1 Euklidische Approximation

Sei V ein reeller euklidischer Vektorraum.
Das Skalarprodukt in V wird mit $\langle\cdot,\cdot\rangle_V$ und die Norm mit $\|\cdot\|_V$ bezeichnet.
$V_N \subset V$ sei ein Teilraum der Dimension $N < \infty$ mit Basis $\{\phi_n\}_{n=1,\dots,N}$.

(1.1) Problemstellung: Sei $v \in V$. Bestimme $v_N \in V$ mit

$$\|v - v_N\|_V = \min_{w_N \in V_N} \|v - w_N\|_V \,.$$

(1.2) Die Matrix

$$A = \Big(\langle\phi_n, \phi_k\rangle_V\Big)_{n,k=1,\dots,N} \in \mathbb{R}^{N\times N}$$

ist symmetrisch und positiv definit.

(1.3) Problem (1.1) ist eindeutig lösbar. Es gilt

$$v_N = \sum_{n=1}^{N} x_n \phi_n \,,$$

wobei $x^* \in \mathbb{R}^N$ die eindeutige Lösung des linearen Gleichungssystems

$$Ax = b$$

mit $b = \Big(\langle v, \phi_k\rangle_V\Big)_{k=1,\dots,N} \in \mathbb{R}^N$ ist.

2 Direkte Lösungsverfahren für lineare Gleichungen

(2.1) Sei $L \in \mathbb{R}^{N\times N}$ eine normierte untere Dreiecksmatrix und $b \in \mathbb{R}^N$, d.h. $\operatorname{diag} L = I_N$ und $L[1:n, n+1] = 0_n$ für $n = 1, ..., N-1$. Dann ist L regulär und das lineare Gleichungssystem $Ly = b$ ist mit $O(N^2)$ Operationen lösbar.
Entsprechend ist für eine reguläre obere Dreiecksmatrix $R \in \mathbb{R}^{N\times N}$ (d.h. $R[n,n] \neq 0$ für alle n und $R[n+1, 1:n] = 0_n^T$ für $n < N$) das LGS $Rx = y$ in $O(N^2)$ Operationen lösbar.

(2.2) Die normierten unteren Dreiecksmatrizen bilden eine Gruppe.
Die regulären oberen Dreiecksmatrizen bilden eine Gruppe.

(2.4) Wenn eine Matrix $A \in \mathbb{R}^{N\times N}$ eine LR-Zerlegung $A = LR$ mit einer normierten untere Dreiecksmatrix L und einer regulären obere Dreiecksmatrix R besitzt, dann ist A regulär und das LGS $Ax = b$ ist mit $O(N^2)$ Operationen lösbar.

(2.5) Eine Matrix $A \in \mathbb{R}^{N\times N}$ besitzt genau dann eine LR-Zerlegung von A, wenn alle Hauptuntermatrizen $A[1:n, 1:n]$ regulär sind.
Die LR-Zerlegung ist eindeutig und lässt sich mit $O(N^3)$ Operationen berechnen.

(2.6) Eine Matrix $A \in \mathbb{R}^{N\times N}$ heißt *strikt diagonal-dominant*, falls $|a_{nn}| > \sum_{k=1}^{n-1} |a_{nk}| + \sum_{k=n+1}^{N} |a_{nk}|$.

(2.7) Wenn A strikt diagonal dominant ist, dann existiert eine LR-Zerlegung.

(2.8) Sei $A \in \mathbb{R}^{N\times N}$ symmetrisch und positiv definit. Dann existiert genau eine Cholesky-Zerlegung $A = LL^T$ mit einer regulären unteren Dreiecksmatrix L.

2 Direkte Lösungsverfahren für lineare Gleichungen

(2.9) Sei $\pi \in S_N$ eine Permutation. Dann heißt $P_\pi = \big(e_{\pi^{-1}(1)}|\ldots|e_{\pi^{-1}(N)}\big) \in \mathbb{R}^{N\times N}$ Permutationsmatrix zu π. Es gilt $(P_\pi A)[n,k] = A[\pi(n),k]$ und $(AP_\pi)[n,k] = A[n,\pi^{-1}(k)]$.

(2.10) Die Permutationsmatrizen in $\mathbb{R}^{N\times N}$ bilden eine Gruppe. Es gilt $P_\sigma P_\pi = P_{\pi\circ\sigma}$ und $P_{\pi^{-1}} = P_\pi^T$.

(2.11) Sei $A \in \mathbb{R}^{N\times N}$ regulär. Dann existiert eine Permutationsmatrix P, so dass PA eine LR-Zerlegung $PA = LR$ besitzt und für die Einträge $|L[m,n]| \le 1$ gilt.
Sie lässt sich mit $O(N^3)$ Operationen berechnen.
Die Lösung von $Ax = b$ berechnet sich aus $Ly = Pb$ und $Rx = y$.

Sei $|\cdot|$ eine Vektornorm, und sei $\|\cdot\|$ eine zugeordete Matrixnorm, d. h.,

$$|Ax| \le \|A\|\,|x|, \qquad x \in \mathbb{R}^N,\ A \in \mathbb{R}^{M\times N}.$$

(2.12) Sei $A \in \mathbb{R}^{N\times N}$ regulär, und sei $\Delta A \in \mathbb{R}^{N\times N}$ so klein, dass $\|\Delta A\| < \|A^{-1}\|^{-1}$ gilt. Dann ist die Matrix $\tilde{A} = A + \Delta A$ regulär.
Sei $b \in \mathbb{R}^N$, $b \ne 0_N$, $\Delta b \in \mathbb{R}^N$ klein und $\tilde{b} = b + \Delta b$.
Dann gilt für die Lösungen $x \in \mathbb{R}^N$ von $Ax = b$ und $\tilde{x} \in \mathbb{R}^N$ von $\tilde{A}\tilde{x} = \tilde{b}$

$$\frac{|\Delta x|}{|x|} \le \frac{\kappa(A)}{1 - \kappa(A)\frac{\|\Delta A\|}{\|A\|}} \left(\frac{|\Delta b|}{|b|} + \frac{\|\Delta A\|}{\|A\|}\right).$$

Dabei ist $\Delta x = \tilde{x} - x$, $\frac{|\Delta x|}{|x|}$ der *relative Fehler*, und $\kappa(A) = \|A\|\,\|A^{-1}\|$ die *Kondition* von A.

2 Arithmetische Grundlagen

(2.13) a) Eine Gleitkommazahlen zur Basis $B \in \{2,3,...\}$ der Mantissenlänge M und Exponentenlänge E ist die Menge

$$\mathrm{FL} = \Big\{ \pm B^e \sum_{m=1}^{M} a_m B^{-m} \colon e = e^- + \sum_{k=0}^{E-1} c_k B^k,\ a_m, c_k \in \{0,1,...,B-1\} \Big\}$$

b) Eine Gleitkommaarithmetik wird durch eine Abbildung $\mathrm{fl}\colon \mathbb{R} \longrightarrow \mathrm{FL}$ mit $\mathrm{fl}(x) = x$ für $x \in \mathrm{FL}$ definiert. Sei bestimmt die Rundung: $x \oplus y = \mathrm{fl}(x+y)$, $x \odot y = \mathrm{fl}(x \cdot y)$, etc. Die zugehörige Maschinengenauigkeit ist $\mathrm{eps} := \max \Big\{ \frac{|x - \mathrm{fl}(x)|}{|x|} \colon x \notin \mathrm{FL} \Big\}$.

(2.14) Sei $f\colon \mathbb{R}^N \longrightarrow \mathbb{R}^K$ eine differenzierbare Funktion und $x \in \mathbb{R}^N$. Dann heißt

a) $\kappa_{\mathrm{abs}}^{kn} = \Big| \frac{\partial}{\partial x_n} f_k(x) \Big|$ *absolute Konditionszahl.*

b) $\kappa_{\mathrm{rel}}^{kn} = \Big| \frac{\partial}{\partial x_n} f_k(x) \Big| \frac{|x_n|}{|f_k(x)|}$ *relative Konditionszahl.*

2 Kondition und Stabilität

(2.15) a) Ein Problem heißt *sachgemäß gestellt*, wenn es eindeutig lösbar ist und die Lösung stetig von den Daten abhängt.

b) Die *Kondition* eines Problems ist eine Maß dafür, wie stark die Abhängigkeit der Lösung von den Daten ist.

c) Die *Stabilität* eines numerischen Algorithmus ist eine Maß dafür, wie stark die Daten-Abhängigkeit der numerischen Lösung im Vergeich zu der exakten Lösung ist.

(2.16) Wir verwenden für $x \in \mathbb{R}^N$ und $A \in \mathbb{R}^{M\times N}$

$$|x|_1 = \sum_{n=1}^{N} |x[n]| \qquad |x|_2 = \sqrt{x^T x} \qquad |x|_\infty = \max_{n=1,\dots,N} |x[n]|$$

und die zugeordnete Operatornorm $\|A\|_p = \sup_{x\neq 0_N} \frac{|Ax|_p}{|x|_p}$, d.h.

$$\|A\|_1 = \max_{n=1,\dots,N} \sum_{m=1}^{M} |A[m,n]|\,, \quad \|A\|_2 = \sqrt{\rho(A^T A)}\,, \quad \|A\|_\infty = \max_{m=1,\dots,M} \sum_{n=1}^{N} |A[m,n]|$$

mit *Spekralradius* $\rho(A) = \max\{|\lambda|\colon \lambda \in \sigma(A)\}$ und Spektrum $\sigma(A)$.

3 Ausgleichsrechnung

Sei $Q \in \mathbb{R}^{N\times N}$ orthogonal, d.h. $Q^T Q = I_N$. Dann ist $\kappa_2(Q) = 1$.

(3.1) Zu $v \in \mathbb{R}^N$ und $k \neq n$ mit $v[k]^2 + v[n]^2 > 0$ existiert eine *Givens-Rotation* $G \in \mathbb{R}^{N\times N}$ mit

$$\begin{pmatrix} G[k,k] & G[k,n] \\ G[n,k] & G[n,n] \end{pmatrix} = \begin{pmatrix} c & s \\ -s & c \end{pmatrix}, \qquad c^2 + s^2 = 1\,,$$

und $G[j][j] = 1$ für $j \neq k,n$ und $G[i][j] = 0$ sonst, so dass $e_n^T G v = 0$ gilt:
Für $|v[n]| > |v[k]|$ setze $\tau = \frac{v[k]}{v[n]}$, $s = \frac{1}{\sqrt{1+\tau^2}}$, $c = s\tau$, sonst setze $\tau = \frac{v[n]}{v[k]}$, $c = \frac{1}{\sqrt{1+\tau^2}}$, $s = c\tau$.
Es gilt $G = c(e_k e_k^T + e_n e_n^T) + s(e_k e_n^T - e_n e_k^T) + \sum_{j\neq k,n} e_j e_j^T$.

(2.10) Zu $v \in \mathbb{R}^N$, $v \neq 0_N$, existiert eine *Householder-Spiegelung* $H = I_N - \dfrac{2}{w^T w} w w^T \in \mathbb{R}^{N\times N}$ mit $w \in \mathbb{R}^N$, $w[1] = 1$, so dass $Hv = \sigma e_1$ mit $\sigma \in \mathbb{R}$ und $Hw = -w$ gilt:
Falls $v[1] > 0$, setze $\sigma = -|v|_2$, sonst setze $\sigma = |v|_2$. Dann definierte $w = \frac{1}{v[1]-\sigma}(v - \sigma e_1)$.

(3.3) Zu $A \in \mathbb{R}^{K\times N}$ existiert eine QR-Zerlegung $A = QR$ mit einer orthogonalen Matrix $Q \in \mathbb{R}^{K\times K}$ und eine oberen Dreiecksmatrix $R \in \mathbb{R}^{K\times N}$, d.h. $QQ^T = I_K$ und $R[k+1:K,k] = 0_{K-k}$ für $k = 1,\dots,K$.

(3.4) Sei $A \in \mathbb{R}^{K\times N}$ und $b \in \mathbb{R}^K$. Dann gilt:

$$x \in \mathbb{R}^N \text{ minimiert } |Ax-b|_2 \qquad \Longleftrightarrow \qquad A^T A x = A^T b\,.$$

Berechnung der Householder-Vektoren

```
function [v,beta] = householder(y)
N = length(y);
s = y(2:N)' * y(2:N);
if N == 1
  s = 0;
end;
v = [1;y(2:N)];
if s == 0
  beta = 0;
else
  mu = sqrt(y(1)^2 + s);
  if y(1) <= 0
    v(1) = y(1) - mu;
  else
    v(1) = -s/(y(1) + mu);
  end;
  beta = 2*v(1)^2/(s + v(1)^2);
  v = v / v(1);
end;
return;
```

QR-Zerlegung

```
[M,N] = size(A);
for m = 1:min(N,M-1)
  [v,beta] = householder(A(m:M,m));
  if beta ~= 0
    w = beta * v' * A(m:M,m:N);
    A(m:M,m:N) = A(m:M,m:N) - v * w;
    A(m+1:M,m) = v(2:M-m+1);
  end;
end;

for m = 1:min(N,M-1)
  v = [1;A(m+1:M,m)];
  beta = 2 / (v' * v);
  if beta ~= 2
     b(m:M) = b(m:M) - beta*(v'*b(m:M)) * v;
  end;
end;
for n=min(N,M):-1:1
  x(n) = (b(n) - A(n,n+1:N) * x(n+1:N)) / A(n,n);
end;
```

3 Ausgleichsrechnung

(3.5) Zu $A \in \mathbb{R}^{K\times N}$ mit $R = \mathrm{rang}(A)$ existiert eine Singulärwertzerlegung

$$A = V\Sigma U^T$$

mit Matizen $V = (v_1|\cdots|v_R) \in \mathbb{R}^{K\times R}$, $U = (u_1,|\cdots|u_R) \in \mathbb{R}^{N\times R}$, $\Sigma = \mathrm{diag}(\sigma_1,...,\sigma_R) \in \mathbb{R}^{R\times R}$ mit $V^TV = U^TU = I_R$ und den Singulärwerten $\sigma_1,...,\sigma_R > 0$.

Es gilt $A = \sum_{r=1}^{R} \sigma_r u_r v_r^T$ und $Ax = \sum_{r=1}^{R} \sigma_r (v_r^T x) u_r$.

(2.22) $A^+ = U\Sigma^{-1}V^T$ ist die *Pseudo-Inverse*. Es gilt $A^+ = \sum_{r=1}^{R} \sigma_r^{-1} v_r u_r^T$ und $Ax = \sum_{r=1}^{R} \sigma_r^{-1}(u_r^T x) v_r$.

(2.22) $x = A^+b$ löst die Normalengleichung $A^TAx = A^Tb$.

(2.23) Sei $A \in \mathbb{R}^{K\times N}$ und $b \in \mathbb{R}^M$. Dann gilt für die Tikhonov-Regularisierung mit $\alpha > 0$:

$$x \in \mathbb{R}^N \text{ minimiert } |Ax - b|_2^2 + \alpha|x|_2^2 \qquad \Longleftrightarrow \qquad (A^TA + \alpha I_N)x = A^Tb.$$

(2.24) Es gilt $\lim\limits_{\alpha\longrightarrow 0}(A^TA + \alpha I_N)^{-1}A^Tb = A^+b$.

Diskrepanzprinzip: Sei $b \in \mathrm{Bild}(A)$, $x = A^+b$ und b^δ eine Störung mit $|b - b^\delta| < \delta < |b|_2$.
Dann existiert ein $\alpha = \alpha(\delta) > 0$ mit $|Ax^\alpha - b^\delta|_2 = \delta$ für $x^\alpha = (A^TA + \alpha I_N)^{-1}A^Tb^\delta$.
Es gilt $\alpha(\delta) \longrightarrow 0$ für $\delta \longrightarrow 0$.

4 Eigenwertberechnung

(4.3) $H \in \mathbb{R}^{N\times N}$ heißt *Hessenberg*-Matrix, wenn $H[n+2:N,n] = 0_{N-n-1}$ für $n = 1,...,N-2$.

(4.4) Sei $A \in \mathbb{R}^{N\times N}$. Dann existiert eine orthogonale Matrix $Q \in \mathbb{R}^{N\times N}$, so dass $H = QAQ^T$ eine Hessenberg-Matrix ist. Die Berechnung benötigt $O(N^3)$ Operationen.

Wenn A symmetrisch ist, dann ist H eine Tridiagonalmatrix.

(4.8) Sei $A \in \mathbb{R}^{N\times N}$ symmetrisch, tridiagonal, und irreduzibel, d.h. $A[n-1,n] = A[n,n-1] \neq 0$ und $A[n+2:N,n] = A[n,n+2:N]^T = 0_{N-n-1}$.
Die charakteristischen Polynome $P_n(t) = \det\big(A[1:n,1:n] - tI_n\big)$ der Hauptuntermatrizen lassen sich durch eine Dreitermrekursion berechnen:
Setze $P_0 \equiv 1$. Dann gilt $P_1(t) = A[1,1] - t$ und

$$P_n(t) = (A[n,n]-t)P_{n-1}(t) - A[n-1,n]^2 P_{n-2}(t)\,.$$

Sie bilden eine *Sturmsche Kette*: Für die Nullstellen $\lambda_1^n \le \lambda_2^n \le \cdots \le \lambda_n^n$ von P_n gilt

$$\lambda_{k-1}^{n-1} < \lambda_k^n < \lambda_k^{n-1}\,, \qquad k = 1,...,n$$

(mit $\lambda_0^n = -\|A\|_\infty$ und $\lambda_{n+1}^n = \|A\|_\infty$), und es gilt für $t \in (-\|A\|_\infty, \|A\|_\infty)$

$$\lambda_k^n < t \le \lambda_{k+1}^n$$

mit $k = W_n(t)$ und $W_n(t) = \#\big\{j \in \{1,...,n\} \colon P_j(t)P_{j-1}(t) < 0 \text{ oder } P_j(t) = 0\big\}$.

4 Eigenwertberechnung

Sei $A \in \mathbb{R}^{N\times N}$ symmetrisch mit Eigenwerten $\lambda_1,\dots,\lambda_N$ und ONB aus Eigenvektoren $v_1,\dots,v_N$. Dann gilt

$$A = \sum_n \lambda_n v_n (v_n)^T \qquad \text{und} \qquad Ax = \sum_n \lambda_n (v_n^T x) v_n\,.$$

(4.9) Der *Rayleigh-Quotient* ist

$$r(A,x) = \frac{x^T A x}{x^T x}\,, \qquad x \in \mathbb{R}^N,\ x \neq 0_N\,.$$

(4.10) Sei $|\lambda_1| = \rho(A)$ und $|\lambda_n| < |\lambda_1|$ für $n = 2,\dots,N$. Dann gilt für alle $w \in \mathbb{R}^N$ mit $w^T v_1 > 0$

$$\lim_{k\longrightarrow\infty} r(A, A^k w) = \lambda_1\,, \qquad \lim_{k\longrightarrow\infty} \frac{1}{|A^k w|_2} A^k w = v_1\,.$$

(4.11) Sei $|w|_2 = 1$ und $\mu = r(A,w)$. Dann gilt

$$\min_{n=1,\dots,N} |\lambda_n - \mu| \le |Aw - \mu w|_2\,.$$

Eine konvergente Folge (d^k) in $\mathbb{R}$ mit Grenzwert d^* konvergiert

a) *linear*, wenn $c \in (0,1)$ und $k_0 > 0$ existieren mit
$$|d^{k+1} - d^*| \le c\,|d^k - d^*| \qquad \text{für } k \ge k_0$$

b) *superlinear*, wenn zu jedem $\varepsilon > 0$ ein $k_0 > 0$ existiert mit
$$|d^{k+1} - d^*| \le \varepsilon\,|d^k - d^*| \qquad \text{für } k \ge k_0$$

c) *von der Ordnung $p > 1$*, wenn wenn $C > 0$ existiert mit
$$|d^{k+1} - d^*| \le C\,|d^k - d^*|^p\,.$$

(4.12) Inverse Iteration mit variablem shift

S0) Wähle $z_0 \in \mathbb{R}^N$, $z_0 \neq 0_N$, $\varepsilon \geq 0$. Setze $k = 0$.
S1) Setze $w_k = \frac{1}{|z_k|_2} z_k$, $\mu_k = r(A, w_k)$.
S2) Falls $|Aw_k - \mu_k w_k|_2 \leq \varepsilon$ STOP.
S3) Berechne $z_{k+1} = (A - \mu_k I_N)^{-1} w_k$.
S4) Setze $k := k+1$, gehe zu S1).

Wenn der Startvektor z_0 hinreichend nahe bei einem Eigenvektor v_m mit isoliertem Eigenwert λ_m liegt, konvergiert die Iteration kubisch (d.h. von der Ordnung $p = 3$).

(4.13) QR-Iteration mit shift ($A \in \mathbb{R}^{N \times N}$ symmetrisch)

S0) Berechne $A_0 = QAQ^T$ tridiagonal (Hessenberg-Transformation).
Wähle $\varepsilon \geq 0$. Setze $k = 0$.

S1) Falls $|A_k[n+1, n]| \leq \varepsilon$ für ein n:
getrennte Eigenwertberechnung für $A_k[1:n, 1:n]$ und $A_k[n+1:N, n+1:N]$.

S2) Berechne $d_k = \frac{1}{2}(A_k[N-1, N-1] - A_k[N, N])$ und

$$s_k = A_k[N, N] + d_k - \operatorname{sgn}(d_k)\sqrt{d_k^2 + A_k[N-1, N]^2}\,.$$

S3) Berechne QR-Zerlegung $Q_k R_k = A_k - s_k I_N$ und setze $A_{k+1} = R_k Q_k + s_k I_N$.
S4) Setze $k := k+1$, gehe zu S1).

Es gilt $A_{k+1} = Q_k^T A_k Q_k$. Falls der shift $\mu_k = A_k[N, N]$ gewählt wird, entspricht die QR-Iteration der Inversen Iteration mit variablem shift und Startvektor $z_0 = e_N$.

4 Eigenwertberechnung

(4.14) Gershgorin
Zu $A \in \mathbb{R}^{N\times N}$ sind die Gershgorin-Kreise durch

$$K_n = \{\lambda \in \mathbb{C} \colon |\lambda - A[n,n]| \le \sum_{k\neq n} |A[n,k]|\}, \qquad n = 1,\dots,N$$

defininiert. Dann gilt

$$\sigma(A) \subset \bigcup_{n=1}^{N} K_n .$$

(4.15) Sei $A \in \mathbb{R}^{N\times N}$ symmetrisch mit Eigenwerten $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_N$.

$$\begin{aligned} \lambda_n &= \max_{\dim S = n} \min_{0_N \neq x \in S} r(A,x), \\ \lambda_{N+1-n} &= \min_{\dim S = n} \max_{0_N \neq x \in S} r(A,x). \end{aligned}$$

5 Iterative Lösungsverfahren für lineare Gleichungen

(5.1) Sei $A, B \in \mathbb{R}^{N\times N}$ mit $\rho(I_N - BA) < 1$.
Dann ist A invertierbar, und es gilt für alle $b \in \mathbb{R}^N$ und alle Startvektoren $x^0 \in \mathbb{R}^N$ konvergiert die Iteration

$$x^{k+1} = x^k + B(b - Ax^k)\,, \qquad k = 0, 1, 2, \ldots$$

gegen $\lim\limits_{k\longrightarrow\infty} x^k = A^{-1}b$.

(5.2) Sei $K \in \mathbb{R}^{N\times N}$ und $\varepsilon > 0$
Dann existiert eine Vektor-Norm $|\cdot|$ und eine zugeordnete Matrix-Norm $\|\cdot\|$, so dass $\|K\| \le \rho(K) + \varepsilon$ gilt.

Anwendung: Es gilt $|x - x^k| \le \|I_N - BA\|^k |x - x^0|$ (lineare Konvergenz).

(5.3) Konvergenz des Gauß-Seidel-Verfahrens
Sei $A = L + D + R \in \mathbb{R}^{N\times N}$ symmetrisch positiv definit und sei $B = (L + D)^{-1}$.
Dann gilt bezüglich der Energienorm $|x|_A = \sqrt{x^T Ax}$ und $\|K\|_A = \sup\limits_{x\neq 0_N} \frac{|Kx|_A}{|x|_A}$

$$\|I_N - BA\|_A < 1\,.$$

Anwendung der Neumannschen Reihe ergibt dann $A^{-1} = \sum_{k=0}^{\infty}(I_N - BA)^k B$.

5 Iterative Lösungsverfahren für lineare Gleichungen

(5.4) Eine Matrix $A \in \mathbb{R}^{N\times N}$ heißt stark diagonal-dominant, wenn

$$\sum_{\substack{k=1\\k\neq n}}^{N} |A[n,k]| \leq |A[n,n]|\,, \qquad n = 1,...,N\,,$$

und wenn ein $j \in \{1,\dots,N\}$ existiert $\sum_{\substack{k=1\\k\neq n}}^{N} |A[j,k]| < |A[j,j]|$.

(5.5) Eine Matrix $A \in \mathbb{R}^{N\times N}$ sei irreduzibel. Dann existiert zu jedem Paar $j \neq n$ eine Folge $j = j_0, j_1, j_2, \dots, j_R = n$ mit $A[j_1, j_0] \neq 0, ..., A[j_R, j_{R-1}] \neq 0$.

(5.6) Eine Matrix $A \in \mathbb{R}^{N\times N}$ sei stark diagonal-dominant und irreduzibel. Dann gilt:

a) A ist regulär und das Jacobi-Verfahren $x^{k+1} = x^k + \mathrm{diag}(A)^{-1}(b - Ax^k)$ konvergiert.

b) Sei $A[n,n] > 0$ für alle n. Dann ist A positiv definit.

c) Sei $A[n,n] > 0$ und $A[n,k] \leq 0$ für $n \neq k$. Dann ist $A^{-1}[n,k] \geq 0$ für alle n,k.

5 Iterative Lösungsverfahren: Krylov-Verfahren

(5.7) Zu $C \in \mathbb{R}^{N\times N}$ und $d \in \mathbb{R}^N$ ist k-te *Krylov-Raum*

$$\mathcal{K}_k(C,d) = \operatorname{span}\{d, Cd, ..C^{k-1}d\} = \{P(C)d\colon P \in \mathbb{P}_{k-1}\}.$$

(5.8) Zu einer regulären Matrix $A \in \mathbb{R}^{N\times N}$, einer rechten Seite $b \in \mathbb{R}^N$ und einem Startwert $x^0 \in \mathbb{R}^N$ sei $x \in \mathbb{R}^N$ die Lösung von $Ax = b$ und $r^0 = b - Ax^0$.
Sei $B \in \mathbb{R}^{N\times N}$ eine reguläre Matrix (*Vorkonditionierer*).
Wenn $\dim \mathcal{K}_k(AB, r^0) < k$ für ein k gilt, dann ist $x \in x^0 + \mathcal{K}_{k-1}(BA, Br^0)$.

Sei $\langle\cdot,\cdot\rangle$ ein Skalarprodukt in $\mathbb{R}^N$.

Gram-Schmidt-Verfahren zur Berechnung einer Orthonormalbasis $v^1, ..., v^k$ von

$$\mathcal{K}_k(BA, Br^0) = \operatorname{span}\{Br^0, BABr^0, ..., (BA)^{k-1}Br^0\} = \{V_k y\colon y \in \mathbb{R}^k\}, \quad V_k = (v^1|....|v^k).$$

S0) Wähle $x^0 \in \mathbb{R}^N$, setze $r^0 = b - Ax^0$, $z^1 = Br^0$, $h_{10} = |z^1|_V$ und $v^1 = \frac{1}{h_{10}} z^1$.

S1) Für $k = 1, 2, 3, ...$ berechne $w^k = BAv^k$,

$$\begin{aligned} z^{k+1} &= w^k - \sum_{j=1}^k h_{jk} v^j \text{ mit } h_{jk} = \langle v^j, w^k\rangle_V \\ v^{k+1} &= \frac{1}{h_{k+1,k}} z^{k+1} \text{ mit } h_{k+1,k} = |z^{k+1}|_V \end{aligned}$$

Dann gilt $BAv^k = \sum_{j=1}^{k+1} h_{jk} v^j$, also $BAV_k = V_{k+1} H_k$ mit $H_k = (h_{jm}) \in \mathbb{R}^{k+1,k}$.

5 Iterative Lösungsverfahren: GMRES-Verfahren

S0) Wähle $x^0 \in \mathbb{R}^N$, $\varepsilon > 0$.
Berechne $r^0 = b - Ax^0$, $z^1 = Br^0$, $h_{10} = |z^1|_2$ und $v^1 = \frac{1}{h_{10}} z^1$. Setze $k = 1$.

S1) Berechne $w^k = BAv^k$ und

$$\begin{aligned} z^{k+1} &= w^k - \sum_{j=1}^{k} h_{jk} v^j \text{ mit } h_{jk} = (v^j)^T w^k \\ v^{k+1} &= \frac{1}{h_{k+1,k}} z^{k+1} \text{ mit } h_{k+1,k} = |z^{k+1}|_2 \end{aligned}$$

S2) Berechne $y^k \in \mathbb{R}^k$ mit $\rho_k = |H_k y_k - h_{10} e^1|_2 = \min!$
Dabei ist $H_k = (h_{jm})_{j=1,\dots,k+1,\ m=1,\dots,k} \in \mathbb{R}^{k+1,k}$.

S3) Wenn $\rho_k < \varepsilon$, setze $x^k = x^0 + \sum_{j=1}^{k} y_j^k v^j$ STOP.

S4) Setze $k := k+1$ und gehe zu S1).

(5.4) Es gilt $\rho_k = \min\limits_{z \in x^0 + \operatorname{span}\{v^1,\dots,v^k\}} |B(b - Az)|_2$.

Das GMRES-Verfahren ist wohldefiniert, und wenn es abbricht, gilt $|x - x^k|_2 \le \|(BA)^{-1}\|_2 \varepsilon$.

(5.5) Für $C \ge \alpha > 0$ gelte $z^T BAz \ge \alpha z^T z$ und $\|BA\|_2 \le C$.

Dann gilt für das GMRES-Verfahren $|x^k - x|_2 \le \kappa_2(BA) \left(1 - \frac{\alpha^2}{C^2}\right)^{k/2} |x^0 - x|_2$.

5 Iterative Lösungsverfahren: CG-Verfahren

S0) Wähle $x^0 \in \mathbb{R}^N$, $\varepsilon > 0$.
Berechne $r^0 = b - Ax^0$, $y^0 = Br^0$, $\rho_0 = (y^0)^T r^0$ und $d^1 = y^0$. Setze $k = 0$.

S1) Falls $\rho_k \leq \varepsilon$ STOP

S2) Setze $k := k+1$ und berechne

$$\begin{aligned} u^k &= Ad^k \\ \alpha_k &= \frac{\rho_{k-1}}{(u^k)^T d^k} \\ x^k &= x^{k-1} + \alpha_k d^k \\ r^k &= r^{k-1} - \alpha_k u^k \\ y^k &= Br^k \\ \rho_k &= (y^k)^T r^k \\ d^{k+1} &= y^k + \frac{\rho_k}{\rho_{k-1}} d^k \end{aligned}$$

Gehe zu S1).

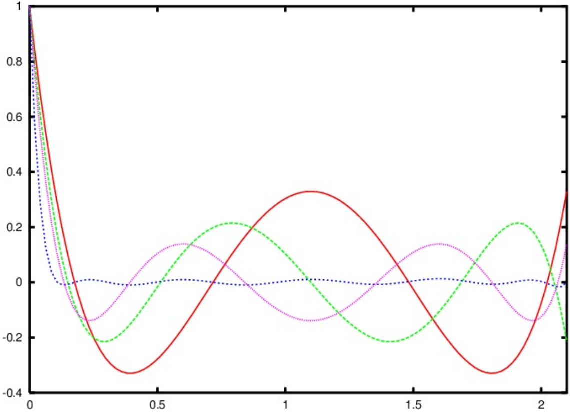
(5.6)

$$\begin{aligned} |x^k - x|_A &= \min_{z \in x^0 + \operatorname{span}\{d^1,\dots,d^k\}} |z - x|_A \\ &\leq \min_{P \in \mathbb{P}_k,\ P(0)=1} \max_{\lambda \in \sigma(BA)} |P(\lambda)| |x^0 - x|_A \leq 2 \left(\frac{\sqrt{\kappa(BA)} - 1}{\sqrt{\kappa(BA)} + 1} \right)^k |x^0 - x|_A . \end{aligned}$$

5 Transformierte Chebychev-Polynome

(5.6)

$$\min_{P\in\mathbb{P}_k,\ P(0)=1}\ \max_{t\in[a,b]} |P(t)| \le 2\left(\frac{\sqrt{\frac{b}{a}}-1}{\sqrt{\frac{b}{a}}+1}\right)^k$$



P_4, P_5, P_6, P_{12} mit $P_k(0) = 1$ zu $[a,b] = [0.1, 2.1]$

6 Lösungsverfahren für nichtlineare Gleichungen

(6.1) Sei $D \subset \mathbb{R}^N$ konvex und $F\colon D \longrightarrow \mathbb{R}^N$ ein stetig differenzierbares Vektorfeld.
Sei $x^* \in D$ eine Nullstelle von F, und sei $B \in \mathbb{R}^{N\times N}$.
Wenn $\rho\big(I_N - BJ_F(x^*)\big) < 1$ gilt, dann existiert ein $\delta > 0$, so dass für alle $x^0 \in B(x^*,\delta)$ die Fixpunktiteration $x^{k+1} = \Phi(x^k)$ mit $\Phi(x) = x - BF(x)$ linear gegen x^* konvergiert.

(6.2) Banachscher Fixpunktsatz
Sei $D \subset \mathbb{R}^N$ konvex und $\phi\colon D \longrightarrow D$ kontrahierend mit $q \in (0.1)$, d.h.

$$|\phi(y) - \phi(z)| \leq q\,|y - z| \quad \text{für} \quad y,z \in D\,.$$

Dann existiert genau ein Fixpunkt $x^* \in D$ von ϕ, d.h. $\phi(x^*) = x^*$, und es gelten für die Fixpunktiteration $x^{k+1} = \Phi(x^k)$ die Abschätzungen

$$|x^k - x^*| \leq \frac{q^k}{1-q}\,|x^0 - x^1| \qquad \text{und} \qquad |x^k - x^*| \leq \frac{q}{1-q}\,|x^k - x^{k-1}|\,.$$

(6.3) Seien $\alpha,\beta,\gamma > 0$ mit $2\alpha\gamma < \beta^2$ und $P(t) = \alpha - \beta t + \frac{\gamma}{2}t^2$.
Dann konvergiert für $t_0 = 0$ das Newton-Verfahren

$$t_{k+1} = t_k - P'(t_k)^{-1}P(t_k)$$

quadratisch gegen die kleinste Nullstelle t^* von P, und $\{t_k\}$ ist monoton steigend mit

$$t_0 = 0 < t_1 < \cdots < t_k < t_{k+1} = t_k + \frac{\gamma}{2}\,\frac{(t_k - t_{k-1})^2}{\beta - \gamma t_k} \leq t^* = \frac{2\alpha}{\beta + \sqrt{\beta^2 - 2\alpha\gamma}}\,.$$

6 Lösungsverfahren für nichtlineare Gleichungen

(6.3) Sei $D \subset \mathbb{R}^N$ offen und $F\colon D \longrightarrow \mathbb{R}^N$ ein stetig differenzierbares Vektorfeld. Sei $x^0 \in D$ mit

a) $|F(x^0)| \leq \alpha$

b) $|J_F(x^0)y| \geq \beta\,|y|$ für $y \in \mathbb{R}^N$

c) $B(x^0, 2\alpha/\beta) = \{z \in \mathbb{R}^N \colon |z - x^0| \leq 2\alpha/\beta\} \subset D$

d) $\|J_F(y) - J_F(z)\| \leq \gamma|y - z|$ für $y, z \in B(x^0, 2\alpha/\beta)$

e) $2\alpha\gamma < \beta^2$.

Dann ist das Newton-Verfahren $x^{k+1} = x^k - J_F(x^k)^{-1}F(x^k)$ wohldefiniert und konvergiert quadratisch gegen $x^* \in D$ mit

$$|x^* - x^0| \leq \frac{2\alpha}{\beta + \sqrt{\beta^2 - 2\alpha\gamma}}\,.$$

Gedämpftes Newton-Verfahren

S0) Wähle $x^0 \in D$, $\varepsilon > 0$, $\theta \in (0,1)$. Setze $k = 0$.

S1) Falls $|F(x^k)| \leq \varepsilon$ STOP

S2) Löse $J_F(x^k)d^k = -F(x^k)$.

S3) Bestimme $t_k \in \{1, \theta, \theta^2, ..., \theta^r\}$ mit $|F(x^k + t_k d^k)|$ minimal.

S4) Setze $x^{k+1} = x^k + t_k d^k$, $k := k + 1$ und gehe zu S1).

7 Orthogonalpolynome

Zu $-\infty \leq a < b \leq \infty$ und einem Gewicht $W \in C(a,b)$, $W > 0$ definiere $V = C(a,b)$ als Euklidischen Vektorraum mit dem Skalarprodukt

$$(u,v)_V = \int_a^b u(t)v(t)W(t)\,dt\,.$$

Aufgabe: Zu $f \in C(a,b)$ bestimme $P \in \mathbb{P}_N$ mit

$$\|P - f\|_V = \min_{Q \in \mathbb{P}_N} \|Q - f\|_V\,.$$

Sei $Q_0,\ldots,Q_N \in \mathbb{P}_N$ eine Orthogonalbasis. Dann gilt $P(t) = \sum_{k=0}^{N} (f, Q_k)_V\; Q_k(t)$.

(7.1) Eine orthogonale Basis Q_n, $n = 0,1,2,...$ berechnet sich mit $Q_{-1} \equiv 0$, $Q_0 \equiv \frac{1}{\rho_0}$ und

$$\rho_{n+1}Q_{n+1}(t) = (t - \beta_n)\; Q_n(t) - \rho_n\; Q_{n-1}(t) \qquad n = 0,1,2,...$$

mit $\rho_0 = \|1\|_V$, $\rho_n = (Q_n, tQ_{n-1})_V > 0$, und $\beta_n = (Q_n, tQ_n)_V$. Es gilt $\operatorname{grad} Q_n = n$.

(7.2) Das Orthogonalpolynom Q_n besitzt n verschiedene Nullstellen in (a,b).

Die Orthogonalpolynome $Q_0, Q_1, Q_2, ...$ bilden eine Sturmsche Kette.

7 Orthogonalpolynome – Chebychev-Polynome

Wir betrachten den Fall $(a,b) = (-1,1)$ und $W(t) = \dfrac{1}{\sqrt{1-t^2}}$.

Definiere $T_0 \equiv 1$, $T_1(t) = t$ und $T_{n+1} \in \mathbb{P}_{n+1}$ mit $T_{n+1}(t) = 2tT_n(t) - T_{n-1}(t)$.
Für $t \in [-1,1]$ gilt

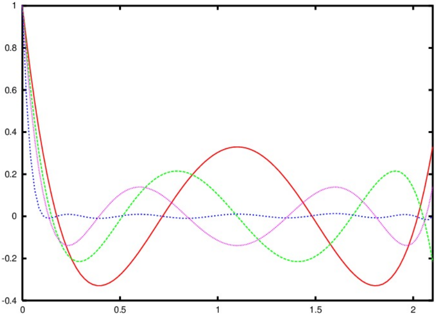
$$T_n(t) = \cos(n \arccos t)\,,$$

$(T_n, T_k)_V = 0$ für $n \neq k$, und für $|t| \geq 1$ gilt $T_n(t) = \dfrac{1}{2}\Big(t + \sqrt{t^2-1}\Big)^n + \dfrac{1}{2}\Big(t - \sqrt{t^2-1}\Big)^n$.

(7.3) Sei $\xi \notin [-1,1]$ und $P(t) = T_n(t)/T_n(\xi)$.
Dann gilt $\max\limits_{t\in[-1,1]} |P(t)| \leq \max\limits_{t\in[-1,1]} |Q(t)|$ für alle $Q \in \mathbb{P}_n$ mit $Q(\xi) = 1$.

(7.4) Sei $0 < \alpha < \beta$ und $\kappa = \beta/\alpha$.
Dann existiert ein $P \in \mathbb{P}_n$ mit $P(0) = 1$

und $\max\limits_{x\in[\alpha,\beta]} |P(x)| \leq 2\left(\dfrac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^n$.



Transformierte Chebychev-Polynome
P_4, P_5, P_6, P_{12} mit $P_n(0) = 1$
zu $[\alpha,\beta] = [0.1, 2.1]$

7 Orthogonalpolynome

(7.5) Approximationssatz von Weierstraß
Sei $[a,b]$ ein kompaktes Intervall. Dann gilt:
Jede stetige Funktion $f \in \mathrm{C}([a,b])$ lässt sich gleichmäßig durch Polynome approximieren.

Konstruktion: O.E. sei $f \in C(\mathbb{R})$ mit $f(t) = 0$ für $|t| > 0.5$ und $[a,b] = [-0.25, 0.25]$. Dann wähle

$$P_{2n}(t) = \frac{1}{r_n} \int_{-1}^{1} (1-t^2)^n f(s-t)\,\mathrm{d}s\,, \qquad r_n = \int_{-1}^{1} (1-t^2)^n\,\mathrm{d}t\,.$$

Für $|t| \le 0.25$ gilt $P_{2n}(t) = \frac{1}{r_n} \sum\limits_{j,k=0}^{n} a_{nkj} t^k \int\limits_{-1}^{1} s^j f(s)\,\mathrm{d}t \in \mathbb{P}_{2n}$ mit $(1-(s-t)^2)^n = \sum\limits_{j,k=0}^{2n} a_{nkj} t^k s^j$.

Nun sei $(v,w)_0 = \int\limits_{-1}^{1} v(t) w(t)\,\mathrm{d}t$ das Skalarprodukt in $\mathrm{L}_2(-1,1)$, seien $Q_0, Q_1, Q_2, \ldots$ die Orthogonalpolynonme, und sei $p_n(f) = \sum\limits_{k=0}^{n} (f, Q_k)_0 Q_k$ die Orthogonalprojektion auf $\mathbb{P}_n$.

(7.6) Für jede stetig Funktion $f \in \mathrm{C}[-1,1]$ gilt $\lim\limits_{n \longrightarrow \infty} \|p_n(f) - f\|_0 = 0$.

(7.7) Sei $f \in \mathrm{C}^k[-1,1]$. Dann existiert $C > 0$ mit $\|p_n(f) - f\|_0 \le C n^{-k} \left\| \left(\frac{\mathrm{d}}{\mathrm{d}t}\right)^k f \right\|_0$.

8 Interpolation

(8.1) *Lagrange-Interpolation*
Zu Stützstellen $t_0 < t_1 < \cdots < t_N$ und Werten $f_0, f_1, \ldots, f_N \in \mathbb{R}$ bestimme $P \in \mathbb{P}_N$ mit $P(t_n) = f_n$.

(8.2) Die Interpolationsaufgabe (8.1) ist eindeutig lösbar.

(8.3) *Hermite-Interpolation*
Zu $t_0 \le t_1 \le \cdots \le t_N$ definiere $d_n = \max\{n-k\colon\ t_k = \cdots = t_n\}$ und $d = \max d_n$.
Zu $f \in C^d(\mathbb{R})$ bestimme $P \in \mathbb{P}_N$ mit

$$\left(\frac{d}{dt}\right)^{d_n} P(t_n) = \left(\frac{d}{dt}\right)^{d_n} f(t_n) \quad \text{für} \quad n = 0, \ldots, N.$$

(8.4) Die Interpolationsaufgabe (8.3) ist eindeutig lösbar.

(8.5) Der eindeutig bestimmte höchste Koeffizient $b_N \equiv \frac{1}{N!}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)^N P(t)$ der Lösung der Interpolationsaufgabe (8.1) wird mit $b_N = f[t_0, \ldots, t_N]$ bezeichnet.

(8.6) Zu $t_0 \le t_1 \le \cdots \le t_N$ definiere die Newton-Basis von $\mathbb{P}_N$ durch $\omega_0 \equiv 1$ und

$$\omega_{k+1}(t) = \omega_k(t)(t - t_k), \qquad k = 0, \ldots, N-1.$$

a) $P(t) = \sum_{k=0}^{N} f[t_0, \ldots, t_k]\omega_k(t)$ löst die Interpolationsaufgabe (8.3).

b) Für $f \in \mathrm{C}(\mathbb{R})$ gilt $f(t) = P(t) + f[t_0, \ldots, t_N, t]\ \omega_{N+1}(t)$.

8 Interpolation

(8.7) Für $t_k < t_n$ gilt $f[t_k,\dots,t_n] = \dfrac{f[t_{k+1},\dots,t_n] - f[t_k,\dots,t_{n-1}]}{t_n - t_k}$.

Sei $f \in \mathrm{C}^{n-k}(\mathbb{R})$. Für $t_k = t_{k+1} = \dots = t_n$ gilt $f[t_k,\dots,t_n] = \dfrac{1}{(n-k)!}\left(\dfrac{d}{dt}\right)^{n-k} f(t_k)$.

Sei $P_{kn}(t) = \sum\limits_{j=k}^{n} f[t_k,\dots,t_j]\omega_{kj}(t)$ mit $\omega_{kk} \equiv 1$ und mit $\omega_{k,j+1}(t) = \omega_{kj}(t)(t-t_j)$ die Interpolation an $t_k \le \dots \le t_n$ zu $f_k,\dots,f_n$. Für $t_k < t_n$ gilt $P_{kn}(t) = \frac{t-t_k}{t_n-t_k} P_{k+1,n}(t) + \frac{t_n-t}{t_n-t_k} P_{k,n-1}(t)$.

(8.9) Für den Interpolationsfehler gilt

$$f(t) - P_N(t) = \frac{1}{(N+1)!}\left(\frac{d}{dt}\right)^{N+1} f(\tau)\ \omega_{N+1}(t)$$

mit einer Zwischenstelle $\tau = \tau(t) \in I := [\min\{t,t_0\}\ \max\{t,t_N\}]$.

(8.11) Es gilt $f[t_0,\dots,t_N] = \displaystyle\int_{\Sigma^N} \left(\frac{d}{dt}\right)^N f\Big(\sum_{n=0}^{N} s_n t_n\Big)\, ds$

$$\text{mit}\quad \Sigma^N = \left\{ \begin{pmatrix} s_0 \\ \vdots \\ s_n \end{pmatrix} \in \mathbb{R}^{N+1} :\ \sum_{n=0}^{N} s_n = 1, s_n \ge 0 \right\} \qquad \text{(Standard-Simplex)}$$

(8.12) $f[t_0,\dots,t_N] = \frac{1}{N!}\left(\frac{d}{dt}\right)^N f(\tau)$ mit $\tau \in [\min t_n, \max t_n]$.

9 Splines

(9.1) Zu einer Zerlegung $\Delta\colon\ a = t_0 < t_1 < \cdots < t_N = b$ von $[a,b]$ definiere den Spline-Raum von Grad k

$$\mathscr{S}_k(\Delta) = \left\{ S \in C^{k-1}[a,b]\colon\ S_n := S|_{[t_{n-1},t_n]} \in \mathbb{P}_k, \qquad n = 1,\dots,N \right\}.$$

$S \in \mathscr{S}_k(\Delta)$ heißt interpolierender Spline zu $f \in C^0[a,b]$ wenn $S(t_n) = f(t_n)$.

(9.2) Es gilt $\dim \mathscr{S}_k(\Delta) = k + N$.

Wir betrachten interpolierende Splines von Grad $k = 2r+1$ mit einer der Randbedingungen

(I) Natürliche Randbedingung

$$\left(\frac{d}{dt}\right)^m S(a) = 0 \quad \text{und} \quad \left(\frac{d}{dt}\right)^m S(b) = 0 \quad \text{für} \quad r+1 \le m \le 2r$$

(II) Hermite-Randbedingung zu $f \in C^r[a,b]$

$$\left(\frac{d}{dt}\right)^m S(a) = \left(\frac{d}{dt}\right)^m f(a) \quad \text{und} \quad \left(\frac{d}{dt}\right)^m S(b) = \left(\frac{d}{dt}\right)^m f(b) \ \text{ für } \ 1 \le m \le r$$

(III) periodische Randbedingungen

$$\left(\frac{d}{dt}\right)^m S(a) = \left(\frac{d}{dt}\right)^m S(b) \quad \text{für} \quad 1 \le m \le 2r$$

9 Splines

Zu $g,h \in C^m[a,b]$ definiere

$$(g,h)_m = \int_a^b \Big(\frac{d}{dt}\Big)^m g(t) \Big(\frac{d}{dt}\Big)^m h(t)\, dt\,, \qquad |g|_m = \sqrt{(g,g)_m}\,.$$

(9.3) Zu $a \le t_0 < t_1 < \cdots < t_N \le b$ und $f_0,\ldots,f_N$ definiere $M = \big\{ g \in C^2[a,b] \colon g(t_n) = f_n \big\}$.
Dann existiert genau ein $S \in M$ mit $|S|_2 \le |g|_2$ für alle $g \in M$.
Es gilt $S''(t) = 0$ für $t \in [a,t_0] \cup [t_N,b]$ und $S|_{[t_{n-1},t_n]} \in \mathbb{P}_3$.

(9.4) Sei $S \in \mathscr{S}_{2r+1}(\Delta)$ ein interpolierender Spline zu f und sei zusätzlich eine der Randbedingungen (I), (II) oder (III) erfüllt.
Sei $g \in C^{r+1}[a,b]$ mit $g(t_n) = f(t_n)$ für $n = 0,\ldots,N$, und zusätzlich gelte
im Fall (II)

$$\Big(\frac{d}{dt}\Big)^m g(a) = \Big(\frac{d}{dt}\Big)^m f(a) \quad \text{und} \quad \Big(\frac{d}{dt}\Big)^m g(b) = \Big(\frac{d}{dt}\Big)^m f(b) \;\; \text{für} \;\; 1 \le m \le r\,,$$

im Fall (III)

$$\Big(\frac{d}{dt}\Big)^m g(a) = \Big(\frac{d}{dt}\Big)^m g(b) \quad \text{für} \quad 1 \le m \le r\,.$$

Dann gilt

$$(S-g,S)_{r+1} = 0 \quad \text{und} \quad |S|_{r+1} \le |g|_{r+1}\,.$$

(9.5) Die Spline-Interpolation $S \in S_{2r+1}(\Delta)$ zu f mit einer der Randbedingungen (I), (II), (III) ist eindeutig lösbar.

(9.6) Die *Momente* $M_n = S''(t_n)$ von $S \in \mathscr{S}_3(\Delta)$ zu f sind eindeutig durch

$$\mu_n M_{n-1} + M_n + \lambda_n M_{n+1} = 3f[t_{n-1}, t_n, t_{n+1}]$$

und $\mu_n = \dfrac{h_n}{2(h_n + h_{n+1})}$, $\quad \lambda_n = \dfrac{h_{n+1}}{2(h_n + h_{n+1})}$ mit $h_n = t_n - t_{n-1}$ bestimmt und es gilt

$$\begin{aligned} S_n(t) \quad = \quad & \frac{M_n(t-t_{n-1})^3 + M_{n-1}(t_n - t)^3}{6h_n} + \frac{f(t_n) + f(t_{n-1})}{2} - \frac{h_n^2}{12}(M_n + M_{n-1}) \\ & + \Big(\frac{f(t_n) - f(t_{n-1})}{h_n} - \frac{h_n}{6}(M_n - M_{n-1}) \Big) \Big(t - \frac{t_n + t_{n-1}}{2} \Big), \qquad t \in [t_{n-1}, t_n]. \end{aligned}$$

(9.7) Es gilt für $m = 0, \dots, r$ und $h = \max\limits_{i=n,\dots,N} h_i$, $h_n = t_n - t_{n-1}$

$$|S - f|_m \le 2^{-(r+1-m)/2} \frac{(r+1)}{m!} h^{r+1-m} |f|_{r+1}$$

(9.8) Es gilt für $f \in C^2[a,b]$

$$\|S''\|_\infty \le 3\,\|f''\|_\infty.$$

(9.9) Es gilt für $f \in C^4[a,b]$

$$\|S - f\|_\infty \le h^4\, \Big\| \Big(\frac{d}{dt}\Big)^4 f \Big\|_\infty.$$

10 Numerische Integration – Newton-Cotes-Formeln

(10.1) Eine Quadraturformel $I_\Xi : C[a,b] \longrightarrow \mathbb{R}$ mit $I_\Xi(f) = \sum_{\xi\in\Xi} \omega_\xi f(\xi)$ zu Stützstellen $\Xi \subset [a,b]$ und Gewichten ω_ξ ist eine Linearform zur Approximation des Integrals $I(f) = \int_a^b f(t)dt$.
Sie heißt *exakt von der Ordnung p*, wenn der Fehler $I(P) = I_\Xi(P)$ für Polynome $P \in \mathbb{P}_p$.

(10.2) Sei $|\Xi| = N+1$. Wenn I_Ξ exakt von der Ordnung N ist, dann gilt

$$\omega_\xi = \int_a^b L_\xi(t)dt \text{ mit } L_\xi(t) = \prod_{\eta\in\Xi\setminus\{\xi\}} \tfrac{t-\eta}{\xi-\eta} \in \mathbb{P}_N.$$

(10.3) Die Quadraturformeln von der Ordnung N zu äquidistanten Stützstellen $\xi_n = a+nh$, $h = \frac{b-a}{N}$ heißen *Newton-Cotes-Quadraturformeln*. Es gilt:

$$\omega_n = \omega_{N-n} = \frac{b-a}{N} \, \frac{(-1)^{N-n}}{n!(N-n)!} \int_0^N \prod_{j=0,\ j\neq n}^{N} (t-j)dt.$$

(10.4) Die Newton-Cotes-Formeln sind für gerade N exakt von der Ordnung $N+1$.

(10.5) Seien $g,h \in C[a,b]$ mit $g(t) \geq 0$ für $t \in [a,b]$. Dann gilt:

$$\int_a^b g(t)h(t)dt = h(\tau)\int_a^b g(t)dt \text{ für eine Zwischenstelle } \tau \in (a,b).$$

(10.6) Sei $f \in C^4[a,b]$. Dann gilt für die Simpsonregel:

$$I_2(f) - I(f) = \left(\frac{b-a}{2}\right)^5 \frac{1}{90} f''''(\tau) \text{ für ein } \tau \in (a,b).$$

10 Numerische Integration – Gauß-Quadatur

(10.7) Die maximale Ordnung einer Quadratur mit N Stützstellen ist $2N-1$.

(10.8) Seien $\xi_1,\dots,\xi_N$ die Nullstellen des Orthogonalpolynoms Q_N in (a,b) bzgl. des Gewichts W.
Dann ist die zugehörige *Gauß-Quadratur* $G_N(f) = \sum_{n=1}^{N} \omega_n f(\xi_n)$ exakt für $P \in \mathbb{P}_{2N-1}$.
Die Gewichte der Gauß-Quadratur sind positiv.

(10.9) $$\int_a^b f(t)W(t)\,dt - G_N(f) = \frac{1}{(2N)!}\Big(\frac{d}{dt}\Big)^{2N} f(\tau)\ (R_N,R_N)_W \ \text{ mit } \ R_N(t) = \prod_{n=1}^{N}(t-\xi_n).$$

Eine Folge $\{I_{\Xi^N}\}_{N\in\mathbb{N}}$ von Quadraturen mit Stützstellen $\Xi^N \subset [a,b]$ und Ge- wichten ω_ξ^N heißt konvergent, wenn $\lim\limits_{N\to\infty} |I(f) - I_{\Xi^N}(f)| = 0$ für alle $f \in C[a,b]$.

(10.10) Sei $\{I_{\Xi^N}\}$ für alle Polynome P konvergent (d.h. $\lim_{N\to\infty} I_{\Xi^N}(P) = I(P)$), und seien die Gewichte uniform beschränkt durch $\sum\limits_{\xi\in\Xi^N} |\omega_\xi^N| \le C$ für alle N.
Dann ist $\{I_{\Xi^N}\}$ konvergent für alle $f \in C[a,b]$.

(10.11) Wenn alle Gewichte positiv sind und I_{Ξ^N} von der Ordnung N exakt ist, ist $\{I_{\Xi^N}\}$ konvergent.

10 Numerische Integration

(10.12) Sei $I_h - I = Ch^p + O(h^q)$ mit $q > p$. Dann gilt für die Extrapolation $\hat{I}_h = I_h + \dfrac{1}{2^p - 1}(I_h - I_{2h})$

a) $|I - \hat{I}_h| = O(h^q)$
b) $|I - I_h| = \frac{1}{2^p-1}(I_h - I_{2h}) + O(h^q)$

(10.13) Sei $g \in C^{2m+2}[0,1]$. Es existieren Konstanten b_j und eine Zwischenstelle $\xi \in (0,1)$ mit

$$\int_0^1 g(t)\,dt = \frac{1}{2}g(0) + \frac{1}{2}g(1) + \sum_{j=1}^m b_j\Big(\Big(\frac{d}{dt}\Big)^{2j-1}g(0) - \Big(\frac{d}{dt}\Big)^{2j-1}g(1)\Big) + b_{m+1}\Big(\frac{d}{dt}\Big)^{2m+2}g(\xi).$$

(10.14) Für $f \in C^{2m+2}[a,b]$ und $h = \dfrac{b-a}{n}$ gilt

$$T_h(f) - I(f) = \sum_{j=1}^m c_j h^{2j} + O(h^{2m+2}) \text{ mit } c_j = b_j\Big(\Big(\frac{d}{dt}\Big)^{2j-1}f(a) - \Big(\frac{d}{dt}\Big)^{2j-1}f(b)\Big).$$

(10.15) Zu $T_{k0} = T_{h_k}(f)$ mit $h_k = 2^{-k}h$ für $k = 0,1,\dots,m$ definiere rekursiv

$$T_{ki} = T_{k,i-1} + \frac{1}{4^i - 1}(T_{k,i-1} - T_{k-1,i-1}), \qquad i = 1,\dots,m \text{ und } k = i,\dots,m.$$

Dann gilt: $|T_{mm} - I| \le Ch^{2m+2}\Big\|\Big(\frac{d}{dt}\Big)^{2m+2}f\Big\|_\infty$ mit $C > 0$ unabhängig von f.

(10.16) a) $|I - T_{kk}| = O(h^{2k+2})$
b) $|I - T_{k,k-1}| = |T_{kk} - T_{k,k-1}| + O(h^{2k+2})$.

10 Numerische Integration – Romberg-Quadratur

```
using namespace std;

#include <iostream>
#include <cmath>
#include <ext/numeric>
using __gnu_cxx::power;

const int K = 6;
const int N = power(2,K);

double Romberg (double F(double), double a, double b, double Eps) {
    double f[N+1];
    double h = (b-a) / N;
    for (int i=0; i<=N; i+=2) f[i] = F(a+i*h);
    return Romberg(F,a,b,f,Eps);
}

double F (double x) {
    return 4.0 / (x*x + 1);
}

int main () { cout << "Q = " << Romberg(F,0,1,1e-10) << endl; }
```

10 Numerische Integration – Romberg-Quadratur

```
double Romberg (double F(double), double a, double b,
                double* f, double Eps) {
    double T[K+1];
    double h = (b - a) * 0.5;
    T[0] = h * (f[0] + f[N]);
    for (int i=1, m=N; i<=K; ++i, m/=2, h /= 2) {
        double s = 0.0;
        if (i<K) for (int j=m/2; j<N; j+=m) s += f[j];
        else     for (int j=1;   j<N; j+=2) s += (f[j] = F(a+j*h));
        T[i] = 0.5 * T[i-1] + h * s;
        for (int j=1, q=4; j<=i; ++j, q *= 4) {
            double e = (T[i-j+1] - T[i-j]) / (q - 1);
            T[i-j] = T[i-j+1] + e;
            if (abs(e) < Eps) return T[i-j];
        }
    }
    double f2[2*N+1];
    for (int j=0; j<=N; ++j) f2[2*j] = f[j];
    Eps *= 0.5;
    return Romberg(F,a,(a+b)*0.5,f2,Eps)
           + Romberg(F,(a+b)*0.5,b,f2+N,Eps);
}
```

11 Trigonometrische Interpolation

(11.1) Die 2π-periodischen komplexen trigonometrischen Polynome $\mathbb{T}_n$ sind von der Form

$$T(t) = \sum_{k=-n}^{n} \alpha_k \exp(\mathrm{i}kt)\,.$$

(11.2) Die Funktionen $Q_k(t) = \frac{1}{\sqrt{2\pi}} \exp(\mathrm{i}kt)$ bilden eine Orthogonalbasis von $\mathbb{T}_n$ bzgl.

$$(f,g)_0 = \int_0^{2\pi} f(t)\overline{g(t)}\,dt\,.$$

(11.3) Sei $f \in L_2(0,\pi)$. Dann ist $p_n(f)(t) = \sum_k (f, Q_k)_0 Q_k$ Orthogonalprojektion von f mit

$$\|f - p_n(f)\|_0 \le \|f - T\|_0\,, \qquad T \in \mathbb{T}_n\,.$$

(11.4) Sei $N \in \mathbb{N}$, $\omega = \omega_N = \exp\left(\mathrm{i}\frac{2\pi}{N}\right)$ die N-te Einheitswurzel und

$$F_N = \left(\omega_N^{nk}\right)_{n,k=0,\cdots,N-1} = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \omega & \omega^2 & \cdots & \omega^{N-1} \\ 1 & \omega^2 & \omega^4 & \cdots & \omega^{2(N-1)} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \omega^{N-1} & \omega^{2(N-1)} & \cdots & \omega^{(N-1)^2} \end{pmatrix} \in \mathbb{C}^{N\times N}$$

die *Fouriermatrix*. Dann gilt $F^{-1} = \frac{1}{N}\overline{F}$, d.h $\sqrt{\frac{1}{N}}F$ ist unitär.

(11.7) Sei $\hat{\alpha}_k = \frac{1}{N}\sum_{j=0}^{N-1} f(t_j)\exp(-\mathrm{i}kt_j)$ mit $t_j = \frac{2\pi j}{N}$. Dann gilt $f(t_j) = \sum_{k=0}^{N-1} \hat{\alpha}_k \exp(\mathrm{i}kt_j)$.

(11.6) Sei $N = 2M \in \mathbb{N}$ gerade und $y = F_N x \in \mathbb{C}^N$. Dann gilt

$$
\begin{aligned}
y_{2k} &= \sum_{n=0}^{M-1} x_n^+ \omega_M^{kn} \quad \text{mit} \quad x_n^+ = x_n + x_{M+n},\\
y_{2k+1} &= \sum_{n=0}^{M-1} x_n^- \omega_M^{kn} \quad \text{mit} \quad x_n^- = (x_n - x_{M+n})\omega_N^n.
\end{aligned}
$$

FFT-Algorithmus für die Berechnung von $y = F_N x$ für $N = 2^K$

S1) Wenn $N = 1$ und $y = x$: STOP

S2) Definiere

$$
\begin{aligned}
x^+ &= x\left[0 : \frac{N}{2} - 1\right] + x\left[\frac{N}{2} : N - 1\right] \in \mathbb{C}^{\frac{N}{2}}\\
x^- &= \operatorname{diag}(\omega_N^n)\left(x\left[0 : \frac{N}{2} - 1\right] - x\left[\frac{N}{2} : N - 1\right]\right) \in \mathbb{C}^{\frac{N}{2}}
\end{aligned}
$$

S3) Berechne $y^+ = F_{\frac{N}{2}} x^+,\ y^- = F_{\frac{N}{2}} x^-$

S4) Setze $y = (y^+[0], y^-[0], y^+[1], y^-[1], \cdots) \in \mathbb{C}^N$. STOP

(11.7) Der FFT-Algorithmus benötigt $N \log(N)$ Operationen.

11 Trigonometrische Interpolation

Zu $s \geq 0$ definiere $\|f\|_s = \left(\sum_{k\in\mathbb{Z}}(1+|k|^{2s})|\alpha_k|^2\right)^{1/2}$ mit $\alpha_k = \dfrac{1}{2\pi}\displaystyle\int_0^{2\pi} f(t)\exp(-\mathrm{i}kt)\,dt$ und

$$H^s = \big\{u \in L_2(0,2\pi)\colon\ f \text{ periodisch fortsetzbar und } \|f\|_s < \infty\big\}\,.$$

(11.8) Es gilt

a) $H^s \subset C^0_{\mathrm{per}}[0,2\pi]$ für $s > 1/2$.

b) $C^m_{\mathrm{per}}[0,2\pi] \subset H^m$ für $m \in \mathbb{N}$.

Im Folgenden sei $f \in H^s$ für $s > 1/2$.

(11.9) Es gilt für $\hat{\alpha}_k = \frac{1}{N}\displaystyle\sum_{j=0}^{N-1} f(t_j)\exp(\mathrm{i}kt_j)$ mit $t_j = \frac{2\pi j}{N}$ und $-N/2 < k \leq N/2$

$$\hat{\alpha}_k = \sum_{j\in\mathbb{Z}} \alpha_{k+jN}\,.$$

(11.10) Sei $n < N/2$ und $T_n(t) = \displaystyle\sum_{k=-n}^{n} \hat{\alpha}_k \exp(\mathrm{i}kt)$. Dann existiert eine Konstante $C > 0$ mit

$$\|f - T_n\|_0 \leq \frac{C}{n^s}\,\|f\|_s\,.$$