Numerische Mathematik für das Lehramt - Formelsammlung

von Julian Merkert, Sommersemester 2006, Prof. Alefeld

Polynominterpolation

$n+1$ reele Zahlen $x_0, ..., x_n$ und $f_0, ..., f_n \Rightarrow (x_i, f_i)$

- x_i heißen Stützstellen
- f_i heißen Stützwerte

$\Pi_n = \{p | p \text{ ist ein Polynom höchsten n-ten Grades}\}$

Interpolation: Polynom $p \in \Pi_n$ gesucht, für das $p(x_i) = f_i$ gilt

Horner-Schema: $p(x) = (... (((a_n x + a_{n-1}) x + a_{n-2}) x + a_{n-3}) x + ... + a_1) x + a_0$

Lagrange-Darstellung des Interpolationspolynoms

- Lagrange-Grundpolynome: $L_k(x) = \Pi_{s=0, s\neq k}^{n} \frac{x - x_s}{x_k - x_s}$ $k = 0, 1, ..., n$
- Eigenschaften:
  - $L_k(x) \in \Pi_n$
  - $L_k(x_j) = \delta_{kj} = \begin{cases} 1 & j = k \\ 0 & j \neq k \end{cases}$
- $(x_0, f_0)...(x_n, f_n)$ mit $x_0, ..., x_n$ paarweise verschieden $\Rightarrow$ es existiert genau ein Interpolationspolynom $p \in \Pi_n$ mit folgender Darstellung:

$$p(x) = \sum_{k=0}^{n} f_k \cdot L_k(x)$$

Newton'sche Basis- oder Grundpolynome: $1, (x - x_0), (x - x_0)(x - x_1), ..., (x - x_0)(x - x_1) \cdot ... \cdot (x - x_{n-1})$

Steigung:

- 0. Ordnung: $f[x_k] := f_k$
- 1. Ordnung: $f[x_i, x_j] = \frac{f_i - f_j}{x_i - x_j}$
- höherer Ordnung: $f[x_k, x_{k+1}, ..., x_{k+m-1}, x_{k+m}] := \frac{f[x_k, x_{k+1}, ..., x_{k+m-1}] - f[x_{k+1}, ..., x_{k+m}]}{x_k - x_{k+m}}$
  - $f[x_0, x_1, x_2] = \frac{f[x_0, x_1] - f[x_1, x_2]}{x_0 - x_2}$
  - $f[x_1, x_2, x_3] = \frac{f[x_1, x_2] - f[x_2, x_3]}{x_1 - x_3}$
  - $f[x_0, x_1, x_2, x_3] = \frac{f[x_0, x_1, x_2] - f[x_1, x_2, x_3]}{x_0 - x_3}$

Newton'sche Interpolationsformel: $p(x) = f[x_0] + f[x_0, x_1](x - x_0) + ... + f[x_0, ..., x_n](x - x_0) \cdot ... \cdot (x - x_{n-1})$

Restglied / Fehler einer durch p angenäherten Funktion $f : [a, b] \to \mathbb{R}$: $f(\overline{x}) - p(\overline{x}) = \frac{1}{(n+1)!} \cdot \omega(\overline{x}) \cdot f^{(n+1)}(\xi)$

- $\omega(x) = (x - x_0)(x - x_1) \cdot ... \cdot (x - x_n)$
- $\xi = \xi(\overline{x}) \in [a, b]$

Zerlegung: $\Delta = \left\{ a = x_0^{(\Delta)} < x_1^{(\Delta)} < x_2^{(\Delta)} < ... < x_{n(\Delta)}^{(\Delta)} \right\}$

Satz von Faber:
In jeder Folge $\Delta^{(0)}, \Delta^{(1)}, ...$ von Zerlegungen von $[a, b]$ gibt es eine stetige Funktion, für die die Folge der Interpolationspolynome NICHT glm. gegen f konvergiert.

Tschebyscheff-Polynome: $T_n(x) = \cos(n \cdot \arccos(x)) \quad n = 0, 1, 2, ..., \ x \in [-1, 1]$

- $T_{n+1}(x) = 2 \cdot x \cdot T_n(x) - T_{n-1}(x), \ n = 1, 2, ...$
- Der führende Koeffizient von T_n ist 2^{n-1}
- T_n ist gerade für gerades n, T_n ist ungerade für ungerades n
- $T_n, \ n \geq 1$ hat die reellen Nullstellen $x_j = \cos\left(\frac{(2j-1)\cdot\pi}{2\cdot n}\right) \quad j = 1, 2, ..., n$
- $|T_n(x)| \leq 1, \ x \in [-1, 1] \ \forall \ n \geq 0$
- Für $n \geq 1$ gilt: $T_n(x) = \pm 1$ abwechselnd an den $n+1$ Stellen $x_k = \cos\left(\frac{k\cdot\pi}{n}\right) \quad n = 0, 1, ...$
- $\min_{x_0, x_1, ..., x_n \in [-1,1]} \max_{x \in [-1,1]} |(x - x_j) \cdot ... \cdot (x - x_n)| = \frac{1}{2^n}$ mit x_j Nullstelle von T_n (s.o.)

Splineinterpolation

Zerlegung: $\Delta = \{a = x_0 < x_1 < ... < x_{N-1} < x_N = b\}$

Splinefunktion der Ordnung $l \in \mathbb{N}$ zur Zerlegung Δ: Funktion $s \in C^{l-1}[a, b]$, die auf jedem Intervall $[x_{j-1}, x_j], \ j = 1, 2, ..., N$ mit einem Polynom l-ten Grades übereinstimmt.

- Bezeichnung: $s \in S_{\Delta,l} = \left\{s \in C^{l-1}[a, b] : s|_{[x_{j-1}, x_j]} \in \Pi_l\right\}$

Maximale Intervalllänge: $h_{max} = \max_{j=0,1,...,N-1}(x_{j+1} - x_j)$

Unendlich-Norm von u: $u \in C[a, b] \Rightarrow ||u||_\infty := \max_{x \in [a,b]} |u(x)|$

2-Norm von u: $u \in C[a, b] \Rightarrow ||u||_2 := \left(\int_a^b |u(x)|^2 \, \mathrm{dx}\right)^{\frac{1}{2}}$

Lineare Splines: $s \in S_{\Delta,1}$

- Ansatz: $s_j(x) = f_j + \frac{f_{j+1} - f_j}{x_{j+1} - x_j}(x - x_j) \quad x \in [x_j, x_{j+1}], \ j = 0, 1, ..., N_1$
- Zu jeder Zerlegung Δ von $[a, b]$ und Werten $f_0, f_1, ..., f_N$ gibt es genau einen interpolierenden linearen Spline
- Fehlerabschätzung: $||s - f||_\infty \leq \frac{1}{8} \cdot ||f''||_\infty \cdot h_{max}^2$

Kubische Splines: $s \in S_{\Delta,3}$

- $||f'' - s''||_2^2 = ||f''||_2^2 - ||s''||_2^2 - 2 \cdot [(f'(x) - s'(x)) \cdot s''(x)]_{x=a}^{x=b}$
- $||f''||_2^2 - ||s''||_2^2 = ||f'' - s''||_2^2$, falls eine der folgenden Bedingungen erfüllt wird:
  - Natürliche Randbedingung: $s''(a) = s''(b) = 0$
  - Vollständige Randbedingung: $s'(a) = f'(a), \ s'(b) = f'(b)$
  - Periodische Randbedingung: $f'(a) = f'(b), \ s'(a) = s'(b), \ s''(a) = s''(b)$

  Dann: $||s''||_2 \leq ||f''||_2$
- Berechnung: $s_j(x) = \frac{(x_j - x)^3}{6 \cdot h_j} \cdot \sigma_{j-1} + \frac{(x - x_{j-1})^3}{6 \cdot h_j} \cdot \sigma_j + \left(\frac{f(x_j)}{h_j} - \frac{1}{6} \cdot h_j \cdot \sigma_j\right)(x - x_{j-1}) + \left(\frac{f(x_{j-1})}{h_j} - \frac{1}{6} \cdot h_j \cdot \sigma_{j-1}\right)(x_j - x)$
- $h_j = x_j - x_{j-1}$
- Bestimmung der σ_j:
  1. Natürliche Randbedingung: $\sigma_0 = \sigma_N = 0$

$$\begin{pmatrix} 2(h_1 + h_2) & h_2 & & 0 \\ h_2 & 2(h_2 + h_3) & h_3 & \\ & \ddots & \ddots & \ddots \\ & h_{N-2} & 2(h_{N-2} + h_{N-1}) & h_{N-1} \\ 0 & & h_{N-1} & 2(h_{N-1} + h_N) \end{pmatrix} \cdot \begin{pmatrix} \sigma_1 \\ \vdots \\ \vdots \\ \sigma_{N-1} \end{pmatrix} = \begin{pmatrix} g_1 \\ \vdots \\ \vdots \\ g_{N-1} \end{pmatrix}$$

  $g_j = 6 \cdot \left[\frac{f(x_{j+1}) - f(x_j)}{h_{j+1}} - \frac{f(x_j) - f(x_{j-1})}{h_j}\right]$
  $\Rightarrow$ LGS ergibt die σ_i

2. Vollständige Randbedingung:

$$\begin{pmatrix} 2h_1 & h_1 & & 0 \\ h_1 & 2(h_1+h_2) & h_2 & \\ & \ddots & \ddots & \ddots \\ & h_{N-1} & 2(h_{N-1}+h_N) & h_N \\ 0 & & h_N & 2h_N \end{pmatrix} \cdot \begin{pmatrix} \sigma_0 \\ \vdots \\ \vdots \\ \sigma_N \end{pmatrix} = \begin{pmatrix} g_0 \\ \vdots \\ \vdots \\ g_N \end{pmatrix}$$

$g_j = 6 \cdot \left[\frac{f(x_{j+1})-f(x_j)}{h_{j+1}} - \frac{f(x_j)-f(x_{j-1})}{h_j} \right]$ für $j = 1, ..., N-1$

$g_0 = 6 \cdot \left[-f'(x_0) + \frac{f(x_1)-f(x_0)}{h_1} \right]$

$g_N = 6 \cdot \left[f'(x_N) - \frac{f(x_N)-f(x_{N-1})}{h_N} \right]$

3. Periodische Randbedingung: $\sigma_0 = \sigma_N$

$$\begin{pmatrix} 2(h_1+h_N) & h_1 & 0...0 & h_N \\ h_1 & 2(h_1+h_2) & h_2 & \\ & \ddots & \ddots & \ddots \\ & h_{N-2} & 2(h_{N-2}+h_{N-1}) & h_{N-1} \\ h_N & 0...0 & h_{N-1} & 2(h_{N-1}+h_N) \end{pmatrix} \cdot \begin{pmatrix} \sigma_0 \\ \vdots \\ \vdots \\ \sigma_{N-1} \end{pmatrix} = \begin{pmatrix} g_0 \\ \vdots \\ \vdots \\ g_{N-1} \end{pmatrix}$$

$g_j = 6 \cdot \left[\frac{f(x_{j+1})-f(x_j)}{h_{j+1}} - \frac{f(x_j)-f(x_{j-1})}{h_j} \right]$ für $j = 1, ..., N-1$

$g_0 = 6 \cdot \left[\frac{f(x_1)-f(x_0)}{h_1} - \frac{f(x_N)-f(x_{N-1})}{h_N} \right]$

Numerische Integration (Quadratur)

Newton-Cotes (Näherungs-)Formel: $\int_a^b f(x)\,\mathrm{dx} \approx \sum_{k=0}^{n} w_k \cdot f(x_k)$

- Gewichte der Quadraturformel: $w_k = \int_a^b L_k(x)\,\mathrm{dx}$
- Nur für gleichabständige Punkte
- Fehlerabschätzung: $E_n(f) = \int_a^b f(x)\,\mathrm{dx} - \sum_{k=0}^{n} w_k \cdot f(x_k)$
  - Falls f $(n+1)$-mal stetig db $\Rightarrow |E_n(f)| \leq \frac{M_{n+1}}{(n+1)!} \int_a^b |\Pi_{n+1}(x)|\,\mathrm{dx}$
  - $\Pi_{n+1}(x) = (x-x_0)(x-x_1) \cdot ... \cdot (x-x_n)$
  - $M_{n+1} = \max_{\xi \in [a,b]} |f^{(n+1)}(\xi)|$

Trapezregel $(n=1)$: $\int_a^b f(x)\,\mathrm{dx} \approx \frac{b-a}{2} \cdot [f(a)+f(b)] =: T$

- Fehler: $|E_1(f)| \leq \frac{(b-a)^3}{12} \cdot M_2$
- $M_2 = \max_{x \in [a,b]} f''(x)$

Simpson-Regel $(n=2)$: $\int_a^b f(x)\,\mathrm{dx} \approx \frac{b-a}{6} \cdot \left[f(a) + 4 \cdot f\left(\frac{a+b}{2}\right) + f(b) \right] =: S$

- Fehler: $|E_2(f)| \leq \frac{(b-a)^4}{196} \cdot M_3$
- $M_3 = \max_{x \in [a,b]} f'''(x)$

Summierte Trapezregel: $\int_a^b f(x)\mathrm{dx} = \sum_{i=1}^{m} \int_{x_{i-1}}^{x_i} f(x)\mathrm{dx} \approx \frac{h}{2} \cdot [f(x_0) + 2 \cdot f(x_1) + 2 \cdot f(x_2) + ... + 2 \cdot f(x_{m-1}) + f(x_m)]$

- Fehler: $|E_n(f)| \leq \frac{(b-a)^3}{12 \cdot m^2} \cdot M$
- $M = \max_{\xi \in [a,b]} f''(\xi)$
- $x_i = a + i \cdot h$
- $h = \frac{b-a}{m}$

Summierte Simpsonregel:

$$\int_a^b f(x)\,\mathrm{dx} \approx \frac{h}{3}\cdot[f(x_0)+4\cdot f(x_1)+2\cdot f(x_2)+4\cdot f(x_3)+...+2\cdot f(x_{2m-2})+4\cdot f(x_{2m-1})+f(x_{2m})] =: Q(f)$$

- Fehler: $|E_n(f)| \leq \frac{(b-a)^5}{2880\cdot m^4}\cdot M$
- $|f^{(4)}(x)| \leq M$
- $x_i = a + i\cdot h,\ i = 0,...,2\cdot m$
- $h = \frac{b-a}{2\cdot m}$

Bernoulli-Polynome: $q_r(t),\ r = 1,2,...$ mit

1. q_r ist ein Polynom r-ten Grades
2. $q'_{r+1} = q_r$
3. q_r ist ungerade ($q_r(t) = -q_r(-t)$) für ungerades r und gerade ($q_r(t) = q_r(-t)$) für gerades r
4. Ist $r > 1$ und ungerade, so gilt $q_r(-1) = 0$ bzw. $q_r(1) = 0$
5. $q_1(t) = -t$

Euler-MacLaurin'sche Summenformel:
$f : [a,b] \to \mathbb{R}$ 2k-mal stetig db, $[a,b]$ sei in $m \geq 1$ Teilintervalle $[x_{i-1}, x_i],\ i = 1,2,...,m$ und $x_i = a + i\cdot h,\ h = \frac{b-a}{m}$ zerlegt, $I := \int_a^b f(x)\,\mathrm{dx}$, $T(h)$ summierte Trapezregel. Dann:

$$I - T(h) = \sum_{i=1}^{k} c_r h^{2r}\left[f^{(2r-1)}(b) - f^{(2r-1)}(a)\right] - \left(\frac{h}{2}\right)^{2k}\sum_{i=1}^{m}\int_{x_{i-1}}^{x_i} q_{2k}(t) f^{(2k)}(x)\,\mathrm{dx}$$

- $t = -1 + \frac{2}{h}(x - x_{i-1})$
- $c_r = \frac{q_{2r}(1)}{2^{2r}},\ r = 1,2,...,k$

Restglied $\mathcal{O}(h)$: $F(h) = \mathcal{O}(h^n) \Leftrightarrow |\frac{F(h)}{h^n}| \leq c$ für $h \to 0+$

Romberg-Verfahren / Extrapolationsverfahren

- Schema:

m	$T_0(m)$	$T_1(m)$	$T_2(m)$	$T_3(m)$
1	$T_0(1) \to$	$T_1(1) \to$	$T_2(1) \to$	$T_3(1)$
2	$T_0(2) \nearrow \to$	$T_1(2) \nearrow \to$	$T_2(2) \nearrow$	
4	$T_0(4) \nearrow \to$	$T_1(4) \nearrow$		
8	$T_0(8) \nearrow$			

- $T_0(m)$: berechnet mit summierter Trapezregel oder $h_m = \frac{b-a}{2^{m-1}}$,
  $T_0(m) = h_m \cdot \left[\frac{f(a)}{2} + f(a+h_m) + f(a+2h_m) + ... + f(a+2^{m-1}h_m) + \frac{f(b)}{2}\right]$
- $T_k(m) = \frac{4^k \cdot T_{k-1}(2m) - T_{k-1}(m)}{4^k - 1}$

Numerische Integration von Differentialgleichungen

Satz von Picard

- $\left.\begin{array}{l} y' = f(x,y) \\ y(x_0) = y_0 \end{array}\right\}$ Anfangswertproblem
- $D = \{(x,y) \mid x_0 \le x \le x_M,\ y_0 - C \le y_0 \le y_0 + C\}$
- $f(x,y)$ sei stetig in D
- $|f(x,y_0)| \le K$ für $x_0 \le x \le x_M$
- $|f(x,u) - f(x,v)| \le L|u-v|$ für $(x,u),\ (x,v) \in D$
- $C \ge \frac{K}{L}\left(e^{L(x_M - x_0)} - 1\right)$

Dann gibt es genau eine Funktion $y(x) \in C^1[x_0, x_M]$, die dem AWP genügt.

Picard-Iterationsverfahren: $y_n(x) = y_0 + \int_{x_0}^{x} f(s, y_{n-1}(s))\ ds$

Einschrittverfahren für das AWP $\left.\begin{array}{l} y' = f(x,y) \\ y(x_0) = y_0 \end{array}\right\}$

- $x_n = x_0 + n \cdot h,\ n = 0,1,...,N$: Gitterpunkte
- $h = \frac{x_M - x_0}{N}$: Schrittweite
- Näherungswerte y_n: $y_{n+1} = y_n + h \cdot \phi(x_n, y_n; h) \quad n = 0,1,...,N-1$

Euler-Verfahren: $y_{n+1} = y_n + h \cdot f(x_n, y_n)$

Globaler Fehler: $e(x_n) = e_n = y(x_n) - y_n$

Abschneidefehler: $T_n = \frac{y(x_{n+1}) - y(x_n)}{h} - \phi(x_n, y(x_n); h)$

Fehlerabschätzung Einschrittverfahren: $|\phi(x,u;h) - \phi(x,v;h)| \le L_\phi |u-v|,\ |y_n - y_0| \le C$ Dann:

$$|e_n| \le \frac{T}{L_\phi}\left(e^{L_\phi(x_n - x_0)} - 1\right),\ n = 0,1,...,N,\ T = \max_{0 \le n \le N-1} |T_n|$$

Fehlerabschätzung Euler-Verfahren:

- $T_n \le \frac{1}{2} \cdot h \cdot M_2$ mit $M_2 = \max_{x \in [x_0, x_M]} |y''(x)|$
- $|e_n| \le \frac{1}{2} \cdot \frac{M_2}{L} \cdot (e^{L(x_n - x_0)} - 1) \cdot h$

Konstistenz eines Einschrittverfahrens:

- $\forall\ \varepsilon > 0\ \exists\ h(\varepsilon) > 0$ mit $|T_n| < \varepsilon\ \forall\ 0 < h < h(\varepsilon)$ und für jedes Paar $(x_n, y(x_n)),\ (x_{n+1}, y(x_{n+1}))$ auf einer Lösungskure der DGL in D
- $\Leftrightarrow \phi(x, y(x), 0) = f(x, y(x))$

Verfahren p-ter Ordnung: Einschrittverfahren mit p größte positive Zahl, für die $|T_n| \le k \cdot h^p$ $(k > 0)$ gilt.

Konvergenzsatz:

- AWP genüge den Voraussetzungen des Picard'schen Satzes
- ϕ stetig in $D \times [0, h_0]$ $(h \le h_0)$
- Konsistenzbedingung $\phi(x,y;0) = f(x,y)$ erfüllt
- Lipschitzbedingung $|\phi(x,u;h) - \phi(x,v;h)| \le L_\phi |u-v|$ auf $D \times [0, h_0]$

Dann gilt für die mit dem Einschrittverfahren berechneten Näherungen an den Stellen $x_n = x_0 + n \cdot h,\ n = 1,2,...,N$ für $x_n \to x \in [x_0, x_M]$ wenn $h \to 0$ und $n \to \infty$: $\lim_{n \to \infty} y_n = y(x)$

Implizites Verfahren: $y_{n+1} = y_n + \frac{h}{2}\left[f(x_n, y_n) + f(x_{n+1}, y_{n+1})\right]$

- $\left| \int_{x_n}^{x_{n+1}} g(x)\,\mathrm{dx} - \frac{h}{2}\left[g(x_{n+1} + g(x_n)\right] \right| \leq \frac{h^3}{12} \cdot M_2,\ |g''| \leq M_2$
- $|T_n| \leq \frac{1}{12} \cdot h^2 \cdot M_3,\ |y'''(x)| \leq M_3$

Runge-Kutta-Verfahren: $y_{n+1} = y_n + h \cdot \{a \cdot K_1 + b \cdot K_2\}$

- $K_1 = f(x_n, y_n)$
- $K_2 = f(x_n + \alpha \cdot h,\ y_n + \beta \cdot h \cdot K_1)$
- Forderung von Konsistenz und Ordnung 2 $\Rightarrow a + b = 1,\ b = \frac{1}{2\alpha},\ \alpha = \beta$

Taylorentwicklung von $y(x_{n+1}) = y(x_n + h)$:

$$y(x_n + h) = y(x_n) + h \cdot y'(x_n) + \frac{h^2}{2} \cdot y''(x_n) + \frac{h^3}{3!} \cdot y'''(x_n) + \frac{h^4}{4!} \cdot y^{(4)}(x_n) + ...$$

Modifiziertes Euler-Verfahren: $y_{n+1} = y_n + h \cdot f\left[x_n + \frac{1}{2}h,\ y_n + \frac{1}{2} \cdot h \cdot f(x_n, y_n)\right]$

Verbessertes Euler-Verfahren: $y_{n+1} = y_n + \frac{h}{2}\left[f(x_n, y_n) + f(x_n + h,\ y_n + h \cdot f(x_n, y_n))\right]$

Klassisches Runge-Kutta-Verfahren: $y_{n+1} = y_n + \frac{1}{6} \cdot h \cdot \{K_1 + 2 \cdot K_2 + 2 \cdot K_3 + K_4\}$

- $K_1 = f(x_n, y_n)$
- $K_2 = f(x_n + \frac{1}{2} \cdot h,\ y_n + \frac{1}{2} \cdot h \cdot K_1)$
- $K_3 = f(x_n + \frac{1}{2} \cdot h,\ y_n + \frac{1}{2} \cdot h \cdot K_2)$
- $K_4 = f(x_n + h,\ y_n + h \cdot K_3)$
- $T_n = \mathcal{O}(h^4)$

Normen von Vektoren und Matrizen

Norm: $||\cdot|| : \mathbb{R}^n \to \mathbb{R}$ mit $v = (v_i) \in \mathbb{R}^n$:

(a) Definitheit: $||v|| \geq 0,\ ||v|| = 0 \Leftrightarrow v = 0$

(b) Homogenität: $||\lambda v|| = |\lambda| \cdot ||v||,\ \lambda \in \mathbb{R}$

(c) Dreiecksungleichung: $||u + v|| \leq ||u|| + ||v||$

Wichtige Normen:

- 1-Norm: $||v||_1 = \sum_{i=1}^{n} |v_i|$
- 2-Norm: $||v||_2 = \left(\sum_{i=1}^{n} |v_i|^2\right)^{\frac{1}{2}}$
- p-Norm: $||v||_p = \left(\sum_{i=1}^{n} |v_i|^p\right)^{\frac{1}{p}}$
- ∞-Norm: $||v||_\infty = \max_{1 \leq i \leq n} |v_i|$
- Im $\mathbb{R}^n$ sind alle Vektornormen äquivalent.

Matrix-Norm: $||A|| := \sup_{x \neq 0} \frac{||Ax||}{||x||}$

- $||A|| = \max_{||x||=1} ||Ax||$
- Verträglichkeitsbedingung: $||Ax|| \leq ||A|| \cdot ||x||$
- Einheitsmatrix: $||I|| = \sup_{x \neq 0} \frac{||Ix||}{||x||} = 1$

- Normaxiome:
  - $||A|| \geq 0,\ ||A|| = 0 \Leftrightarrow A = 0$
  - $||\lambda A|| = |\lambda| \cdot ||A||,\ \lambda \in \mathbb{R}$
  - $||A + B|| \leq ||A|| + ||B||$
  - Multiplikativität: $||AB|| \leq ||A|| \cdot ||B||$, falls $|| \cdot ||$ Vektornorm zugeordnet
- Zeilensummennorm: $||A||_\infty = \max_{1 \leq i \leq n} \left(\sum_{j=1}^{n} |a_{ij}| \right)$
- Spaltensummennorm: $||A||_1 = \max_{1 \leq j \leq n} \left(\sum_{i=1}^{n} |a_{ij}| \right)$

Nichtlineare Gleichungssysteme

Fixpunkt-Iterationsverfahren:

- Iterationsfähige Gestalt von $f(x) = 0 \Leftrightarrow x = g(x)$
- Verfahren: $x^1 := g(x^0),\ x^{k+1} := g(x^k),\ k = 0, 1, ...$
- a priori Fehlerabschätzung: $||\xi - x^k|| \leq \frac{L^k}{1-L} ||x^1 - x^0||$
- a posteriori Fehlerabschätzung: $||\xi - x^{k+1}|| \leq \frac{L}{1-L} ||x^{k+1} - x^k||$

Banach'scher Fixpunktsatz:

- $D \subset \mathbb{R}^n$ abgeschlossen
- $g : D \to \mathbb{R}^n,\ g(D) := \{g(x) | x \in D\} \subset D$
- g sei auf D kontrahierend, d.h. $||g(x) - g(y)|| \leq L \cdot ||x - y||$ mit $0 \leq L < 1,\ x, y \in D$

Dann besitzt g genau einen Fixpunkt $\xi = (\xi_i) \in D$ und $x^{k+1} = g(x^k),\ k = 0, 1, ...$ konv. für jedes $x^0 \in D$ gegen ξ.

Lineare Konvergenz von $\left\{x^k\right\},\ \lim_{k \to \infty} x^k = x^*$: $||x^{k+1} - x^*|| \leq \alpha ||x^k - x^*||$ mit $\alpha < 1$

Bestimmung von Nullstellen:

- Newton-Verfahren: $x^{k+1} = x^k - \frac{f(x^k)}{f'(x^k)}$
- Sekanten-Verfahren: $x^{k+1} = x^k - \frac{f(x^k)}{\frac{f(x^{k-1}) - f(x^k)}{x^{k-1} - x^k}}$
- Newton-Verfahren für $x \in \mathbb{R}^n$:
  - $z^k := x^{k+1} - x^k$
  - Löse Newton-Verfahren (jetzt Matrizen und Vektoren!) nach z^k auf mit Gauß $\Rightarrow f'(x^k) \cdot z^k = -f(x^k)$
  - Berechnung: $x^{k+1} = x^k + z^k$
  - Lokale Konvergenz:
    - $f : D \subset \mathbb{R}^n \to \mathbb{R}^n$, D konvex
    - $f(x^*) = 0$ für ein $x^* \in D$
    - $B(x^*, r) = \{x \mid ||x - x^*|| < r\} \subset D$
    - $f'(x^*)$ nicht singulär mit $||f'(x^*)^{-1}|| \leq \beta$
    - $||f'(x) - f'(y)|| \leq \gamma \cdot ||x - y||$ für $x, y \in D$

    Dann ist für alle $x_0 \in B(x^*, \delta)$ mit $\delta = \min \left\{ r, \frac{1}{2\beta\gamma} \right\}$ das Newton-Verfahren wohldefiniert und $||x^{k+1} - x^*|| \leq \beta \cdot \gamma \cdot ||x^k - x^*||^2 \leq \frac{1}{2} \cdot ||x^k - x^*||$.

Reelle Zahlen, Maschinenzahlen, gerundetes Rechnen

Dezimalsystem: $x = \pm(c_n 10^n + c_{n-1}10^{n-1} + ... + c_0 10^0 + c_{-1}10^{-1} + c_{-2}10^{-2} + ...) = \pm(c_n c_{n-1} c_{n-2} ... c_0, c_{-1} c_{-2} c_{-3} ...)_{10}$

- $0 \leq c_i \leq 9$
- $1 \leq c_n \leq 9$

Dualsystem / Binärsystem: $x = \pm(b_n 2^n + b_{n-1}2^{n-1} + ... + b_0 2^0 + b_{-1}2^{-1} + b_{-2}2^{-2} + ...) = \pm(b_n b_{n-1} b_{n-2} ... b_0, b_{-1} b_{-2} ...)_2$

- $b_i \in \{0, 1\}$
- $b_n = 1$

Umrechnung Binärsystem ↔ Dezimalsystem

- Nicht jeder endliche Dezimalbruch lässt sich als endlicher Binärbruch darstellen. Jedoch lässt sich jeder endliche Binärbruch in einen endlichen Dezimalbruch umwandeln.
- Ganze Binärzahl → Dezimalzahl: Horner-Schema

	a_n	a_{n-1}	a_{n-2}	...	a_1	a_0
β	$\downarrow$	$^{+}\downarrow a'_n \cdot \beta$	$^{+}\downarrow a'_{n-1} \cdot \beta$		$^{+}\downarrow a'_2 \cdot \beta$	$^{+}\downarrow a'_1 \cdot \beta$
	$a'_n \nearrow^{\cdot\beta}$	$a'_{n-1} \nearrow^{\cdot\beta}$			$a'_1 \nearrow^{\cdot\beta}$	$a'_0 = p(\beta)$

  - $p(\beta) = a_n\beta^n + a_{n-1}\beta^{n-1} + ... + a_1\beta^1 + a_0\beta^0, \ \beta \in \mathbb{R}$
  - Dualsystem: $\beta = 2$
  - Oben bei $a_n a_{n-1} ... a_0$ die Binärzahl reinschreiben, rechts unten kommt Dezimalzahl $x = p(2)$ heraus
  - Beispiel:

	1	1	1	1	1	0	0	1	1	1	1
2		2	6	14	30	62	124	248	498	998	1998
	1	3	7	15	31	62	124	249	499	999	**1999**

- Dezimalzahl → Binärzahl
  - $x_0 = x$ (gegeben im Dezimalsystem)
  - $x_{k+1} = \frac{x_k - a_k}{2}$, $k = 0, 1, 2, ...$ bis $x_{k+1} = 0$ mit $a_k = \begin{cases} 1 & x_k \ ungerade \\ 0 & x_k \ gerade \end{cases}$
  - Die $a_k s$ sind die Binärziffern in umgekehrter Reihenfolge
  - Beispiel:

k	0	1	2	3	4	5	6	7	8	9	10	
x_k	1999	999	499	249	124	62	31	15	7	3	1	
a_k	1	1	1	1	0	0	1	1	1	1	1	← von hinten!!!

Maschinenzahlen $\mathbb{R}(t,s)$: $\pm \underbrace{b_{-1}b_{-2}b_{-3}...b_{-t}}_{t\ Kaestchen} \quad \pm \underbrace{c_{s-1}c_{s-2}...c_0}_{s\ Kaestchen}$

- $b_{-i} \in \{0,1\}\,,\ c_i \in \{0,1\}$
- Darstellung: $\pm \underbrace{(0.b_{-1}b_{-2}...b_{-t})_2}_{Matisse} \cdot 2^{\pm \underbrace{(c_{s-1}\cdot 2^{s-1} + c_{s-2}\cdot 2^{s-2} + ... + c_1\cdot 2^1 + c_0\cdot 2^0)}_{=:e\ Exponent}}$
- Gleitpunktdarstellung, falls $b_{-1} = 1$ normalisiert
- Gleitpunktzahlensystem: Alle Gleitpunktzahlen einschließlich 0

Rundung von $x \in \mathbb{R}$: $x = \pm\left(\sum_{k=1}^{\infty} b_{-k}\cdot 2^{-k}\right)\cdot 2^e \Rightarrow x^* \in \mathbb{R}(t,s) : x^* = \pm\left(\sum_{k=1}^{t} b^*_{-k}\cdot 2^{-k}\right)\cdot 2^{e^*}$

- Abschneiden (chopping)
  - $x^* = \text{chop}\,(x) : e^* = e,\ b_{-k} = b_{-k}$ für $k = 1,2,...,t$
    falls $e^* > e_{max} \Rightarrow$ Überlauf (Abbruch)
    falls $e^* < e_{min} \Rightarrow$ Unterlauf ($x^* := 0$)
  - Absoluter Fehler: $|x - \text{chop}\,(x)| \le 2^{-t}\cdot 2^e$
  - Relativer Fehler: $\frac{|x-\text{chop}\,(x)|}{|x|} \le 2^{-t}\cdot 2$
- Symmetrische Rundung:
  - Dezimalsystem: $x = 0.456782\cdot 10^4 \Rightarrow x^* = 0.4568\cdot 10^4$
  - Binärsystem: $b_{-(t+1)} \in \{0,1\}$
    $b_{-(t+1)} = 1 \Rightarrow b^*_{-t} = b_{-t} + 1$
    $b_{-(t+1)} = 0 \Rightarrow b^*_{-t} = b_{-t}$
  - Relativer Fehler:
    Binärsystem: $\frac{|x-\text{rd}\,(x)|}{|x|} \le 2^{-t} =: eps$
    Dezimalsystem: $\frac{|x-\text{rd}\,(x)|}{|x|} \le 5\cdot 10^{-t} =: eps$
  - eps heißt Maschinengenauigkeit
  - $\text{rd}\,(x) := (1+\varepsilon)\cdot x$ mit $|\varepsilon| \le eps$

Fehlerfortpflanzung von $\tilde{x} = (1+\varepsilon_x)\cdot x$ und $\tilde{y} = (1+\varepsilon_y)\cdot y$:

- Multiplikation: $\varepsilon_{xy} = \varepsilon_x + \varepsilon_y$
  „Bei der Multiplikation addieren sich die relativen Fehler“
- Division: $\varepsilon_{\frac{x}{y}} = \varepsilon_x - \varepsilon_y$
- Addition, Subtraktion: $\varepsilon_{x+y} = \frac{x}{x+y}\cdot\varepsilon_x + \frac{y}{x+y}\cdot\varepsilon_y$
  - Auslöschung (großer Fehler) für $x \approx -y$
  - $\Rightarrow$ statt $\sqrt{x+\delta} - \sqrt{x}$ lieber $\frac{\delta}{\sqrt{x+\delta}+\sqrt{x}}$ berechnen
  - $\Rightarrow$ statt $\cos(x+\delta) - \cos(x)$ lieber $-2\sin\frac{\delta}{2}\sin(x+\frac{\delta}{2})$ berechnen (Additionstheoreme...)
  - $\Rightarrow$ statt allgemein $f(x+\delta) - f(x)$ lieber $f(x+\delta) = f(x) + f'(x)\cdot\delta$ verwenden

Kondition: $(\text{cond}\,f)(x) = x\cdot\frac{f'(x)}{f(x)}$

- Allgemein: $(\text{cond}\,f)(x) = \frac{||x||_\infty\cdot||\frac{\partial f}{\partial x}||_\infty}{||f(x)||_\infty}$
- Beliebige Norm: $(\text{cond}\,f)(x) = c\cdot\frac{||x||\cdot||\frac{\partial f}{\partial x}||}{||f(x)||}$
- Die Empfindlichkeit (Kondition) einer Polynomnullstelle kann außerordentlich groß $\Rightarrow$ Eigenwertberechnung nie über charakteristisches Polynom, sondern anderes Verfahren.

Kondition eines Algorithmus: $(\text{cond } A)(x) = \frac{\inf_{x_A} \frac{||x_A - x||}{||x||}}{eps}$

- Grundannahme: für jedes $x \in \mathbb{R}^m (t, s)$ gilt: $f_A(x) = f(x_A)$ mit einem $x_A \in \mathbb{R}^m$

Gesamtfehler: $\frac{||y_A^* - y||}{||y||} \leq (\text{cond } f)(x) \, [\varepsilon + (\text{cond } A)(x^*) \cdot eps]$

- $\frac{||x^* - x||}{||x||} \leq \varepsilon, \ \varepsilon \leq eps$