

Zusammenfassung zu Numerische Mathematik II

Sara Adams

11. August 2004

Diese Zusammenfassung basiert auf der Vorlesung
Numerische Mathematik II
 gehalten im Sommersemester 2004
 von **Prof. Martin Buhmann (Ph.D.)**
 an der Justus-Liebig Universität Gießen

Inhaltsverzeichnis

I	Gewöhnliche Differentialgleichungen	3
1	Vorbemerkungen	3
2	Die ϑ -Methode	4
2.1	Eulersche Methode (Vorwärts-Euler)	4
2.2	Trapezregel	4
3	Allgemeine Mehrschrittverfahren	4
3.1	Adam-Bashforth-Methoden	4
3.2	Definition und Ordnung	5
3.3	Konvergenz und die Dahlquist-Sätze	5
3.4	BDF-Formeln	5
4	Runge-Kutta-Verfahren	5
4.1	explizite Runge-Kutta-Verfahren	6
4.2	implizite Runge-Kutta-Verfahren (IRK)	7
5	Steife Differentialgleichungen	7
5.1	Definitionen	7
5.2	A-Stabilität von MSV	7
5.3	Dissipativität	8
5.4	A-Stabilität von Runge-Kutta-Verfahren	8
6	Schrittweitensteuerung	8
7	Methode nach Milne	8
8	Zadunaiskys Methode	9
9	Prädiktor-Korrektor-Verfahren	9
II	Lösung dünn besetzter, großer linearer Gleichungssysteme	9
10	Cholesky-Verfahren	10
10.1	Matrizen und Bäume	10
11	Iterative Verfahren	10
11.1	Jacobi und Gauss-Seidel	11
11.2	SOR-Verfahren	11
11.3	CG-Verfahren	12
III	Mehrdimensionale Splines	13
12	Wiederholung eindimensionale Splines	13

13 Tensorprodukt	14
14 Splines von totalem Grad k	14
15 Box-Splines	14
16 Polynome im Box-Spline-Raum $S_{X,k}$	15
17 Interpolation durch Box-Splines: Lagrange-Darstellung	16

Teil I

Gewöhnliche Differentialgleichungen

1 Vorbemerkungen

Aufgabenstellung

Problem: Finde eine Funktion $y : [t_0, \infty) \rightarrow \mathbb{R}^n$ diffbar mit

$$y' = f(t, y), \quad t \geq t_0, \quad f : [t_0, \infty) \times \mathbb{R}^n \rightarrow \mathbb{R}^n, \quad \text{Anfangswert } y(t_0) = y_0 \quad (1)$$

- Lipschitzbedingung: $\exists \lambda > 0 : \|f(t, y) - f(t, x)\| \leq \lambda \cdot \|x - y\| \quad \forall x, y \in \mathbb{R}^n$
- f erfüllt die Lipschitzbedingung \Rightarrow (1) besitzt eine eindeutige Lösung (f analytisch $\Rightarrow y$ analytisch)
- Satz von Peano: Seien $a, b > 0, U \subset \mathbb{R}^n$ eine Umgebung, $f : [0, a] \times U \rightarrow \mathbb{R}^n$ stetig in $M = \{(t, x) : t \in [0, a], x \in \mathbb{R}^n, \|x - y_0\| \leq b\}$. Dann hat (1) eine Lösung y in $[0, \min(a, \frac{b}{\mu})]$, wobei $\mu = \max_{(t,x) \in M} \|f(t, x)\|$
- Satz von Picard-Lindelöf: Gilt die Lipschitzbedingung für f , so ist die Lösung aus dem Satz von Peano eindeutig.

Definitionen

- Seien für $h > 0$ $y_{j,h} \approx y(t_j)$ die Näherung an die analytische Lösung y . Eine Methode zur Lösung von (1) heißt **konvergent**, falls

$$\lim_{h \rightarrow 0} \max_{j: jh \leq t^*} \|y_{j,h} - y(t_j)\| = 0$$

für alle Lipschitz-stetigen f und für alle $t^* > 0$

- Der Fehler, der entsteht, wenn $y(t_k)$ usw. statt y_k in eine numerische Methode eingesetzt wird, heißt **lokaler Diskretisierungsfehler (LDF)**.
- Ist der LDF $O(h^{p+1})$, so ist die Methode von der **Ordnung p** .

Im Folgenden setzen wir voraus, dass f die Lipschitzbedingung erfüllt.

2 Die ϑ -Methode

$$y_{k+1} = y_k + h(\vartheta f(t_k, y_k) + (1 - \vartheta)f(t_{k+1}, y_{k+1})) \quad 0 \leq \vartheta \leq 1$$

- $\vartheta = 0$ Vorwärts-Euler (explizites Verfahren)
- $\vartheta = \frac{1}{2}$ Trapezregel (implizites Verfahren)
- $\vartheta = 1$ Rückwärts-Euler (implizites Verfahren)
- Die ϑ -Methode ist min. von Ordnung 1.

2.1 Eulersche Methode (Vorwärts-Euler)

$$y_{k+1} = y_k + h \cdot f(t_k, y_k) \quad t_k = t_0 + kh$$

- Die Eulersche Methode konvergiert.
- Die Eulersche Methode ist von Ordnung 1.

2.2 Trapezregel

$$y_{k+1} = y_k + \frac{h}{2}(f(t_k, y_k) + f(t_{k+1}, y_{k+1})) \quad k \in \mathbb{N}_0$$

- Die Trapezregel konvergiert.
- Die Trapezregel ist von Ordnung 2.

3 Allgemeine Mehrschrittverfahren

3.1 Adam-Bashforth-Methoden

$$y_{k+s} = y_{k+s-1} + h \sum_{m=0}^{s-1} f(t_{k+m}, y_{k+m}) \cdot b_m, \quad b_m = \frac{1}{h} \int_{t_{k+s-1}}^{t_{k+s}} p_m(\vartheta) d\vartheta,$$

wobei p_m die üblichen Lagrangefunktionen sind:

$$p_m(t) = \prod_{l=0, l \neq m}^{s-1} \frac{t - t_{k+l}}{t_{k+m} - t_{k+l}}$$

- Für $s = 1$ ergibt sich die Eulersche Methode (Vorwärts-Euler).
- Die Adam-Bashforth-Methoden sind von Ordnung s .

3.2 Definition und Ordnung

$$\text{Allgemeine MSV: } \sum_{m=0}^s a_m y_{k+m} = h \sum_{m=0}^s b_m f(t_{k+m}, y_{k+m}) \quad \text{o.E. } a_s = 1$$

- Für $b_s = 0$ sind die Verfahren explizit, für $b_s \neq 0$ implizit.
- $\varrho(\omega) := \sum_{m=0}^s a_m \omega^m$
- $\sigma(\omega) := \sum_{m=0}^s b_m \omega^m$
- Die Mehrschrittverfahren hat Ordnung $p \geq 1$ genau dann, wenn es ein $c \neq 0$ gibt mit:

$$\varrho(\omega) - \sigma(\omega) \log(\omega) = c \cdot (\omega - 1)^{p+1} + O(|1 - \omega|^{p+2}) \quad \omega \rightarrow 1$$

3.3 Konvergenz und die Dahlquist-Sätze

- **Wurzelbedingung:** Polynom erfüllt die Wurzelbedingung \Leftrightarrow alle Wurzeln liegen in $B_1(0) = \{z \in \mathbb{C} : |z| \leq 1\}$, alle Wurzeln auf dem Einheitskreis sind einfach
- **Dahlquist-Äquivalenzsatz:** Konvergieren die Fehler in den Startvektoren y_1, \dots, y_{s-1} des MSV gegen 0, so gilt: das MSV konvergiert \Leftrightarrow Ordnung ≥ 1 , ϱ erfüllt die Wurzelbedingung
- **Erste Dahlquist-Schranke:** konvergentes, explizites (bzw. implizites) s -MSV \Rightarrow Ordnung $\leq s$ (bzw. $2\lfloor \frac{s}{2} + 1 \rfloor$)

3.4 BDF-Formeln

“backward differentiation formula”:

$$\sum_{m=0}^s a_m y_{m+k} = h\beta f(t_{k+s}) \quad \left(\varrho(\omega) = \sum_{m=0}^s a_m \omega^m \quad \sigma(\omega) = \beta \omega^s \right)$$

- Die BDF-Formel hat Ordnung s , falls $\beta = (\sum_{m=1}^s \frac{1}{m})^{-1}$, $\varrho(\omega) = \beta \sum_{m=1}^s \frac{1}{m} \omega^{s-m} (\omega - 1)^m$
- $\varrho(\omega) = \beta \sum_{m=1}^s \frac{1}{m} \omega^{s-m} (\omega - 1)^m$ erfüllt Wurzelbedingung $\Leftrightarrow s \leq 6$

4 Runge-Kutta-Verfahren

$$y' = f(t, y), \quad y(t_0) = y_0$$

$$y(t) = y(t_0) + \int_{t_0}^t g(\tau, y(\tau)) d\tau$$

Bei Runge-Kutta-Verfahren benutzt man Quadraturformeln, um das Integral anzunähern.
Spezialfall: Einsetzen der **Gauß-Quadratur**

• Orthogonale Polynome

Sei λ ein absolutstetiges Maß über \mathbb{R} , d.h. $d(\lambda(t)) = \begin{cases} \omega(t) dt & t \in [a, b] \\ 0 & \text{sonst} \end{cases}$, ω integrierbar,

insb. $\int_a^b t^n \omega(t) dt < \infty \quad \forall n \in \mathbb{N}, \omega \geq 0$.

Definiere das **innere Produkt** $(u, v)_\lambda = \int_{\mathbb{R}} u v d\lambda = \int_a^b u v \omega dt$

Orthogonalität: $u \perp v \Leftrightarrow (u, v)_\lambda = 0$

Die **orthogonalen Polynome** $\pi_n \in \mathbb{P}_n$ sind durch die folgenden Bedingungen eindeutig bestimmt:

- $\pi_n \perp \pi_m \quad \forall m \neq n, m, n \in \mathbb{N}$
- $(\pi_n, \pi_n)_\lambda = 1 \quad n \in \mathbb{N}$

Beispiele: Jacobipolynome auf $[-1, 1]$

$$\omega(t) = (1-t)^\alpha (1+t)^\beta, \quad \alpha, \beta > -1$$

- Chebyshev-Polynome $\alpha = \beta = \frac{1}{2} \quad (T_n(x) = \cos(n \arccos x))$
- Legendre-Polynome $\alpha = \beta = 0$
- Gegenbauer-Polynome $\alpha = \beta$

- Alle m Nullstellen eines orthogonalen Polynoms liegen im Intervall (a, b) und sind einfach.
- Die Quadraturformel

$$\int_a^b f(t) d\lambda(t) = \sum_{k=1}^n \lambda_k f(\tau_k) + R_n(f)$$

erfüllt $R_n(p) = 0 \quad \forall p \in \mathbb{P}_{n-1+m} \Leftrightarrow$ [interpoliert für $Grad < n$] $\wedge [q(t) = \prod_{j=1}^n (t - t_j), \int_a^b q(t) p(t) d\lambda(t) = 0 \quad \forall p \in \mathbb{P}_{m-1}]$

4.1 explizite Runge-Kutta-Verfahren

$$k_1 = y_n, \quad k_s = y_n + h \sum_{i=1}^{s-1} a_{si} f(t_n + c_j h, k_i), \quad s > 1$$

$$y_{n+1} = y_n + h \sum_{j=1}^m b_j f(t_n + c_j h, k_j)$$

Beispiele mit $c_1 = 0$:

- **Verfahren von Runge:** $m = 2, c_2 = \frac{1}{2}, a_{21} = \frac{1}{2}, b_1 = 0, b_2 = 1$ mit Ordnung 2
 $[y_{n+1} = y_n + h f(t_n + \frac{h}{2}, y_n + \frac{h}{2} f(t_n, y_n))]$
- **klassisches Runge-Kutta-Verfahren:** $m = 3, c_2 = \frac{1}{2}, c_3 = 1, a_{21} = \frac{1}{2}, a_{31} = -1, a_{32} = 2, b_1 = b_2 = \frac{1}{6}, b_3 = \frac{2}{3}$ mit Ordnung 3
- **Nystrom-Verfahren** $m = 3, c_2 = c_3 = \frac{2}{3}, a_{21} = \frac{2}{3}, a_{31} = 0, a_{32} = \frac{2}{3}, b_1 = \frac{1}{4}, b_2 = b_3 = \frac{3}{8}$ mit Ordnung 3

4.2 implizite Runge-Kutta-Verfahren (IRK)

$$k_j = y_n + h \cdot \sum_{i=1}^{\nu} a_{ji} \cdot f(t_n + c_j h, k_i)$$

$$y_{n+1} = y_n + h \cdot \sum_{j=1}^{\nu} b_j \cdot f(t_n + c_j h, k_j)$$

- Verfahren konsistent $\Rightarrow \sum_{i=1}^{\nu} a_{ji} = c_j \forall j$
- **Kollokationsverfahren:** Suche Polynom $u \in \Pi_{\nu} : u(t_n) = y_n, u'(t_n + c_j h) = f(t_n + c_j h, u(t_n + c_j h)) \forall 1 \leq j \leq \nu$
 $y_{n+1} := u(t_{n+1}) \approx y(t_{n+1})$
 - c_1, \dots, c_{ν} gegeben, $q(t) = \prod_{j=1}^{\nu} (t - c_j)$, $q_i(t) = \frac{q(t)}{t - c_i}$
 $a_{ji} = \int_0^{c_j} \frac{q_i(t)}{q_i(c_i)} dt$, $b_j = \int_0^1 \frac{q_i(\tau)}{q_i(c_j)} d\tau$, $j = 1, \dots, \nu$
 \Rightarrow Die Kollokationsmethode ist identisch zu dem IRK mit a_{ji}, b_j, c_j
 - **Butcher-Tableau:** $\begin{array}{c|c} c & A \\ \hline & b^T \end{array}$
 - c_i pw. verschieden, $\int_0^1 q(\tau) \tau^j d\tau = 0 \forall j < m \leq \nu \Rightarrow$ Kollokationsmethode hat Ordnung $\nu + m$
- $T > 0, j \leq \frac{T}{h}, y_0, \dots, y_j$ errechnete Näherungen von $y' = f(t, y)$, $y(0) = y_0$, f Lipschitzstetig \Rightarrow Verfahren konvergiert mit Fehler $|y_j - y(jh)| = O(h^{m+\nu})$

5 Steife Differentialgleichungen

5.1 Definitionen

- **DG steif** \Leftrightarrow Lösung durch eine Standardmethode erfordert eine wesentliche Einschränkung der Schrittweite h
- **Steifigkeitsverhältnis** := Verhältnis vom größten zum kleinsten Eigenwert der Jacobimatrix von f
- Betrachte $y' = \lambda y$, $y(0) = 1$, h Schrittweite $(y(t) = e^{\lambda t} \xrightarrow{t \rightarrow \infty} 0 \quad \forall \lambda : \Re(\lambda) < 0)$
 - **Lineares Stabilitätsgebiet eines Verfahrens** $\mathcal{D} := \{\lambda h \in \mathbb{C} : y_k \xrightarrow{k \rightarrow \infty} 0\}$
- Numerisches Verfahren **A-stabil** $\Leftrightarrow \mathcal{D} = \mathbb{C}_- = \{z \in \mathbb{C} : \Re(z) < 0\}$

5.2 A-Stabilität von MSV

- $\eta(z, w) := \sum_{m=0}^s (a_m - z b_m) w^m$ char. Polynom der Differenzgleichung $\sum_{m=0}^s (a_m - h \lambda b_m) y_{m+n} = 0$, $w_1, \dots, w_{q(z)}$ die versch. Nullstellen mit Vielfachheit $k_1(z), \dots, k_{q(z)}(z)$: MSV A-stabil $\Leftrightarrow |w_i(z)| < 1 \quad \forall z \in \mathbb{C}_-$
- MSV A-stabil $\Leftrightarrow b_s > 0, |w_j(it)| \leq 1 \quad \forall t \in \mathbb{R}, j = 1, \dots, q(it)$

- **Satz von Cohn und Schur:** Beide Nullstellen der quadratischen Gleichung $\alpha w^2 + \beta w + \gamma = 0$ sind in $\overline{B_1(0)}$ $\Leftrightarrow |\alpha|^2 \geq |\gamma|^2, (|\alpha|^2 - |\gamma|^2)^2 \geq |\alpha \bar{\beta} - \beta \bar{\gamma}|^2, (\alpha = \gamma \neq 0 \Rightarrow |\beta| \leq 2\alpha)$
- Die Trapezregel und das 2-Schritt BDF-Verfahren sind A-stabil, das Eulerverfahren ist nicht A-stabil.
- **2. Dahlquist-Schranke:** Die maximale Ordnung eines A-stabilen MSV ist 2.

5.3 Dissipativität

$y' = f(t, y)$, $y(0) = y_0$ **dissipativ** $\Leftrightarrow [x' = f(t, x), x(0) = x_0 \Rightarrow \|x(t) - y(t)\| \xrightarrow{t \rightarrow \infty} 0]$

- $(x - y)^T (f(t, x) - f(t, y)) \leq 0 \quad \forall t \geq 0 \Rightarrow$ Differentialgleichung $y' = f(t, y)$, $y(0) = y_0 \in \mathbb{R}^n$ dissipativ
- MSV A-stabil $\Leftrightarrow \|y_k - x_k\| \xrightarrow{k \rightarrow \infty} 0$

5.4 A-Stabilität von Runge-Kutta-Verfahren

$y_n = (r(h\lambda))^n y_0$, $r(h\lambda) := 1 + h\lambda b^T (I - h\lambda A)^{-1} \mathbf{1}$, (wobei $\mathbf{1}$ der Einsvektor ist)

- $|r(h\lambda)| < 1 \Rightarrow$ RK A-stabil
- $\mathcal{D} := \{z \in \mathbb{C} : |r(z)| < 1\} = \mathbb{C}_-, r(z)$ rationale Funktion \Rightarrow A-Stabilität
- Keine explizite Methode ist A-stabil.
- $r \neq \text{const.}$ und rational:
 $|r(z)| < 1 \quad \forall z \in \mathbb{C}_- \Leftrightarrow [z \text{ Pol von } r(z) \Rightarrow \Re(z) > 0], |r(it)| \leq 1 \quad \forall t \in \mathbb{R}$
- Padi-Approximation: $r(z) = e^z + O(h^{p+1})$, p Ordnung der RK-Methode

6 Schrittweitensteuerung

kleine Schrittweiten $h > 0 \Rightarrow$ gute Genauigkeit, viele Berechnungsschritte
 großen Schrittweiten $h \gg 0 \Rightarrow$ wenige Berechnungsschritte, schlechte Genauigkeit
 \Rightarrow Schrittweitensteuerung, variable Schrittweiten h_n

7 Methode nach Milne

Sei das MSV $\sum_{m=1}^s a_m y_{k+m} = h \sum_{m=0}^s b_m f(t_m, y_{k+m})$ von Ordnung p und $a_s = 1$. Wir wähles ein weitere MSV von Ordnung p zur Überwachung des ersten MSV: $\sum_{m=q}^s \tilde{a}_m x_{k+m} = h \sum_{m=q}^s \tilde{b}_m f(t_m, x_{k+m})$ mit $q \leq s - 1, \tilde{a}_s = 1$

$$y(t_{k+s}) - y_{k+s} \approx \frac{C}{C - \tilde{C}} (x_{k+s} - y_{k+s})$$

Mögliche Implementierung:

1. Festlegung von $h > 0$

2. Berechnung von y_{k+s}

3. $H := \frac{C}{C-\bar{C}} \cdot \|x_{k+s} - y_{k+s}\| \leq hs$?

(a) ja $\Rightarrow t_{k+s} \geq T$?

i. ja \Rightarrow Stop

ii. nein $\Rightarrow H \leq \frac{hs}{10}$?

A. ja $\Rightarrow h$ verdoppeln, GOTO 3.(a)ii.B

B. nein $\Rightarrow k \rightarrow k+1$, GOTO 2.

(b) nein $\Rightarrow h$ halbieren, Startwerte für neues y_{k+s} durch Interpolation neu berechnen, GOTO 2.

8 Zadunaiskys Methode

Seien y_{k-p}, \dots, y_k die Näherungen der DGL $y' = f(t, y)$, $y(t_0) = y_0$ an den Stellen t_{k-p}, \dots, t_k durch ein MSV von Ordnung p . Sei \underline{p} ein Polynom mit

$$\underline{p}(t) = \sum_{j=k-p}^k y_j L_j(t), \quad L_j(t) := \prod_{l=k-p, l \neq j} \frac{t - t_k}{t_j - t_l}$$

Betrachte nun die DGL:

$$z' = f(t, z) + \underline{p}'(t) - f(t, \underline{p}(t)), \quad z(t_k) = y_k, \quad t \geq t_k$$

Durch die Lösung dieser DGL erhält man einen Fehler

$$\|\underline{p}(t_{k+1}) - z_{k+1}\| = H \approx \|y_{k+1} - y(t_{k+1})\|$$

9 Prädiktor-Korrektor-Verfahren

Startwert mit explizitem Verfahren (Prädiktor) berechnen:

$$y_{n+k}^{(0)} + \sum_{j=0}^{k-1} \alpha_j y_{n+j} = h \sum_{j=0}^{k-1} b_j f(t_{n+j}, y_{n+j})$$

Iteration mit implizitem Verfahren (Korrektor) durchführen:

$$y_{n+k}^{(l+1)} = - \sum_{j=0}^{k-1} \alpha_j y_{n+j} + h \beta_k f(t_{n+k}, y_{n+k}^{(l)}) + h \sum_{j=0}^{k-1} \beta_j f(t_{n+j}, y_{n+j})$$

Für ausreichend kleine Schrittweiten h konvergiert die Methode.

Teil II

Lösung dünn besetzter, großer linearer Gleichungssysteme

10 Cholesky-Verfahren

10.1 Matrizen und Bäume

Man kann die Besetzungsstruktur von symmetrischen Matrizen $A = (a_{ij}) \in \mathbb{R}^{n \times n}$ durch Graphen ausdrücken:

- Ein Graph ist eine Menge von Ecken $E = \{1, \dots, n\}$ und Kanten K
- $(i, j) \in K \Leftrightarrow a_{ij} \neq 0, i < j$
- Ein Baum ist ein Graph, bei dem je zwei Ecken aus E durch eine Kante $k \in K$ verbunden sind.
- Man kann einen Baum (E, K) mit einer Wurzel ausstatten.
- Ein Baum mit Wurzel (E, K, W) kann nach Eltern und Kindern geordnet werden.
- Ein Baum heißt monoton, falls jede Ecke vor allen ihren Vorfahren numeriert ist $((i, k) \in K \Rightarrow (i, q) \notin K \forall q < k)$
- Parter: $A \in \mathbb{R}^{n \times n}$ positiv definit, symmetrisch mit monotonem Baum als Graphen \Rightarrow Cholesky-Zerlegung $A = LL^T$ mit $l_{kj} = \frac{a_{kj}}{l_{jj}}, j < k \leq n, 1 \leq j < n, l_{jj} = \sqrt{a_{jj}}$

11 Iterative Verfahren

Wir wollen das Gleichungssystem

$$Ax = b, \quad A \in \mathbb{R}^{n \times n}, \quad b \in \mathbb{R}^n$$

lösen. Dazu kann man A in zwei Matrizen zerlegen: $A = P - N$, so dass P eine "einfache" Matrix ist und dann eine Iteration durchführen:

$$Px_{k+1} = Nx_k + b \Leftrightarrow x_{k+1} = P^{-1}Nx_k + P^{-1}b \Leftrightarrow x_{k+1} = Hx_k + v$$

- $x_{k+1} = Hx_k + v$ konvergiert für jeden Startwert $x_1 \Leftrightarrow \varrho(H) < 1$
- **Householder-John:** $A \in \mathbb{R}^{n \times n}$ symmetrisch:
 A und $(P + P^T - A)$ positiv definit $\Rightarrow \varrho(H) < 1$

Sei $A = D - L - R$, D Diagonalmatrix, L linke untere, R rechte obere Dreiecksmatrix.

11.1 Jacobi und Gauss-Seidel

Jacobi-Verfahren: $P = D, \quad N = L + R, \quad \tilde{L} := D^{-1} \cdot (L + R)$

Gauss-Seidel-Verfahren: $P = D - L, \quad N = R, \quad B := (D - L)^{-1} \cdot R$

- Starkes Zeilensummenkriterium: A strikt diagonaldominant $\Rightarrow \varrho(\tilde{L}) < 1, \varrho(B) < 1$
- Stein-Rosenberg: $|a_{jj}| \neq 0 \forall j = 1, \dots, n, b_{ij} \geq 0 \forall i, j = 1, \dots, n$
 $\Rightarrow \begin{matrix} \varrho(\tilde{L}) = \varrho(B) = 0 & \vee & \varrho(\tilde{L}) < \varrho(B) < 1 \\ \vee & \varrho(\tilde{L}) = \varrho(B) = 1 & \vee & \varrho(\tilde{L}) > \varrho(B) > 1 \end{matrix}$
- A tridiagonal, $a_{jj} \neq 0 \forall j = 1, \dots, n$
 - $\lambda \in \sigma(B) \Rightarrow \lambda^2 \in \sigma(\tilde{L})$
 - $\mu \in \sigma(\tilde{L}) \setminus \{0\} \Rightarrow \sqrt{\mu} \vee -\sqrt{\mu} \in \sigma(B)$
 - $\varrho(B) < 1 \Rightarrow \varrho(\tilde{L}) < 1$

11.2 SOR-Verfahren

$$P = D - \omega L, \quad D = (1 - \omega)D + \omega R, \quad \omega \in [1, 2)$$

Ordnungsvektoren

- $K := \{(k, l) : 1 \leq k, l \leq n, a_{k,l} \neq 0\}$
- $j \in \mathbb{Z}^n$ Ordnungsvektor zu $A \in \mathbb{R}^{n \times n} : \Leftrightarrow |j_k - j_l| = 1 \forall (k, l) \in K$
- $j \in \mathbb{Z}^n$ kompatibler Ordnungsvektor zu $A \in \mathbb{R}^{n \times n}$
 $\Leftrightarrow |j_k - j_l| = \begin{cases} 1 & k \geq l + 1 \\ -1 & k \leq l - 1 \end{cases} \forall (k, l) \in K$

Konvergenz des SOR-Verfahrens

- j Ordnungsvektor zu $A \Rightarrow \exists P$ Permutationsmatrix: PAP^{-1} besitzt kompatiblen Ordnungsvektor
- A besitzt kompatiblen Ordnungsvektor $\Rightarrow g(s, t) = \det(tL + t^{-1}R - sD)$ unabhängig von $t \in \mathbb{R} \setminus \{0\} \forall s \in \mathbb{R}$
- A besitzt kompatiblen Ordnungsvektor
 - $\mu \in \sigma(B)$ mit Multiplizität $k \Rightarrow -\mu \in \sigma(B)$ mit Multiplizität k
 - $\mu \in \sigma(B), \omega \in (0, 2), \lambda : \lambda + \omega - 1 = \omega\mu\sqrt{\lambda} \Rightarrow \lambda \in \sigma(\tilde{L}_\omega)$
 - $\omega \in (0, 2), \lambda \in \sigma(\tilde{L}_\omega) \Rightarrow \exists \mu \in \sigma(B) : \lambda + \omega - 1 = \omega\mu\sqrt{\lambda}$
- A besitzt kompatiblen Ordnungsvektor $\Rightarrow \varrho(\tilde{L}) = \varrho(B)^2$
- $\varrho(\tilde{L}_\omega) < 1 \Rightarrow \omega \in (0, 2)$
- A besitzt kompatiblen Ordnungsvektor, $\sigma(B) \subset \mathbb{R} \Rightarrow$ [SOR konvergiert $\forall \omega \in (0, 2) \Leftrightarrow \varrho(B) < 1$]

- A besitzt kompatiblen Ordnungsvektor, $\varrho(B) \subset \mathbb{R}, \bar{\mu} := \varrho(B) < 1 \Rightarrow$
 - $\varrho(\tilde{L}_\omega) > \varrho(\tilde{L}_{\omega_{\text{opt}}}) \forall \omega \in (0, 2) \setminus \{\omega_{\text{opt}}\}$
 - $\omega_{\text{opt}} = 1 + \left(\frac{\bar{\mu}}{1 + \sqrt{1 - \bar{\mu}^2}}\right)^2 = \frac{2}{1 + \sqrt{1 - \bar{\mu}^2}} = \frac{2(1 - \sqrt{1 - \bar{\mu}^2})}{\bar{\mu}^2} \in (1, 2)$
 - $\varrho(\tilde{L}_{\omega_{\text{opt}}}) = \left(\frac{\bar{\mu}}{1 + \sqrt{1 - \bar{\mu}^2}}\right)^2 = \omega_{\text{opt}} - 1$

Eigenschaft A

- A besitzt **Eigenschaft A** $\Leftrightarrow \exists S_1, S_2 : S_1 \cup S_2 = \{1, \dots, n\}, S_1 \cap S_2 = \emptyset : [(k, l) \in K \Rightarrow (k, l) \in S_1 \times S_2 \vee (k, l) \in S_2 \times S_1]$
- A besitzt Eigenschaft A $\Leftrightarrow A$ besitzt Ordnungsvektor

11.3 CG-Verfahren

Suche $x \in \mathbb{R}^n : A \cdot x = b \quad A \in \mathbb{R}^{n \times n}, b \in \mathbb{R}^n, \quad A$ symm., positiv definit

Definiere $f : \mathbb{R}^n \rightarrow \mathbb{R}, x \mapsto \frac{1}{2} \cdot x^T \cdot A \cdot x - b^T \cdot x$

Dann gilt: $f(x^*) = \min_{x \in \mathbb{R}^n} f(x) \Rightarrow A \cdot x^* = b$, löse also statt des Gleichungssystems ein Minimierungsproblem

Algorithmus:

- $r^k := A \cdot x^k - b$ Residuum
- $\lambda_k := \frac{(r^k)^T \cdot d_k}{(d^k)^T \cdot A \cdot d^k}$
- 1. Startwert $x_0 \in \mathbb{R}^n, k = 0$
- 2. $x^{k+1} = x^k + \lambda_k \cdot d^k$
- 3. r^k klein genug \Rightarrow fertig
 r^k nicht klein genug $\Rightarrow k \rightarrow k + 1, \text{GOTO } 2.$

Wahl der d^k :

Konjugiertheit der Suchrichtungen: $(d^i)^T \cdot A \cdot d^j = 0 \forall i \neq j$

- $d^0 := -r^0 \quad d^0 \neq 0$ (sonst $A \cdot x^0 = b$ und Lösung gefunden)
- $d^k := -r^k + \frac{(r^k)^T \cdot A \cdot d^{k-1}}{(d^{k-1})^T \cdot A \cdot d^{k-1}} d^{k-1}$

Es gilt:

λ_{\min} kleinster, λ_{\max} größter Eigenwert (> 0) der Matrix A , so dass $\text{cond}_2(A) = \frac{\lambda_{\max}}{\lambda_{\min}}, e^k := x^* - x^k$ Fehler der Näherung an $A \cdot x^* = b, x^k$ mit CG-Verfahren berechnet
 $\Rightarrow (e^k)^T \cdot A \cdot e^k \leq 4 \cdot (e^0)^T \cdot \left(\frac{\sqrt{\lambda_{\max}} - \sqrt{\lambda_{\min}}}{\sqrt{\lambda_{\max}} + \sqrt{\lambda_{\min}}}\right)^{2k}$

Vorkonditionierung:

Angenommen A (symmetrisch und positiv definit) ist schlecht konditioniert.
Ziel: Suche B , so dass $\text{cond}(B) \ll \text{cond}(A)$

$$B := H^{-1}AH^{-1T}, \quad Bx = H^{-1}b =: c$$

Sei H regulär, dann ist B symmetrisch und positiv definit. Also kann man den CG-Algorithmus auf $Bx = c$ anwenden

Damit die Eigenwerte von B nah beieinander liegen, sollte $HH^T \approx A$ sein, etwa durch **unvollständige Cholesky-Zerlegung**: Cholesky-Faktorisierung, aber $h_{ij} = 0$, falls $a_{ij} = 0$.

Algorithmus zum vorkonditionierten CG-Verfahren:

Löse statt $Ax = b$ nun $By = c$ mit Startwert x_0

1. $d_0 = r_0 = Ax_0 - b$
2. $\lambda_k = -\frac{r_k^T d_k}{d_k^T H^{-1} A H^{-1T} d_k}$
3. $x_{k+1} = x_k + \lambda_k d_k$
4. $r_{k+1} = r_k - (H^{-1} A H^{-1T})(\lambda_k d_k)$
5. $d_{k+1} = -r_{k+1} + \frac{r_{k+1}^T H^{-1} A H^{-1T} d_k}{d_k^T H^{-1} A H^{-1T} d_k} d_k$
6. r_{k+1} klein genug?
 - (a) ja \rightarrow STOP
 - (b) nein \rightarrow GOTO 2.

Fazit:

CG ist das modernste und beste Verfahren für symmetrische und positiv definite A . Bei der Vorkonditionierung kann man sogar nichtpositive Eigenwerte entfernen und so eine positiv definite Matrix schaffen.

Teil III

Mehrdimensionale Splines

12 Wiederholung eindimensionale Splines

- $X = \{x_i\}_{i=1}^m$ Knotenmenge, $x_1 < \dots < x_m$
- $S_{X,k} = \{s : [x_0, x_m] \rightarrow \mathbb{R} : s|_{[x_i, x_{i+1}]} \in \mathbb{P}^k, s \in \mathcal{C}^{k-1}(x_0, x_m)\}$
- Basis von $S_{X,k}$: B-Splines $\{B_j\}_{j=-k}^{m-1}$, $\text{supp} B_j = [x_j, x_{j+k+1}]$

13 Tensorprodukt

Im Zweidimensionalen:

$$S_{X \times Y, k} = \{s : [x_0, x_m] \times [y_0, y_n] \rightarrow \mathbb{R} : s|_{[x_i, x_{i+1}] \times [y_j, y_{j+1}]} \in \mathbb{P}_1^k \times \mathbb{P}_1^k, s \in \mathcal{C}^{k-1}([x_0, x_m] \times [y_0, y_n])\}$$

Definition:

$X = \{x_0 < \dots < x_m\}, Y = \{y_0 < \dots < y_n\}$ Knotenmengen, $B_i \in S_{X,k}, \tilde{B}_j \in S_{Y,k}$ B-Splines:

Tensorprodukt B-Splines $B_{ij}(x, y) = B_i(x) \cdot \tilde{B}_j(y)$

Die Tensorprodukt B-Splines bilden eine Basis von $S_{X \times Y, k}$

Analog für höhere Dimensionen

Problem:

Totalgrad der Polynome ist von der Raumdimension abhängig ($k \cdot \dim$)

14 Splines von totalem Grad k

Definitionen

- k **Komponentengrad** von $p = p(x_1, \dots, x_n) = \sum_{\alpha} a_{\alpha} x_1^{\alpha_1} \cdot \dots \cdot x_n^{\alpha_n} \in \mathbb{P}_n^k$: $\Leftrightarrow k = \max_{\alpha, i=1, \dots, n} \alpha_i$
- t **Totalgrad** von $p = p(x_1, \dots, x_n) = \sum_{\alpha} a_{\alpha} x_1^{\alpha_1} \cdot \dots \cdot x_n^{\alpha_n} \in \mathbb{P}_n^k$: $\Leftrightarrow t = \max_{\alpha} \{\sum_{i=1}^n \alpha_i\}$

Sätze

- $V \subset \mathbb{R}^{k+1}$ Polyeder mit $k+2$ Ecken, $0 \neq d \in \mathbb{R}^{k+1}$, $U(\vartheta) := \{x \in \mathbb{R}^{k+1} : x^T d = \vartheta\}$:
 $U(\vartheta)$ enthält nie mehr als eine Ecke von $V \Rightarrow B(\vartheta) := \text{vol}_k(U(\vartheta) \cap V)$ B-Spline von Grad k
- $V \subset \mathbb{R}^{k+n}$ Simplex (komplexe Hülle von $n+k+1$ Punkten), $x \in \mathbb{R}^n$: $M_x := \{y \in V : x = (y_1, \dots, y_n)^T\}$, $B(x) = \text{vol}_k M_x$:
 M_x enthält nie mehr als eine Ecke von $V \Rightarrow B \in \mathbb{R}^n[x]$ Spline vom Totalgrad k

15 Box-Splines

Definitionen

- $k \geq 0$, $X = \{x_1, \dots, x_{n+k}\} \subset \mathbb{Z}^n$ **Richtungsmenge**, $\text{span} X = \mathbb{R}^n$:
 - $X = \{x_1, \dots, x_{n+k}\}$ Menge von **Richtungsvektoren**
 - $X = (x_1 \dots x_{n+k}) \in \mathbb{R}^{n+k \times n}$ Matrix
 - $X : \mathbb{R}^{n+k} \rightarrow \mathbb{R}^n, t \mapsto Xt$ Abbildung
- **Box-Spline** B_X definiert durch $\int_{\mathbb{R}^n} B_X(x) f(x) dx = \int_{[0,1]^{n+k}} f(Xt) dt$
- \hat{f} **Fouriertransformierte** von $f \in L^1$: $\Leftrightarrow \hat{f}(x) = \int_{\mathbb{R}^n} e^{-ixt} f(x) dx$
- $L_{\infty}^d(\mathbb{R}^n) := \{f : \mathbb{R}^n \rightarrow \mathbb{R} : f \text{ } d\text{-mal diff. bar, } f^{(d)} \text{ beschränkt}\}$ (insb. $f^{(d)}$ nicht unbedingt stetig)

Sätze

- $\text{span}(X) = \text{span}(X \setminus \{x_{n+k}\}) = \mathbb{R}^n \Rightarrow B_X(x) = \int_0^1 B_{X \setminus \{x_{n+k}\}}(x - tx_{n+k}) dt$
- $\text{supp}(B_X) = X[0, 1]^{n+k} = \{Xt : t \in [0, 1]^{n+k}\}$
- $\widehat{B}_X(x) = \prod_{j=1}^{n+k} \frac{1 - e^{-i(x \cdot x_j)}}{x \cdot x_j}$
- $B_X(x) > 0 \quad \forall x \in \overset{\circ}{\text{supp}}(B_X)$
- $k = 0 \Rightarrow B_X = \frac{1}{|\det(X)|} \cdot \chi_{X[0,1]^n}$
- $y \in \mathbb{R}^n \Rightarrow \exists \lambda_i : \sum_{i=1}^{n+k} \lambda_i x_i$; Ist D_y die Richtungsableitung $y^T \nabla$, so gilt:
 - $D_y B_X(x) = \sum_{i=1}^{n+k} \lambda_i (B_{X \setminus \{x_i\}}(x) - B_{X \setminus \{x_i\}}(x - x_i))$
 - $B_X(y) = \frac{1}{k} \sum_{i=1}^{n+k} \lambda_i B_{X \setminus \{x_i\}}(y) + (1 - \lambda_i) B_{X \setminus \{x_i\}}(y - x_i)$
- $d := \max\{r : \text{span}(X \setminus Z) = \mathbb{R}^n \quad \forall Z \subset X : |Z| = r\} \Rightarrow B_X \in L_\infty^{(d)}(\mathbb{R}^n) \cap \mathcal{C}^{d-1}(\mathbb{R}^n)$
- $Z \subset X$ Basis von $\mathbb{R}^n \Rightarrow \sum_{j \in Z^n} B_X(x - Zj) = \frac{1}{|\det Z|} \quad \forall x \in \mathbb{R}^n$
- $\sum_{j \in \mathbb{Z}^n} B_X(x - j) = 1$

16 Polynome im Box-Spline-Raum $S_{X,k}$

Definitionen

- \mathbb{P}_n^k Menge aller Polynome in n Variablen mit Totalgrad $\leq k$
- $Qp(x) := \sum_{j \in \mathbb{Z}^n} p(j) B_X(x - j), x \in \mathbb{R}^n$ **Quasiinterpolation (Schoenberg-Operator)**
- $N : \mathbb{P}_n \rightarrow \mathbb{P}_n$ **gradvermindernd** $\Leftrightarrow \text{Grad}(N(p)) < \text{Grad}(p) \quad \forall p \neq 0, N(0) = 0$
- $(v_j)_{j \in \mathbb{Z}^n} \mapsto (\sum_{j \in \mathbb{Z}^n} v_j w_{m-j})_{m \in \mathbb{Z}^n}$ **Faltung** $v \mapsto v * w$

Sätze

- $X^* := \{Z \subset X : \text{span}(X \setminus Z) \neq \mathbb{R}^n\}, D_Z = \prod_{z \in Z} D_z$ (Richtungsableitungen)
 $\Rightarrow \mathbb{P}_n \cap S_{X,k} = \bigcap_{Z \in X^*} \text{Kern}(D_Z)$
- $d := \max\{r : \text{span}(X \setminus Z) = \mathbb{R}^n \quad \forall Z \subset X, |Z| = r\} \Rightarrow [\mathbb{P}_n^k \subset S_X \Leftrightarrow k \leq d]$
- $f : \mathbb{R}^n \rightarrow \mathbb{R}$ gleichmässig stetig, $Q_n f(x) = \sum_{j \in \mathbb{Z}^n} f(jh) B_X(\frac{x}{h} - j), x \in \mathbb{R}^n$
 $\Rightarrow Q_n f \xrightarrow{h \rightarrow 0} f$ gleichmässig
- $(\mathbb{P}_n^k)^* = \{\mu : \mathbb{P}^k \rightarrow \mathbb{R} : \mu \text{ lineares Funktional}\}, k \leq \max\{r : \text{span}(X \setminus Z) = \mathbb{R}^n \quad \forall Z \subset X, |Z| = r\} \Rightarrow \exists \lambda \in (\mathbb{P}_n^k)^* : Q(\lambda p) = p \quad \forall p \in \mathbb{P}_n^k$
- $d := \max\{r : \text{span}(X \setminus Z) = \mathbb{R}^n \quad \forall Z \subset X, |Z| = r\}, f \in \mathcal{C}^{d+1}(\mathbb{R}^n)$ mit gl.m. beschränkten Ableitungen der Ordnung $d + 1 \Rightarrow \exists$ Erweiterung $\tilde{\lambda}$ von $\lambda \in (\mathbb{P}_n^d)^* : \|f(x) - Q_h(\tilde{\lambda} f)(x)\|_\infty = O(h^{d+1})$

17 Interpolation durch Box-Splines: Lagrange-Darstellung

Wir wollen die Funktion $f \in \mathcal{C}(\mathbb{R}^n)$ durch eine Funktion $s(x) = \sum_{j \in \mathbb{Z}^n} a_j B_X$ in allen Punkten aus \mathbb{Z}^n interpolieren. Es sind also Koeffizienten $(a_j)_{j \in \mathbb{Z}^n}$ gesucht, so dass gilt:

$$s(x) = \sum_{j \in \mathbb{Z}^n} f(j) L_j(x), \quad L_j(x) = \sum_{k \in \mathbb{Z}^n} c_{k,j} B_X(x - k) = \begin{cases} 1 & x = j \\ 0 & x \in \mathbb{Z}^n \setminus \{j\} \end{cases}$$

Da wir äquidistante Knoten betrachten, sind die L_j bis auf Verschiebung gleich.

$$L(x) = \sum_{k \in \mathbb{Z}^n} c_k B_X(x - k) = \begin{cases} 1 & x = 0 \\ 0 & x \in \mathbb{Z}^n \setminus \{0\} \end{cases} \Rightarrow L_j(x) = L(x - j)$$

Es sei das **Symbol der Lagrange-Funktion** wie folgt definiert:

$$\sigma(\vartheta) := \sum_{j \in \mathbb{Z}^n} B_X(j) e^{-i\vartheta j}$$

Die c_k sind die Fourierkoeffizienten von $\frac{1}{\sigma(\vartheta)}$:

$$c_k = \frac{1}{(2\pi)^n} \int_{[-\pi, \pi]^n} e^{i\vartheta k} \frac{1}{\sigma(\vartheta)} d\vartheta$$

Dies ist wohldefiniert, da $\sigma(\vartheta)$ keine Nullstellen besitzt.

Man kann dies etwa mit der **Poisson'schen Summationsformel** einsehen:

$$f : \mathbb{R}^n \rightarrow \mathbb{R} \text{ absolut integrierbar, } \sum_{j \in \mathbb{Z}^n} |f(j)|, \sum_{k \in \mathbb{Z}^n} |\hat{f}(\vartheta + 2\pi k)| \text{ beide konvergent } \forall \vartheta \\ \Rightarrow \sum_{j \in \mathbb{Z}^n} f(j) e^{-i\vartheta j} = \sum_{k \in \mathbb{Z}^n} \hat{f}(\vartheta + 2\pi k)$$

Da $\sigma(\vartheta) = \sum_{j \in \mathbb{Z}^n} B_X(j) e^{-i\vartheta j} = \sum_{k \in \mathbb{Z}^n} \widehat{B}_X(\vartheta + 2\pi k)$, muss man nur noch zeigen, dass gilt:

- $\widehat{B}_X \geq 0$
- $\exists \vartheta_0 : \widehat{B}_X(\vartheta_0 + 2\pi k) = 0$