44
- Fehler
 - absoluter, 7
 - Diskretisierungs-, 54
 - Gauss, 63
 - Interpolations-, 14
 - Quadratur, 59
 - quellen, 7
 - relativer, 7
 - wahrer, 7
- Fixpunktsatz, 65
- Frobenius, 28
- Frobenius-Norm, 34
- Funktionen
 - Hut-, 18
- gültige Dezimalen, 7
- Gauß
 - Quadratur, 61
- Gauss
 - Fehler, 63
- Gerschgorin-Kreise, 74
- Gesamtschrittverfahren, 38
- Gieves
 - Matrizen, 84
- GSV
 - nichtlinear, 44
- Hessenberg, 74, 86
- Homogen, 33
- Hutfunktion, 18
- Integrationsintervall
 - Partition, 60
- Interpolation
 - fehler, 14
- Jacobi
 - Matrizen, 84
- Knoten
 - polynom, 14
- Kondition, 9, 36
- konditioniert
 - gut, 9

Index

- schlecht, 9, 37
- konstruktiv
 - Methode, 5
- kontrahierend, 42
- Kontraktion, 65, 67
- Konvergenz
 - Koordinatenweise, 34
- Konvergenzordnung, 68
- Korrektur, 7
- Kronecker
 - Symbol, 10
- kubisch
 - Spline, 20
- Lagrange
 - polynom, 10
- Legendrepolynom, 51
- LR-Zerlegung, 26
- Maschinenzahlen, 8
- Matrizen
 - Tridiagonal, 86
- Methode
 - konstruktive, 5
- Näherung, 7
- Nachiteration, 41
- Newton
 - Polynom, 11
- Norm, 33
 - Frobenius, 34
 - Spaltensummen, 35
 - Spektral-, 35
 - Zeilensummen, 35
- normalisiert, 8
- Partition
 - Intergrationsintervall, 60
- Pivot, 25
- Polynom
 - Knoten-, 14
 - Lagranges, 10
 - Legendre-, 51
 - Newton, 11
 - Tschebyscheff, 15
- Quadratur
 - Gauß, 61
- Quadraturfehler, 59
- Raum
 - unitär, 48
- Rayleigh, 83
- Residuum, 41
- Richardson, 61
- Romberg, 61, 64
- Runden, 8
- Schur, 30
- schwaches
 - Zeilensummenkriterium, 40
- signifikant
 - Ziffern, 7
- Simpson-Regel, 60
- SOR, 38
 - nichtlinear, 44
- Spaltensummennorm, 35
- Spektralnorm, 35
- Spline
 - kubischer, 20
 - Abschätzung, 20
 - Existenz, 20
- Spline-Funktion
 - Definition, 21
- Trapezregel, 60
- Tridiagonalmatrizen, 86
- Tschebyscheff
 - Polynom, 15
- unitär
 - Raum, 48
- Verfahren
 - SOR-, 38
- verträglich, 36
- vorwärtsgerichtet
 - Differenz, 13
- wesentlich
 - Ziffern, 7

Index

Wilkinson, 75
wohlgestellt, 9

Zahlen

 Maschinen-, 8

Zeilensummennorm, 35

zerfallend, 41

Zerlegung

 Cholesky, 31

Ziffern

 signifikante, 7

 wesentliche, 7

zugeordnet, 36