Rechnerarithmetik

Maschinendarstellbare Zahlen

- $M = \left\{x \in R \mid x = \pm 0.m_1m_2 \ldots m_n \cdot B^{\pm 0.e_1e_2 \ldots e_l}\right\} \cap \{0\}$
- $2 \cdot 2^m \cdot \left(2 \cdot 2^l - 1\right) + 1$ verschiedene Zahlen (mit Vorzeichen)
- $x_{\min} = 0.1 \cdot 10^{-e..e}$
- $x_{\max} = 0.99..9 \cdot 10^{e..e}$
- normaisiert = 1. Ziffer keine 0

IEC/IEEE:

- $x = (-1)^V \cdot (1.M..M) \cdot 2^{(E..E)-\text{bias}}$
- single precision: 1 Bit Vorzeichen, 23 Bit Mantisse, 8 Bit Exponent
- double precision: 1 Bit Vorzeichen, 52 Bit Mantisse, 11 Bit Exponent
- $x_{\max} = \left(1 - B^{-n}\right) B^{e_{\max}}$
- $x_{\min} = B^{e_{\min}-1}$
- $m = n - 1$

Fehler

Absoluter Fehler = $|\tilde{x} - x|$

Relativer Fehler = $\frac{|(\tilde{x} - x)|}{x}$

Maximaler abs. Fehler = $\frac{B}{2} \cdot B^{e-n-1}$

Maschinengenauigkeit eps = maximaler rel. Fehler = $\frac{1}{2} \cdot B^{1-n}$

1 + eps = 1

Rundungsfehler

Bsp: 2590 + 4 + 4 in 3 Stellen = rd(2590 + 4) = 2590

Man sollte in aufsteigender Reihenfolge summieren.

Näherungen

Konditionszahl $K = \frac{|f'(x)| \cdot |x|}{|f(x)|}$

Näherung absoluten Fehlers = $\left|f'(x)\right| \cdot |\tilde{x} - x|$

Näherung des relativen Fehlers = $K \cdot \frac{|\tilde{x} - x|}{|x|}$

Näherung zu x mit maximalem abs. Fehler a = $[x - a, x + a]$

Auslöschng

Wenn f(x) = x + c entgegengesetze Vorzeichen haben und etwas gleich gross sind, dann wird das Ergebnis sehr nahe bei 0 sein. Dadurch ist K sehr gross und das Problem schlecht Konditioniert.

Bei der Multiplikation ist K = 1. Dies ist gut Konditioniert weil die Zahl klein ist.

Nichtlineare Gleichungen

Heron-Verfahren

Berechnet eine Quadratwurzel der Zahl $A > 1$

$$x_{n+1} = \frac{x_n + \frac{A}{x_n}}{2}$$

Fixpunktiteration

1. Gleichung stellen in der Form $F(x) = x$
2. Einen Startwert x_0 wählen
3. $x_{k+1} = F(x_k)$

Dies Konvergiert zu einem Fixpunkt, wenn $F'(x) < 1$ (je kleiner, desto schneller)

Banascher Fixpunktsatz

Wenn gilt: $|F(x) - F(y)| \leqslant \alpha \cdot |x - y|$ für alle x, y in [a, b] mit F[a, b] → [a, b], bzw:

- y-Werte müssen zwischen a und b sein
- Der maximale Betrag der Steigung F' darf nicht grösser als 1 sein
- (α = Lipschitz-Konstante < 1)

Dann gilt:

- F hat genau einen Fixpunkt in [a, b]
- Die Fixpunktiteration konvergiert für Startwerte in [a, b]
- a-priori Fehlerabschätzung: $|x_n - \bar{x}| \leqslant \frac{\alpha^n}{1-\alpha} \cdot |x_1 - x_0|$
- a-posteriori: $|x_n - \bar{x}| \leqslant \frac{\alpha}{1-\alpha} \cdot |x_n - x_{n-1}|$
- $\alpha = \max |F'(x)|$

Newton-Verfahren

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

Konvergiert, wenn $\left| \frac{f(x) \cdot f''(x)}{[f'(x)]^2} \right| < 1$

Vereinfachtes Newton-Verfahren

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_0)}$$

Sekantenverfahren

$$x_{n+1} = x_n - \frac{x_n - x_{n-1}}{f(x_n) - f(x_{n-1})} \cdot f(x_n)$$

Konvergenzordnung

Ein Verfahren hat die Konvergenzordnung q, wenn für alle n gilt: $|x_{n+1} - \bar{x}| \leqslant c \cdot |x_n - \bar{x}|^q$

q = 1 = Lineare Konvergenz

Fehlerabschätzung

Falls bei einer Nullstelle ξ und einer Fehlerschranke / Fehlertoleranz ε gilt: $f(x_n - \varepsilon) \cdot f(x_n + \varepsilon) < 0$, dann gilt: $|x_n - \xi| < \varepsilon$

Für Abbruchbedingungen kann auch geschaut werden, ob $|x_{n+1} - x_n| < \varepsilon$.

Lineare Gleichungen

$Ax = b$

Gauss-Algorithmus

Die Matrix wird in eine (unnormierte) obere Dreiecksmatrix gebracht. Es gilt $\det(A) = (-1)^I \cdot \det(A) = (-1)^I \cdot \Pi_{i=1}^n \tilde{a}_{ii}$ wobei I die Anzahl Zeilenvertauschungen und $\tilde{a}_{ii}$ die Diagonalelemente sind. Der Faktor λ ist der Faktor bei der Elimination.

LR-Zerlegung ohne Zeilenvertauschen

$$L = \begin{pmatrix} 1 & 0 & 0 \\ \lambda & 1 & 0 \\ \lambda & \lambda & 1 \end{pmatrix} R = \begin{pmatrix} ? & ? & ? \\ 0 & ? & ? \\ 0 & 0 & ? \end{pmatrix} \lambda = \frac{A_{(i+1)j}}{A_{ij}}$$

1. Man löst das Gleichungssystem $Ly = b$ und erhält y
2. Vorwärtseinsetzen: $Ly = b$
3. Rückwärtseinsetzen: $Rx = y$

LR-Zerlegung mit Zeilenvertauschen

Zusätzlich werden in der Matrix $P = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ die Zeilen vertauscht. Man muss also A→R, L und P gleichzeitig berechnen.

Es gilt neu $Ly = P \cdot b$

Householder-Matrix

$|u| = \sqrt{u_1^2 + \ldots + u_n^2}$

$\tilde{u} = \frac{u}{|u|}$

$H = I_n - 2\tilde{u}\tilde{u}^T$

H ist symmetrisch: $H = H^T$

$$(1,2,3) \cdot \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{pmatrix}$$

QR-Zerlegung

1. $v_n = a_n + sign(a_{n1}) \cdot \|a_n\|_2 \cdot e_n$ wobei $e_1 = (1,0,0,..)^T$ und a_n = 1. Spalte von A_n $u_n = \frac{1}{|v_n|} v_n$ $H_n = I_n - 2u_n u_n^T$
2. $Q_n \cdot \ldots Q_1 \cdot A_n = \begin{pmatrix} * & * & * \\ 0 & A_{n+1} & \\ 0 & & \end{pmatrix}$ $Q_1 = H_1$ ... $Q_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & H_2 & \\ 0 & & \end{pmatrix}$
3. $Q = Q_1 \cdot \ldots \cdot Q_n$ $R = Q_n \cdot \ldots \cdot Q_1 \cdot A$ $QR = A$
4. $Rx = Q^T b$

Vektornorm

1-Norm, Summennorm : $\| \boldsymbol{x} \|_1 = \sum_{i=1}^{n} | x_i |$

2-Norm, euklidische Norm : $\| \boldsymbol{x} \|_2 = \sqrt{\sum_{i=1}^{n} x_i^2}$

∞-Norm, Maximumnorm : $\| \boldsymbol{x} \|_\infty = \max_{i=1,\ldots,n} | x_i |$

Matrixnorm

1-Norm, Spaltensummennorm : $\| \boldsymbol{A} \|_1 = \max_{j=1,\ldots,n} \sum_{i=1}^{n} | a_{ij} |$

2-Norm, Spektralnorm : $\| \boldsymbol{A} \|_2 = \sqrt{\rho(\boldsymbol{A}^T \boldsymbol{A})}$

∞-Norm, Zeilensummennorm : $\| \boldsymbol{A} \|_\infty = \max_{i=1,\ldots,n} \sum_{j=1}^{n} | a_{ij} |$

Abschätzung fehlerbehafteter Vektor

$cond\,(A) = \|A\|_\infty \cdot \|A^{-1}\|_\infty$

$$\frac{\|x-\tilde{x}\|_\infty}{\|x\|_\infty} \leqslant cond\,(A) \cdot \frac{\|b-\tilde{b}\|_\infty}{\|b\|_\infty}$$

Abschätzung für fehlerbehaftete Matrix

$$\frac{\|x-\tilde{x}\|_\infty}{\|x\|_\infty} \leqslant \frac{cond\,(A)}{1 - cond\,(A) \cdot \frac{\|A-\tilde{A}\|_\infty}{\|A\|_\infty}} \cdot \left(\frac{\|A-\tilde{A}\|_\infty}{\|A\|_\infty} + \frac{\|b-\tilde{b}\|_\infty}{\|b\|_\infty} \right)$$

Wenn A (3×3) um maximal 2 elementweise gestört ist, dann ist $\|A-\tilde{A}\|_\infty = 6$

Aufwand

Gauss $O\left(\frac{2}{3}n^3\right)$

LR $O\left(\frac{2}{3}n^3\right)$

QR $O\left(\frac{5}{3}n^3\right)$

Einsetzen $O\left(n^2\right)$

Jakobi-Verfahren / Gesamtschrittverfahren

R

$$\boldsymbol{A} = \begin{pmatrix} 4 & -1 & 1 \\ -2 & 5 & 1 \\ 1 & -2 & 5 \end{pmatrix}$$

L D

$A = L + D + R$

$x^{k+1} = -D^{-1}\,(L+R)x^k + D^{-1}b$

$$x_1^{k+1} = \left(-\left(a_{12}x_2^k + a_{13}x_3^k\right) + b_1\right)\frac{1}{a_{11}}$$

$$x_2^{k+1} = \left(-\left(a_{21}x_1^k + a_{23}x_3^k\right) + b_2\right)\frac{1}{a_{22}}$$

$$x_1^{k+1} = \left(-\left(a_{31}x_1^k + a_{32}x_2^k\right) + b_3\right)\frac{1}{a_{33}}$$

Iteratives Gauss-Seidel-Verfahren

$x^{k+1} = -(D+L)^{-1}Rx^k + (D+L)^{-1}b$

$$x_1^{k+1} = \left(-\left(a_{12}x_2^k + a_{13}x_3^k\right) + b_1\right)\frac{1}{a_{11}}$$

$$x_2^{k+1} = \left(-\left(a_{21}x_1^{k+1} + a_{23}x_3^k\right) + b_2\right)\frac{1}{a_{22}}$$

$$x_1^{k+1} = \left(-\left(a_{31}x_1^{k+1} + a_{32}x_2^{k+1}\right) + b_3\right)\frac{1}{a_{33}}$$

Fehlerabschätzung der Fixpunktiteration

Falls $\|B\|_\infty < 1$

a-priori $\|x^n - \bar{x}\|_\infty \leqslant \frac{\|B\|_\infty^n}{1-\|B\|_\infty}\|x^1 - x^0\|_\infty$

a-posteriori $\|x^n - \bar{x}\|_\infty \leqslant \frac{\|B\|_\infty^n}{1-\|B\|_\infty}\|x^1 - x^0\|_\infty$

Jacobi $B = -D^{-1}\,(L+R)$

Gauss-Seidel $B = -(D+L)^{-1}R$

Konvergenz der Fixpunktiteration

Falls A diagonaldominant ist, kovergiert das Verfahren

= Die Werte in der Diagonalen sind grösser als Werte oben und unten davon ODER rechts und links davon

Eigenwerte und Eigenvektoren

Imaginäre Zahlen

Normalform $z = x + iy$

Trigonometrische Form $z = r\left(\cos\varphi + i \cdot \sin\varphi\right)$ wobei $x = r \cdot \cos\varphi$

Exponentialform $z = re^{i\varphi}$ wobei $e^{i\varphi} = \cos\varphi + i \cdot \sin\varphi$

$r = |z| = \sqrt{y^2 + x^2}$

$z^* = x + -iy$

$\frac{z_1}{z_2} = \frac{x_1x_2 + y_1y_2}{x_2^2 + y_2^2} + i\frac{y_1x_2 - x_1y_2}{x_2^2 + y_2^2}$

$z^n = x$ Lösen:

- $k = 0, 1, \ldots, n-1$
- $r = \sqrt[n]{x}$
- $z^n = x \cdot e^{i(0 + k*2\pi)}$
- $z_k = re^{i\frac{(0+k\cdot 2\pi)}{n}}$

$3 + 2j = \sqrt{3^2 + 2^2} \cdot e^{j \cdot \tan\left(\frac{2}{3}\right)}$

$\varphi = \pi - \arctan\left(\frac{y}{x}\right)$

“Pasted image 20250110165905.png|300” could not be found.

Eigenvektor

Wenn gilt $Ax = \lambda x$ dann ist x ein Eigenvektor. Dieser wird auf Länge 1 normiert.

det(A) = Produkt aller Eigenwerte λ

tr(A) = Summe aller Eigenwerte

Eigenwerte Diagonalmatrix/Dreiecksmatrix = Diagonalelemente

Eigenwerte von $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$: $p(\lambda) = \det\left(A - \lambda I_n\right) = (a - \lambda)(d - \lambda) - bc$

$\det B_{3x3} = b_{11}\left(b_{22}b_{33} - b_{23}b_{32}\right) - b_{12}\left(b_{21}b_{33} - b_{23}b_{31}\right) + b_{13}\left(b_{21}b_{32} - b_{22}b_{31}\right)$

Eigenvektor zu Eigenwert λ von A: $\vec{x} = \left(A - \lambda I_n \mid \vec{0}\right)$

Nicht reelle Nullstellen treten in Paaren auf (+-i)

Eigenraum

Hat die Dimensionen $n - Rg\left(A - \lambda I_n\right)$

Diagonalisierbare Matrixen

- $B = T^{-1}AT$
- Haben gleiche EW/EV

- Die Spalten der Transformationsmatrix sind die EV wenn A symmetrisch ist

Numerische Berechnung

Durch QR-Zerlegung von A_n:

- $A = Q_n R_n$
- $R_n = Q_n^T A$
- $A_{n+1} = R_n Q_n = Q_n^T A Q_n$

Alle A_n sind einender Ähnlich und evt. diagonalisierbar. Wenn die matrix Teilweise in die Dreicecksform gebracht werden kann:

ZB $A = \begin{pmatrix} * & * & * \\ * & * & * \\ 0 & 0 & 1 \end{pmatrix}$ dann gibt es 1 reellen Eigenwert 1. Und 2 Komplexe

Werte, die wie sonst für eine 2×2 Matrix bestimmt werden können. Die Eigenvektoren könnnen dann als 3×3 Gleichungsystem gelöst werden.

Spektralradius

= Grösster Eigenwert

Berechenbar durch Vektoriteration / von-Mises-Iteration:

$$\begin{aligned} \boldsymbol{v}^{(k+1)} &= \frac{\boldsymbol{A}\boldsymbol{v}^{(k)}}{\| \boldsymbol{A}\boldsymbol{v}^{(k)} \|_2} \\ \lambda^{(k+1)} &= \frac{(\boldsymbol{v}^{(k)})^T \boldsymbol{A}\boldsymbol{v}^{(k)}}{(\boldsymbol{v}^{(k)})^T \boldsymbol{v}^{(k)}} \end{aligned}$$

Mitternachtsformel

$$\frac{-B \pm \sqrt{B^2 - 4AC}}{2A}$$

Ableitungen

$$f \cdot g = f' \cdot g + f \cdot g'$$

$$\frac{f}{g} = \frac{f' \cdot g - g' \cdot f}{g^2}$$

$$f(g) = f'(g) \cdot g'$$

$$\cos' = -\sin$$

$$\frac{1}{x}' = -\frac{1}{x^2}$$

sin cos

- cos(0) = 1, sin(0) = 0
- cos(pi) = -1, sin(pi) = 0
- cos(pi/2) = 0, sin(pi/2) = 1
- sin(x) = cos(x - pi/2)