Horner Schema → to Decimal

$11001{,}1011_2$

vor Punkt | nach Punkt

src Base	2^4	2^3	2^2	2^1	2^0		$(\frac{1}{2})^4$	$(\frac{1}{2})^3$	$(\frac{1}{2})^2$	$(\frac{1}{2})^1$	$(\frac{1}{2})^0$
	1	1	0	0	1		1	1	0	1	0
B=2	↓	2	6	12	24	$\frac{1}{B}$	↓	0.5	0.75	0.375	0.6875
	1	3	6	12	25_{10}		1	1.5	0.75	1.375	0.6875_{10}

Min/Max Floatingpoint

$x_{max} = B^e(1-B^{-n})B^{e_{max}}$

$x_{min} = B^{e_{min}-1}$

$-x_{max}$ | $-x_{min}$ 0 $+x_{min}$ | $+x_{max}$

e = Exponent

B = Base

n = Mantissenlänge

Konversion Basen

1006_{10} zu binär vor Punkt

$1006:2 = 503$ Rest 0 ↑ LSB
$503:2 = 251$ Rest 1
$251:2 = 125$ Rest 1
⋮
MSB

$.687_{10}$ zu Binär

$2\cdot0.687 = 1.374$ Ziffer 1 | MSB
$2\cdot0.374 = 0.748$ Ziffer 0
$2\cdot0.748 = 1.496$ Ziffer 1
⋮
↓ LSB

Abs/rel Fehler

$\tilde{x}$ = Näherung von x

$|\tilde{x}-x|$ – abs Fehler

$\frac{|\tilde{x}-x|}{|x|}$ = rel. Fehler

eps

$1+eps \overset{!}{=} 1$ kleinste Zahl

$\frac{B}{2}\cdot B^{-n}$

n = Anzahl Stellen zur Verfügung

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

Fehler Funktionen

rel: $\frac{|f(\tilde{x})-f(x)|}{|f(x)|}$

Näherung rel: $\frac{|f'(x)|\cdot|x|}{|f(x)|}\cdot\frac{|\tilde{x}-x|}{|x|}$

Näherung abs: $|f'(x)|\cdot|\tilde{x}-x|$

abs. Fehler $|f(\tilde{x})-f(x)|$

Konditionszahl

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

groß → schlecht kond, grosse Fehlerfort.

klein: gut kond

Auslöschung

ca gleich grosse fehlerhafte Zahlen werden voneinander abgezogen und resultat wird normiert → hohes K

$K\cdot$ rel. Fehler $\tilde{x} \hat{=}$ rel Fehler $f(\tilde{x})$

Bisektionsverfahren

$f[a;b]$ stetig $f(a)\cdot f(b) < 0$

$[a_0\ b_0] = [a, b]$

$$[a_{i+1}, b_{i+1}] = \begin{cases} [a_i, \frac{a_i+b_i}{2}] & \text{falls } \frac{a_i+b_i}{2}\cdot f(a_i) \le 0 \\ \text{sonst } [\frac{a_i+b_i}{2}, b_i] \end{cases}$$

$\lim_{\to\infty}(b_i - a_i)$

Vorteil Bisektion

funktioniert für allg. stetige Funktionen

konvergiert immer wenn a und b gegeben

Anzahl Schritte für Genauig. hängt nicht von f sondern von a und b ab.

Nullstellenprobleme

Newtonverfahren
Bisektionsverfahren
Fixpunktiteration

Newton

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

Konv falls $\left| \frac{f(x) \cdot f''(x)}{f'(x)^2} \right| < 1$

für alle x im Intervall
→ Stetigkeit beweisen
dann Randstellen

Newton einfach

langsamere Konvergenz

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

Sekanten

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

konvergiert Geschwindigkeit zw. Newton und einf. Newton

Konvergenzgsw

$q \geq 1$ Konvergenz
$q = 1$ lineare Konvergenz
$q = 2$ quad. Konvergenz

$$\underbrace{|x_{n+1} - \bar{x}|}_{\text{nächste Iteration abs Fehler}} \leq c \cdot \underbrace{|x_n - \bar{x}|^q}_{\text{jetz. abs Fehler}}$$

Fixpunktiteration

Gleichung in Form $F(x) = x$
Lösung $\bar{x}$ für $F(\bar{x}) = \bar{x}$ heissen Fixpunkte
$F: [a,b] \to R \quad x_0 \in [a,b]$

$$x_{n+1} = F(x_n)$$

Grafische Nullpunkte → $F(x) = f(x)$
unterschiedliche Konvergenzverhalten, je nach Startwert

$F'(x) < 1$ → konvergiert
$F'(x) > 1$ → divergiert

Banachscher Fixpunktsatz

2 Voraussetzungen
- $F[a,b] \to [a,b]$
- $\frac{F(x) - F(y)}{x - y} \leq \alpha \quad \forall x \in [a,b] \qquad \alpha = \max F'(x_0)$

dann
- genau 1 Fixpunkt in $[a,b]$
- Konvergenz gegen Punkt für alle Startwerte in $[a,b]$
- folgende Fehlerabschätzungen abs. Fehler

$$|x_n - \bar{x}| \leq \frac{\alpha^n}{1-\alpha} |x_1 - x_0| \quad \text{priori}$$

$$|x_n - \bar{x}| \leq \frac{\alpha}{1-\alpha} |x_n - x_{n-1}| \quad \text{posteriori}$$

Fehlerabschätzung

x_n = Näherungswert Nullstelle von ξ (ungerade Ordnung)
$\varepsilon > 0$ vorg. Fehlertoleranz

$f(x_n - \varepsilon) \cdot f(x_n + \varepsilon) < 0$ → $(x_n - \varepsilon, x_n + \varepsilon)$ hat Nullstelle ξ und Fehlerabschätzung $|x_n - \xi| < \varepsilon$

Dreiecksformen

untere normierte $\begin{pmatrix} 1 & 0 \\ L_{n1} & 1 \end{pmatrix} = L$

obere Dreiecksmatrix $\begin{pmatrix} 0 & R_n \\ 0 & 0 \end{pmatrix} = R$

Gauss mit Pivot

Pivot → Rechenschritt repeat

choose max

$\begin{pmatrix} 1 & 2 & -1 \\ 4 & -2 & 6 \\ 3 & 1 & 0 \end{pmatrix} Z_1 \leftrightarrow Z_2$

od. 2. Schritt $\begin{pmatrix} 4 & 2 & -1 \\ 1 & 2 & -1 \\ 3 & 1 & 0 \end{pmatrix}$ 2>1 keine Zeilentausch

Berechnung von L

$L = \begin{pmatrix} 1 & 0 & 0 \\ L_{21} & 1 & 0 \\ L_{31} & L_{32} & 1 \end{pmatrix}$ $A = \begin{pmatrix} -1 & 1 & 1 \\ 1 & -3 & -2 \\ 5 & 1 & 4 \end{pmatrix}$ Gauss 1 Schritt

$L_{21} = \frac{1}{-1} = -1$

$Z_2 = Z_2 + \frac{1}{-1} Z_{max}$

$\tilde{A} = \begin{pmatrix} -1 & 1 & 1 \\ 0 & -2 & -1 \\ 5 & 1 & 4 \end{pmatrix}$ $Z_3 = Z_3 + \frac{5}{-1} Z_1$

$L_{31} = \frac{5}{-1} = -5$

$A = \begin{pmatrix} -1 & 1 & 1 \\ 0 & -2 & -1 \\ 0 & 6 & 9 \end{pmatrix}$ $Z_3 = Z_3 + \frac{6}{-2} Z_2$

$L_{32} = \frac{6}{-2} = -3$

Normen Vektor

$\|x\|_1$ Summennorm betragsmässige Summe

$\|x\|_2$ Euklidsche Norm Länge Vektor

$\|x\|_\infty$ Maximumnorm betragsmässig grösstes Element

Normen Matrix

$\|A\|_1$ Spaltensummennorm Max betragsmässige Summe einer Spalte

$\|A\|_\infty$ Zeilensummennorm Max. betragsmässige Summe einer Zeile

Gauss ohne Pivotisierung

$Ax = b$ nach x lösen

Transformation zu oberer Dreiecksmatrix

$$\left(\begin{array}{ccc|c} 200 & 150 & 100 & 2150 \\ 50 & 30 & 20 & 470 \\ 20 & 10 & 0 & 150 \end{array}\right) \quad Z_2 = Z_2 - \tfrac{1}{4} Z_1$$

auf 0 bringen

$$\begin{array}{ccc|c} x & y & z & \\ 200 & 150 & 100 & 2150 \\ 0 & -7.5 & -5 & -67.5 \\ 20 & 10 & 0 & 150 \end{array}$$

etc.

dann nach x, y, z auflösen

Dreieckszerlegung

Annahme: keine Zeilenvertauschung

bisher: $Ax = b$ neu $A = LR$

$LRx = b \rightarrow$ Sub $Rx = y \rightarrow Ly = b$

Lösen von $Ly = b$ in 2 Schritten

1. $Ly = b$ // man erhält y L, b bekannt
2. $Rx = y$ // R und y bekannt

LR Zerlegung mit Zeilenvertauschung

$P \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ $A \cdot P_n = A^*$ bei Zeilenvertauschung muss $A \cdot P_n$ gerechnet werden, bei L wird zudem unterhalb der Diagonale auch vertauscht

1, 2 Zeile vertauschen: P 1 und 2 vertauschen

$P_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \rightarrow A^* \cdot P_1$

Fehlerrechnung lin. Gls

$Ax = b$ korr. Erg. $\| \ \|$ = Norm

$A\tilde{x} = \tilde{b}$ gest. Erg.

abs Fehler: $\|x - \tilde{x}\| \le \|A^{-1}\| \cdot \|b - \tilde{b}\|$

rel Fehler $\frac{\|x - \tilde{x}\|}{\|x\|} \le \underbrace{\|A\| \cdot \|A^{-1}\|}_{\text{Konditionszahl } K \text{ bezg. gewählter Norm}} \cdot \frac{\|b - \tilde{b}\|}{\|b\|}$

Absch. Fehlerhafte Matrizen

falls $\operatorname{cond}(A) \cdot \frac{\|A-\tilde{A}\|}{\|A\|} < 1$

$$\rightarrow \frac{\|x-\tilde{x}\|}{\|x\|} \leq_{\infty} \frac{\operatorname{cond}(A)}{1-\operatorname{cond}(A)\cdot\underbrace{\frac{\|A-\tilde{A}\|}{\|A\|}}} \cdot \left(\underbrace{\frac{\|A-\tilde{A}\|}{\|A\|}} + \frac{\|b-\tilde{b}\|}{\|b\|}\right)$$

0 falls kein Fehler

bei Matrizen muss gestört Faktor mit Dimensionsfaktor multipliziert werden

BSP $A = \underbrace{\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}}_{n=2}$ A um max 1 gestört $\rightarrow \|A-\tilde{A}\| = 1 \cdot \underbrace{2}_{n}$

Aufwandabschätzung

Anzahl Gleitkommaoperationen

Gauss $O(n^3)$

LR Gauss $O(n^3)$

Choleski $O(n^3)$

Cramersche Regel $O(n!)$

Vor/Rückwärtseinsetzen $O(n^2)$

Iterative Verfahren

Andere L und R als bei Zerlegung

Ziel $x_{n+1} = F(x) \rightarrow$ neu $x^{(n+1)} = F(x^{(n)})$

neue LDR

$$A \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \qquad \text{tril}(A,-1) \quad L = \begin{pmatrix} 0 & 0 & 0 \\ 4 & 0 & 0 \\ 7 & 8 & 0 \end{pmatrix}$$

$$D \begin{pmatrix} 1 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 9 \end{pmatrix} \qquad \text{triu}(A,1) \quad R = \begin{pmatrix} 0 & 2 & 3 \\ 0 & 0 & 6 \\ 0 & 0 & 0 \end{pmatrix}$$

diag(diag(A))

Generell

$A = L + D + R$

$Ax = B$

$(L + D + R)x = B$

Inverse D

$$D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix} \quad D^{-1} = \begin{pmatrix} 1/1 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1/3 \end{pmatrix}$$

Jacobi/Gesamtschrittverfahren

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

Gauss/Seidel/Einzelschrittverfahren

= Jacobi, aber jede Komponente Einzel

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

Konvergenz

geg Fixpunktit. $x^{(n+1)} = Bx^{(n)} + b = F(x^{(n)})$

$\bar{x}$ erfüllt $\bar{x} = B\bar{x} + b = F(\bar{x})$

$\bar{x}$ anziehend Fixpunkt falls $\|B\| < 1$

$\bar{x}$ abstossend Fixpunkt falls $\|B\| > 1$

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

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

Iterationen ~~Fehler~~ abschätzungen

$$\|x^{(n)} - \bar{x}\| \leq \frac{\|B\|^n}{1-\|B\|} \cdot \|x^{(1)} - x^{(0)}\| \quad \text{priori}$$

$$\|x^n - \bar{x}\| \leq \frac{\|B\|}{1-\|B\|} \|x^{(n)} - x^{(n-1)}\|$$

Diagonalendominanz

Für jede Zeile: Diagonalenelement grösser als Summe anderer

oder

Für jede Spalte: Diagonalenelement grösser als Summe anderer

A diagonaldominant $\rightarrow$ Jacobi/Gauss-Seidel konvergieren

$\|B\| < 1 \rightarrow$ A Diagonaldominant

Funktionen mit mehreren Variablen

Skalarwertig $R^n \rightarrow R$ $\quad f(x_1, x_2, x_3) = 1$

Resultat genau 1 Zahl

vektorwertig $R^n \rightarrow R^m$ $\quad f(x_1, x_2, x_3) = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$

Resultat ist ein Vektor

Explizite Darstellung: Funktion nach Variable aufgelöst $\quad y = 2x + 3$

implicite Darstellung: Funktion nicht nach Variable aufgelöst $\quad F(x_1, x_2, x_3) = 0$

Linearisierung

$$g(x) = f(x^{(0)}) + Df(x^{(0)}) \cdot (x - x^{(0)})$$

$x^{(0)}$ = gegebener Vektor

Jacobi Matrix

geg: Gleichungssystem bsp 3 Gleichungen mit 3 Variablen I II III

$$Df = \begin{pmatrix} I\,dx & I\,dy & I\,dz \\ II\,dx & II\,dy & II\,dz \\ III\,dx & III\,dy & III\,dz \end{pmatrix}$$

Tangentialebene Spezialfall

$f: R^2 \rightarrow R$ $\quad f(x_1, x_2)$ liefert Gleichung Tangentialebene

$$g(x_1, x_2) = f(x_1^{(0)}, x_2^{(0)}) + \frac{\partial f}{\partial x_1}(x_1^{(0)}, x_2^{(0)}) \cdot (x_1 - x_1^{(0)}) + \frac{\partial f}{\partial x_2}(x_1^{(0)}, x_2^{(0)}) \cdot (x_2 - x_2^{(0)})$$

enthält alle im Flächenpunkt $P = (x_1^{(0)}, x_2^{(0)}, f(x_1^{(0)}, x_2^{(0)}))$ an die Bildfläche $y = f(x_1, x_2)$ angelegte Tangenten

Nullstellenbestimmung nicht linearer ~~Bestimmung~~ Systeme

Problem $f: R^n \to R^n$

Gls $\bar{x}$ mit $f(\bar{x}) = 0$

$$f(x) = f(x_1, x_2, x_3) = \begin{pmatrix} f_1(x_1, x_2, x_3) \\ f_2(x_1, x_2, x_3) \\ f_3(x_1, x_2, x_3) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Gedämpftes Newton

wenn Df schlecht konditioniert oder nur schwer/gar nicht invertierbar ist (Näherung kann sich von Nullstelle entfernen)

1. $\delta^{(n)}$ berechnen aus $Df(x^{(n)})\delta^{(n)} = -f(x^{(n)})$

2. Finde minimales $k = \{\overset{\text{start}}{0}, 1, \dots, k_{max}\}$ wofür gilt

$$\left\| f\left(x^{(n)} + \frac{\delta^{(n)}}{2^k}\right) \right\|_2 < \| f(x^{(n)}) \|_2$$

falls kein $k \to k = 0$

3. $$x^{(n+1)} = x^{(n)} + \frac{\delta^{(n)}}{2^k}$$

gedämpftes Newton im allg besser als norm. Newton

k_{max} problemabhängig, kann einmal divergieren und einmal konvergieren

für verschiedene k_{max} ~~können~~ können möglicherweise unterschiedliche Nullstellen angesteuert werden

$k_{max} = 4$ für uns

Newton

quad Konv. falls Df reg und f 3x ableitbar

evtl. wird lokales Min $\neq 0$ angesteuert, in diesem fall ist $Df(x_{min})$ nicht regulär

$$x^{(n+1)} = x^{(n)} + \delta^{(n)}$$

aus

$$Df(x^{(n)}) \cdot \delta^{(n)} = -f(x^{(n)})$$

Vereinfachtes Newton

weniger Aufwand/Schritt, nur lineare Konvergenz

$$x^{(n+1)} = x^{(n)} + \delta^{(n)}$$

aus

$$Df(x^{(0)})\delta^{(n)} = -f(x^{(n)})$$