Vorlesung 401-1652-10L : Numerische Mathematik I

Frühlingssemester 2014 : Prof. Ralf Hiptmair (SAM, D-MATH)

[Vorlesungsniederschrift, Rohformat, URL: http://www.math.ethz.ch/education/bachelor/lectures/fs2014/math/nm]

VI. Iterative Lösung linearer Gleichungssysteme : Das Verfahren der konjugierten Gradienten (CG)

Focus : LGS $A\underline{x} = \underline{b}$, $\underline{A} \in \mathbb{R}^{n\times n}$ s.p.d., $n \in \mathbb{N}$

↓

A regulär

6.0. Minimierungsproblem

Lemma 6.0.B (Äquivalentes Minimierungsproblem)
Für s.p.d. $\mathbf{A} \in \mathbb{R}^{n\times n}$ und beliebiges $\underline{b} \in \mathbb{R}^n$ gilt

$$\mathbf{A}\underline{x}^* = \underline{b} \iff \underline{x}^* = \operatorname*{argmin}_{\underline{x}\in\mathbb{R}^n} J(\underline{x}) ,$$

mit dem quadratischen Funktional

$$J : \mathbb{R}^n \to \mathbb{R} \quad , \quad J(\underline{x}) = \tfrac{1}{2}\underline{x}^\top \mathbf{A}\underline{x} - \underline{b}^\top \underline{x} .$$

Def. 6.0.A (Energienorm)
Für eine symmetrisch positiv definite (s.p.d.) Matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ definiert

$$\|\underline{x}\| := \left(\underline{x}^\top \mathbf{A}\underline{x}\right)^{1/2}$$

eine Norm auf $\mathbb{R}^n$, die von $\mathbf{A}$ induzierte Energienorm.

Beweis : $J(x) - J(x^*) = \tfrac{1}{2}\|x - x^*\|_A^2$ (Δ) □

Bsp : Visualisierung quadratisches Funktional

$A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$ s.p.d., $b = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ ⇒ J strikt konvex

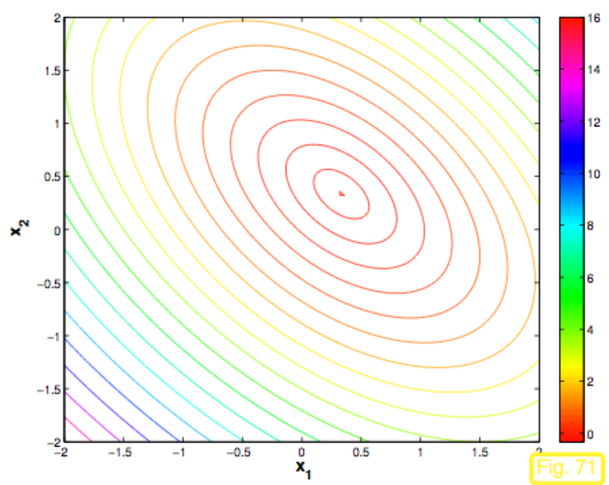


Fig. 71

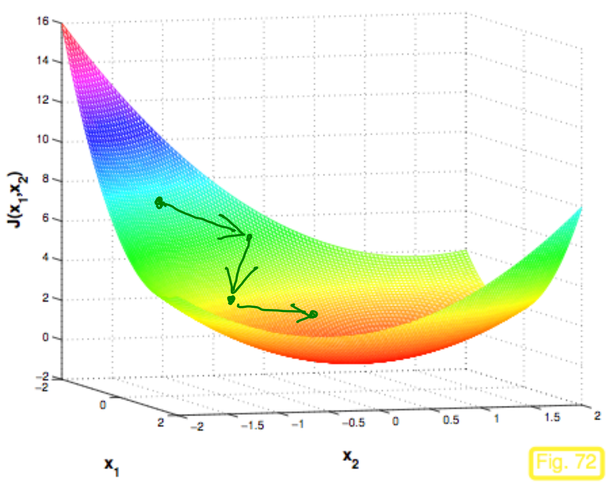


Fig. 72

6.1. Steilster Abstieg (Gradientenverfahren)

6.1.1. Algorithmus (zur lokal Minimierung von $J: \mathbb{R}^n \to \mathbb{R}$)

Abstiegsrichtung in $x \in \mathbb{R}^n$: $-\operatorname{grad} J(x)$

Anfangsnäherung $\underline{x}^{(0)} \in D$, $k = 0$
repeat
$\underline{d}_k := -\mathbf{grad}\, J(\underline{x}^{(k)})$
$\tau^* := \underset{t \in \mathbb{R}}{\operatorname{argmin}}\, J(\underline{x}^{(k)} + \tau \underline{d}_k)$ (Liniensuche)*
$\underline{x}^{(k+1)} := \underline{x}^{(k)} + \tau^* \underline{d}_k$
$k := k + 1$
until $\left(\left\| \underline{x}^{(k)} - \underline{x}^{(k-1)} \right\| \leq \mathrm{tol}_{\mathrm{rel}} \left\| \underline{x}^{(k)} \right\| \text{ or } \left\| \underline{x}^{(k)} - \underline{x}^{(k-1)} \right\| \leq \mathrm{tol}_{\mathrm{abs}} \right)$

→ 1D Minimierung

* i.d.R. approximativ

```
1  function x = gradit(A,b,x,rtol,atol,maxit)
2  r = b-A*x; % residual = search direction
3  for k=1:maxit
4    p = A*r;
5    ts = (r'*r)/(r'*p); % optimal line search parameter   (*)
6    x = x + ts*r;       % update approximate solution
7    cn = (abs(ts)*norm(r)); % norm of correction
8    if ((cn < rtol*norm(x)) || (cn < atol)), return; end
9    r = r - ts*p; % update residual
10 end
11
```

↓ Abbruch

Pro Schritt: 1x Matrix * Vektor
2 Skalarprodukte
2 Vektoroperation
} $O(n)$

6.1.2. Gradientenverfahren für s.p.d. LGS

$A \in \mathbb{R}^{n \times n}$ s.p.d. $\Rightarrow J(\underline{x}) = \frac{1}{2} \underline{x}^T A \underline{x} - \underline{b}^T \underline{x}$
$\operatorname{grad} J(x) = A\underline{x} - \underline{b} = -\underline{r}$ (Residuum)
$\left[\operatorname{grad} J(x) = \left(\frac{\partial J}{\partial x_i}(x) \right)_{i=1}^n \right]$ Liniensuche: $(\underline{r} \neq 0)$

Anfangsnäherung $\underline{x}^{(0)} \in \mathbb{R}^n$, $k = 0$
$\underline{r}_0 := \underline{b} - \mathbf{A}\underline{x}^{(0)}$
repeat
$\tau^* := \frac{\underline{r}_k^\top \underline{r}_k}{\underline{r}_k^\top \mathbf{A} \underline{r}_k}$ (*)
$\underline{x}^{(k+1)} := \underline{x}^{(k)} + \tau^* \underline{r}_k$
$\underline{r}_{k+1} := \underline{r}_k - \tau^* \mathbf{A} \underline{r}_k$ (*)
$k := k + 1$
until $\left(\left\| \underline{x}^{(k)} - \underline{x}^{(k-1)} \right\| \leq \mathrm{tol}_{\mathrm{rel}} \left\| \underline{x}^{(k)} \right\| \text{ or } \left\| \underline{x}^{(k)} - \underline{x}^{(k-1)} \right\| \leq \mathrm{tol}_{\mathrm{abs}} \right)$

$J(\underline{x} + \tau \underline{r}) = J(x) - \tau \underline{r}^T \underline{r} + \frac{1}{2} \tau^2 \underbrace{\underline{r}^T A \underline{r}}_{>0}$
(Parabel in τ)
$\Rightarrow \tau^* = \frac{\underline{r}^T \underline{r}}{\underline{r}^T A \underline{r}}$

→ neue Näherung

(*) $\underline{r}_{k+1} = \underline{b} - A\underline{x}^{(k+1)}$
$= \underline{b} - A\underline{x}^{(k)} - \tau^* A \underline{r}_k$
$= \underbrace{\underline{b} - A\underline{x}^{(k)}}_{\underline{r}_k} - \tau^* A \underline{r}_k$

6.1.3. Gradientenverfahren für s.p.d. LGS: Konvergenz

Iterationsfehler: $\underline{e}^{(k)} = \underline{x}^{(k)} - \underline{x}^*$, $\underline{x}^* := A^{-1}\underline{b}$

$\| \underline{e}^{(k+1)} \|_A^2 \overset{(A)}{=} 2\left(J(\underline{x}^{(k+1)}) - J(\underline{x}^*) \right)$
$\leq 2\left(J(\underline{x}^{(k)} + \tau \underline{r}_k) - J(\underline{x}^*) \right)$ $\forall \tau \in \mathbb{R}$ wg. Liniensuche
$\overset{(A)}{=} \| \underline{x}^{(k)} + \tau \underline{r}_k - \underline{x}^* \|_A^2$
$= \| \underline{e}^{(k)} - \tau A \underline{e}^{(k)} \|_A^2$ $A\underline{e}^{(k)} = -\underline{r}_k$
$\leq \| 1 - \tau A \|_A^2 \, \| \underline{e}^{(k)} \|_A^2$ $\forall \tau \in \mathbb{R}$

↑ Matrixnorm induziert durch $\| \cdot \|_A$

Satz 6.1.3.A (Orthogonale Diagonalisierbarkeit symmetrischer Matrizen)
Zu jeder symmetrischen Matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ gibt es eine orthogonale Matrix $\mathbf{Q} \in \mathbb{R}^{n \times n}$ und reelle Zahlen (Eigenwerte) $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ so, dass

$$\mathbf{Q}^T \mathbf{A} \mathbf{Q} = \operatorname{diag}(\lambda_1, \ldots, \lambda_n) . =: \mathbb{D}$$

Die Menge $\sigma(\mathbf{A}) := \{\lambda_1, \ldots, \lambda_n\}$ heisst das Spektrum von $\mathbf{A}$.

Satz 6.1.3.B (Norm von Matrixpolynomomen)
Sei $\mathbf{A} \in \mathbb{R}^{n\times n}$ s.p.d. und $p \in \mathbb{P}_r$, $r \in \mathbb{N}_0$, ein Polynom. Dann gilt

$$\|p(\mathbf{A})\|_A = \max\{|p(\lambda)|,\ \lambda \in \sigma(\mathbf{A})\}\,,$$

wobei $\|\cdot\|_A$ die von der Energienorm induzierte Matrixnorm ist.

$p(t) = \sum_{j=0}^{l} c_j t^j \Rightarrow p(A) := \sum_{j=0}^{l} c_j A^j$ (Matrixpolynom)

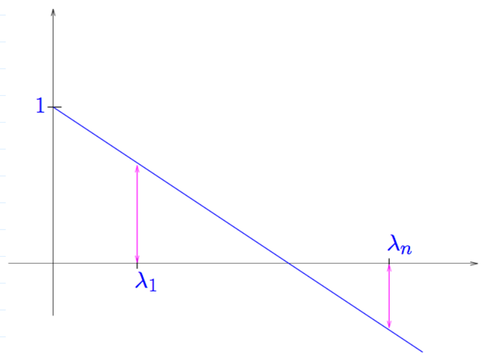
Beweisskizze: $Q^T A Q = D \Rightarrow p(A) = Q\,p(D)\,Q^T$ □

▷ $\|I - \tau A\|_A = \max\{|1-\tau\lambda|,\ \lambda \in \sigma(A)\}$

$= \max\{|1-\tau\lambda_1|, |1-\tau\lambda_n|\}$

$\sigma(A) = \{0 < \lambda_1 < \lambda_2 < \cdots < \lambda_n\}$

▷ $\|e^{(k+1)}\|_A \le \inf_{\tau\in\mathbb{R}} \max\{|1-\tau\lambda_1|, |1-\lambda_n\tau|\}\,\|e^{(k)}\|_A$



$1 - \lambda_1\tau^* = \lambda_n\tau^* - 1$

▷ $\tau^* = \frac{2}{\lambda_1+\lambda_n}$

$\min = \frac{\lambda_n-\lambda_1}{\lambda_n+\lambda_1} = \frac{K-1}{K+1}$

$K = \frac{\lambda_n}{\lambda_1} = \mathrm{cond}_2(A)$

▷ $\|e^{(k+1)}\|_A \le \frac{K-1}{K+1}\,\|e^{(k)}\|_A$

≙ Lineare Konvg. mit Rate $\frac{K-1}{K+1} < 1$
(in der Energienorm!)

Nachtrag: Ableitung eines quadratischen Funktionals

$J(\underline{x}) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_i x_j - \sum_i b_i x_i$

$\frac{\partial J}{\partial x_l}(\underline{x}) = a_{ll}x_l + \frac{1}{2}\sum_{\substack{i=1\\ i\neq l}}^{n} a_{il}x_i + \frac{1}{2}\sum_{\substack{j=1\\ j\neq l}}^{n} a_{lj}x_j - b_l$

$a_{ij} = a_{ji}$

$= a_{ll}x_l + \sum_{\substack{j=1\\ j\neq l}}^{n} a_{lj}x_j - b_l = (A\underline{x}-\underline{b})_l$

$J(\underline{x}+\underline{h}) = \frac{1}{2}(\underline{x}+\underline{h})^T A(\underline{x}+\underline{h}) - (\underline{x}+\underline{h})^T\underline{b}$

$= J(x) + \underline{h}^T(A\underline{x}-\underline{b}) + \underbrace{\frac{1}{2}\underline{h}^T A\underline{h}}_{O(\|\underline{h}\|^2)}$

▷ $\lim_{h\to 0} \frac{J(\underline{x}+\underline{h}) - J(x) - \underline{h}^T(A\underline{x}-\underline{b})}{\|\underline{h}\|} = 0$

6.2. Das Verfahren der konjugierten Gradienten

6.2.1 Die CG-Iteration

↳ "Gradientenverfahren mit Gedächtnis"

Idee: Minimierung von J in Unterräumen

Gegeben: $U_R \subset \mathbb{R}^n$: $\mathcal{U}_1 \subset \mathcal{U}_2 \subset \cdots \subset \mathcal{U}_n = \mathbb{R}^n$, $\dim \mathcal{U}_k = k$

$\underline{x}^{(k)} = \underset{\underline{x}\in \underline{x}^{(0)}+\mathcal{U}_k}{\operatorname{argmin}} J(\underline{x})$ (a)

▷ $\|e^{(k+1)}\|_A \le \|e^{(k)}\|_A$ aus (Δ) & "c"

Idee: Akkumulation lokaler Abstiegsrichtungen

$\mathcal{U}_0 = \{0\}$; $\mathcal{U}_{k+1} = \mathrm{Span}\{\mathcal{U}_k, -\mathrm{grad}J(x^{(k)}) = r_k\}$ (b)

$r_k \neq 0$ sonst fertig

(a) & (b) definieren CG-Verfahren

Lemma 6.2.1.A (Darstellungsformel für U_k)
Für $\underline{x}^{(0)} \in \mathbb{R}^n$ setze $U_0 := \{\mathbf{0}\}$,

$$\begin{cases} U_k := \mathrm{span}\{U_{k-1}, \underline{r}_{k-1} := \underline{b} - \mathbf{A}\underline{x}^{(k-1)}\} , \\ \underline{x}^{(k)} := \underset{\underline{x} \in x^{(0)} + U_k}{\mathrm{argmin}} \; J(\underline{x}) . \end{cases} \qquad k = 1, \ldots, n .$$

Falls $\underline{r}_0, \ldots, \underline{r}_{k-1} \neq 0$, dann gilt

$$U_k = \mathrm{span}\{\underline{r}_0, \mathbf{A}\underline{r}_0, \ldots, \mathbf{A}^{k-1}\underline{r}_0\} , \quad k = 1, \ldots, n .$$

Beweis: $\underline{r}_k = \underline{b} - A\underline{x}^{(k)} = \underline{b} - A(\underline{x}^{(0)} + \underline{u}) \;, \underline{u} \in \mathcal{U}_k$

(Induktion) $= \underline{r}_0 + A\underline{u} \in \mathrm{Span}\{\mathcal{U}_k, A\mathcal{U}_k\} = \mathcal{U}_{k+1}$ □

Beweis: $\forall \underline{u} \in \mathcal{U}_k$: $\frac{d}{dt} J(\underline{x}^{(k)} + t\underline{u})\big|_{t=0} \overset{\text{Kettenregel}}{=} (\underline{b} - A\underline{x}^{(k)})^T \underline{u} = 0$

$\underline{r}_k$ □

Beachte: $\mathcal{U}_k = \mathrm{Span}\{\underline{r}_0, \underline{r}_1, \ldots, \underline{r}_{k-1}\}$

Korollar 6.2.1.D
Seien $\underline{r}_0, \ldots, \underline{r}_k$, $0 \leq k < n$, die vom CG-Verfahren erzeugten Residuen. Falls $\underline{r}_\ell \neq 0$, $\ell = 0, \ldots, k$, dann ist $\{\underline{r}_0, \ldots, \underline{r}_k\}$ eine Orthogonalbasis des Krylov-Raumes $\mathcal{K}_{k+1}(\mathbf{A}, \underline{r}_0)$.

▷ Fall $\underline{r}_k \neq 0 \;\forall k \Rightarrow \dim \mathcal{U}_k = k$

Satz 6.2.1.E (Abbruch des CG-Verfahrens)
Sei $\underline{x}^{(k)}$, $k = 0, \ldots, n$, die durch das CG-Verfahren erzeugte Folge. Dann gibt es ein $m \in \{0, \ldots, n\}$ so, dass $\underline{x}^{(m)} = \mathbf{A}^{-1}\underline{b}$.

Def. 6.2.1.B (Krylov-Raum)
Für eine Matrix $\mathbf{M} \in R^{n \times n}$ und einen Vektor $\underline{z} \in R^n$ heisst

$$\mathcal{K}_\ell(\mathbf{M}, \underline{z}) = \mathrm{span}\left\{\underline{z}, \mathbf{M}\underline{z}, \mathbf{M}^2\underline{z}, \ldots, \mathbf{M}^{\ell-1}\underline{z}\right\} , \quad \ell \in \mathbb{N} ,$$

der ℓ-te Krylov-Raum zu $\mathbf{M}$ und $\underline{z}$.

$\underline{y} \in \mathcal{K}_\ell(M, \underline{z}) \Rightarrow \exists\, p \in \mathbb{P}_{\ell-1} : \underline{y} = p(M)\underline{z}$

▷ Lemma 6.2.1.A $\Rightarrow$ $\mathcal{U}_k = \mathcal{K}_k(A, \underline{r}_0)$

Notation: $V \subset \mathbb{R}^n$: $\underline{z} \perp V \;:\Leftrightarrow\; \underline{z}^\top \underline{v} = 0 \;\; \forall \underline{v} \in V$

Lemma 6.2.1.C (Orthogonalität der CG-Residuen)
Sei $\underline{x}^{(k)}$, $k = 0, \ldots, n$, die durch das CG-Verfahren erzeugte Folge. Dann gilt

$$\underline{r}_k \perp U_k \quad \text{für die Residuen} \quad \underline{r}_k := \underline{b} - \mathbf{A}\underline{x}^{(k)} , \quad k = 0, \ldots, n .$$

6.2.2. "Konvergenz" des CG-Verfahrens

↳ Für $n \gg 1$, wenige Schritte

$$\|\underline{x}^{(k+1)} - \underline{x}^*\|_A^2 \overset{(A)}{=} 2\left(J(\underline{x}^{(k+1)}) - J(\underline{x}^*)\right)$$

$$= \inf_{z \in x^{(0)} + \mathcal{U}_k} \|z - x^*\|_A^2$$

$\mathcal{U}_k = \mathcal{K}_k(A, \underline{r}_0)$

$$\leq \inf\{\|\underbrace{x^{(0)} + p(A)r_0 - \underline{x}^*}_{= \underline{e}^{(0)} - p(A)A\underline{e}^{(0)} = (\underbrace{1 - p(A)A}_{=:q(A)})e^{(0)}}\|_A^2 , \; p \in \mathbb{P}_{k-1}\}$$

$$\leq \inf\{\|q(A)\|_A^2 \|e^{(0)}\|_A^2 : q \in \mathbb{P}_k, q(0) = 1\}$$

(6.1.3.B)

$$\leq \inf\{|q(\lambda)|^2, \lambda \in \sigma(A), q \in \mathbb{P}_k, q(0) = 1\} \cdot \|e^{(0)}\|_A^2$$

[grob!]

$$\leq \inf\{|q(\lambda)|^2; \lambda \in \Lambda := [\lambda_1, \lambda_n], q \in \mathbb{P}_k, q(0) = 1\} \cdot \|e^{(0)}\|_A^2$$

$q^* := \operatorname{argmin} \{ \|q\|_{\infty,\Lambda}, q \in \mathbb{P}_k, q(0) = 1 \}$

$p^* := \{ \quad // \quad p$ hat führenden Koeffizienten $= 1 \}$

Satz 3.26 (min-max Eigenschaft der Tschbyscheffpolynome)

Für alle $k \in \mathbb{N}_0$: $\left\| 2^{1-k} T_k \right\|_{\infty,[-1,1]} = \inf_{q \in \mathbb{P}_k^1} \|q\|_{\infty,[-1,1]}$.

$$p^*(t) = 2^{1-k} \left(\frac{\lambda_n - \lambda_1}{2}\right)^k T_k\left(2\frac{t-\lambda_1}{\lambda_n-\lambda_1} - 1\right)$$

$$\Rightarrow q^* = \frac{p^*}{p^*(0)} \Rightarrow \text{Formel für } q^*$$

$$\|e^{(k+1)}\|_A \leq \|q^*\|_{\infty,\Lambda} \cdot \|e^{(0)}\|_A$$

$$\|T_k\|_{\infty,\Lambda} = 1 \Rightarrow \|q^*\|_{\infty,\Lambda} = \frac{1}{T_k\left(-\frac{2\lambda_1}{\lambda_n-\lambda_1} - 1\right)} = \frac{1}{T_k\left(-\frac{\kappa+1}{\kappa-1}\right)} \; ; \kappa := \frac{\lambda_n}{\lambda_1}$$

Satz 6.2.2.B (Konvergenzabschätzung für das CG-Verfahren)

Sei $\underline{x}^{(k)}$, $k = 0, \ldots, n$, die durch das CG-Verfahren (zum linearen Gleichungssystem $\mathbf{A}\underline{x} = \underline{b}$ mit s.p.d. Matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$) erzeugte Folge und $x^* := \mathbf{A}^{-1}\underline{b}$. Dann gilt

$$\left\| \underline{x}^{(k+1)} - \underline{x}^* \right\|_A \leq 2 \left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^k \left\| \underline{x}^{(0)} - \underline{x}^* \right\|_A , \quad \kappa := \operatorname{cond}(\mathbf{A}) .$$

$\neq$ lineare Konvergenz

$$\kappa = \kappa(A) = \frac{\lambda_n(A)}{\lambda_1(A)} = \operatorname{cond}_2(A)$$

Wiederholung: $A\underline{x} = \underline{b}$, $A \in \mathbb{R}^{n,n}$ s.p.d.

$\Rightarrow x^* = A^{-1}\underline{b} = \operatorname*{argmin}_{x \in \mathbb{R}^n} J(\underline{x})$, $J(\underline{x}) = \frac{1}{2}\underline{x}^T A \underline{x} - \underline{b}^T \underline{x}$

CG: $x^{(k)} = \operatorname*{argmin}_{x \in \mathcal{U}_k} J(x^{(0)} + X)$, $\mathcal{U}_k = \mathcal{K}_{k-1}(A, \underline{r}_0)$

Lemma 6.2.1.E (Darstellungsformel für Tschebyscheff-Polynome)

Für das k. Tschebyscheff-Polynom ($\rightarrow$ Def. 3.27) T_k gilt für $|t| \geq 1$

$$T_k(t) = \frac{1}{2}\left(\left(t + \sqrt{t^2-1}\right)^k + \left(t - \sqrt{t^2-1}\right)^k\right) .$$

Beweis über Dreitermrekursion $\rightarrow$ Lemma

$$t^* = \frac{\kappa+1}{\kappa-1} \Rightarrow t^* \pm \sqrt{t^{*2}-1} = \left(\frac{\sqrt{\kappa}+1}{\sqrt{\kappa}-1}\right)^{\pm 1} \Rightarrow |T_k(-t^*)| \geq \frac{1}{2}\left(\frac{\sqrt{\kappa}+1}{\sqrt{\kappa}-1}\right)^k$$

6.2.3. CG - Algorithmus

Annahme: (i) $\underline{r}_k \neq 0 \Rightarrow \dim \mathcal{U}_k = k$

(ii) Geschachtelte Basis von $\mathcal{U}_k$ verfügbar

$\mathcal{U}_k = \operatorname{Span}\{p_1, \ldots, p_k\}$

$P = [p_1, \ldots, p_k] \in \mathbb{R}^{n,k}$

$\Rightarrow \exists z_k \in \mathbb{R}^k : \quad x^{(k)} = x^{(0)} + P\underline{z}_k$ (d)

$$\underline{z}_k = \operatorname*{argmin}_{\underline{z} \in \mathbb{R}^k} J(x^{(0)} + P\underline{z})$$

$$\Rightarrow 0 = \operatorname{grad}_z J(x^{(0)} + P\underline{z}_k) = P^T(A(\underline{x}^{(0)} + P\underline{z}_k) - \underline{b})$$

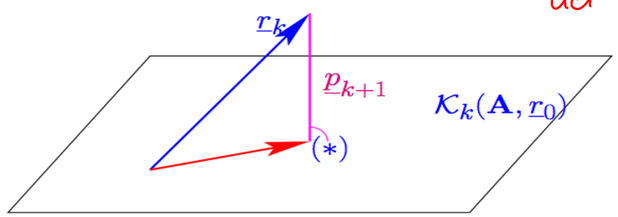
$$\Leftrightarrow \text{LGS}: \quad P^T A P \underline{z}_k = P^T(b - A x^{(0)}) = P^T \underline{r}_0 \quad \text{(e)}$$

Idee: Wenn $p_j^T A p_\ell = 0$ für $\ell \neq j$ (A-Orthogonalität)

$\Rightarrow$ $P^T A P$ diagonal $\overset{(e)}{\Rightarrow}$ $z_k = \begin{pmatrix} z_{k-1} \\ \frac{p_k^T r_0}{p_k^T A p_k} \end{pmatrix}$

(d) ▷ $$x^{(k)} = x^{(k-1)} + \underbrace{\frac{p_k^T r_0}{p_k^T A p_k}}_{=: \alpha_k \in \mathbb{R}} p_k \quad (f)$$ (Updateformel)

Konstruktion der p_k: Gram-Schmidt A-Orthogonalisierung der Residuenvektoren



$$p_{k+1} = r_k - \sum_{\ell=1}^{k} \frac{r_k^T A p_\ell}{p_\ell^T A p_\ell} p_\ell$$

Erinnerung Lemma 6.2.1.C: $r_k \perp \mathcal{U}_k \Rightarrow r_k^T A p_\ell = 0$ für $\ell < k$

($A p_\ell \in \mathcal{U}_k$ wenn $p_\ell \in \mathcal{U}_{k-1}$)

▷ $$p_{k+1} = r_k - \frac{r_k^T A p_k}{p_k^T A p_k} p_k \quad (g)$$

▭▭ ≙ Kurze Rekursionen

Aus $r_k = b - A x^{(k)}$

$\overset{(f)}{=} r_{k-1} - \alpha_k A p_k \quad (h)$

CG-Algorithmus

Eingabe : Anfangsnäherung $\underline{x}^{(0)} \in \mathbb{R}^n$
Ausgabe : Näherungslösung $\underline{x}^{(l)} \in \mathbb{R}^n$

$\underline{p}_1 = \underline{r}_0 := \underline{b} - \mathbf{A}\underline{x}^{(0)}$;
for $j = 1$ **to** l **do** {

(f): $\underline{x}^{(j)} := \underline{x}^{(j-1)} + \frac{\underline{p}_j^\top \underline{r}_{j-1}}{\underline{p}_j^\top \mathbf{A}\underline{p}_j} \underline{p}_j$;

(h): $\underline{r}_j = \underline{r}_{j-1} - \frac{\underline{p}_j^\top \underline{r}_{j-1}}{\underline{p}_j^\top \mathbf{A}\underline{p}_j} \mathbf{A}\underline{p}_j$;

(g): $\underline{p}_{j+1} = \underline{r}_j - \frac{(\mathbf{A}\underline{p}_j)^\top \underline{r}_j}{\underline{p}_j^\top \mathbf{A}\underline{p}_j} \underline{p}_j$;

}

Eingabe: Anfangsnäherung $\underline{x} \hat{=} \underline{x}^{(0)} \in \mathbb{R}^n$
Toleranz $\tau > 0$
Ausgabe: Näherungslösung $\underline{x} \hat{=} \underline{x}^{(l)}$

$\underline{p} := \underline{r}_0 := \underline{r} := \underline{b} - \mathbf{A}\underline{x}$;
for $j = 1$ **to** $l_{\max}$ **do** {
$\beta := \underline{r}^\top \underline{r}$;
$\underline{h} := \mathbf{A}\underline{p}$;
$\alpha := \frac{\beta}{\underline{p}^\top \underline{h}}$;
$\underline{x} := \underline{x} + \alpha \underline{p}$;
$\underline{r} := \underline{r} - \alpha \underline{h}$;
if $\|\underline{r}\| \leq \tau \|\underline{r}_0\|$ **then stop**;
$\beta := \frac{\underline{r}^\top \underline{r}}{\beta}$;
$\underline{p} := \underline{r} + \beta \underline{p}$;
}

Rechenaufwand pro Schritt: 3 Skalarprodukte, 3 Vektoroperationen } $O(n)$
& 1x Ax Vektor

Beachte: (h) $\Rightarrow$ $r_k^T p_{k+1} = \left(r_0 + \sum_{j=1}^{k} \pm \alpha_j \cdot A p_j\right)^T p_{k+1} = r_0^T p_{k+1}$

[A-Orth.: $p_j^T A p_{k+1} = 0$ für $j \leq k$]

```
function x = cg(evalA,b,x,tol,maxit)
% x supplies initial guess, maxit maximal number of CG steps
% evalA must pass a handle to a MATLAB function realizing A*x
r = b - evalA(x); rho = 1; n0 = norm(r);
for i = 1 : maxit
    rho1 = rho; rho = r' * r;
    if (i == 1),
      p = r;
    else
      beta = rho/rho1;
      p = r + beta * p;    % update of A-orthogonal basis vectors
    end
    q = evalA(p);
    alpha = rho/(p' * q);
    x = x + alpha * p;     % update of approximate solution
    if (norm(b-A*x) <= tol*n0) % termination criterion
      return;
    end
    r = r - alpha * q;     % update of residual
end
```

Zum Abbruchkriterium:

$e^{(k)} = x^* - x^{(k)} \Rightarrow Ae^{(k)} = r_k$ (Fehlergleichung)

▷ $\|r_k\| \leq \|A\| \, \|e^{(k)}\|$

$\|e^{(k)}\| \leq \|A^{-1}\| \, \|r_k\|$

$$\mathrm{cond}(A)^{-1} \frac{\|r_k\|}{\|r_0\|} \leq \frac{\|e^{(k)}\|}{\|e^{(0)}\|} \leq \|A^{-1}\| \|A\| \frac{\|r_k\|}{\|r_0\|}$$

[Allgemein: $\|Mx\| \leq \|M\| \, \|x\|$

↳ von $\|\cdot\|$ induzierte Matrixnorm]

$R = [r_1, \ldots, r_{10}]$

$R^T R =$

1.000000	−0.000000	0.000000	−0.000000	0.000000	−0.000000	0.016019	−0.795816	−0.430569	0.348133
−0.000000	1.000000	−0.000000	0.000000	−0.000000	0.000000	−0.012075	0.600068	−0.520610	0.420903
0.000000	−0.000000	1.000000	−0.000000	0.000000	−0.000000	0.001582	−0.078664	0.384453	−0.310577
−0.000000	0.000000	−0.000000	1.000000	−0.000000	0.000000	−0.000024	0.001218	−0.024115	0.019394
0.000000	−0.000000	0.000000	−0.000000	1.000000	−0.000000	0.000000	−0.000002	0.000151	−0.000118
−0.000000	0.000000	−0.000000	0.000000	−0.000000	1.000000	−0.000000	0.000000	−0.000000	0.000000
0.016019	−0.012075	0.001582	−0.000024	0.000000	−0.000000	1.000000	−0.000000	−0.000000	0.000000
−0.795816	0.600068	−0.078664	0.001218	−0.000002	0.000000	−0.000000	1.000000	0.000000	−0.000000
−0.430569	−0.520610	0.384453	−0.024115	0.000151	−0.000000	−0.000000	0.000000	1.000000	0.000000
0.348133	0.420903	−0.310577	0.019394	−0.000118	0.000000	0.000000	−0.000000	0.000000	1.000000

▷ Orthogonalitätsverlust der Residuen wegen Rundungsfehlern

▷ CG instabil als direkter Löser

Satz 6.2.2.B bleibt gültig für $k \ll n$ auch in Gleitpunktarithmetik!

▷ CG als iterativer Löser für grosse s.p.d. LGS

6.2.4. CG in Gleitpunktarithmetik

Bsp: $(A)_{ij} = \frac{1}{i+j+1}$, $1 \leq i,j \leq n$, $n = 20$ (Hilbertmatrix)

$b = (1, \ldots, 1)^T$

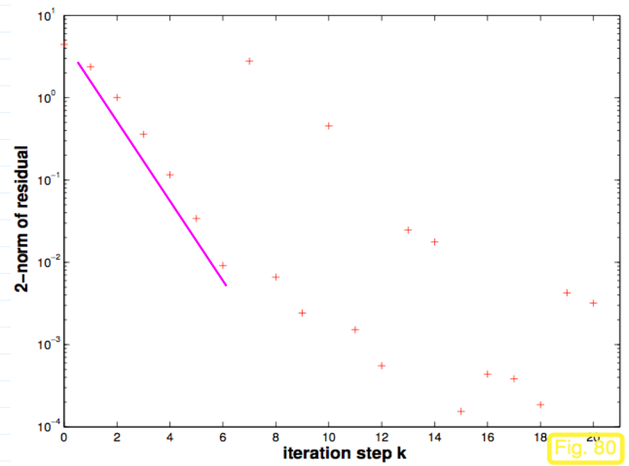


Fig. 80

△ $r_{20} \neq 0$!

Gründe für den Einsatz eines iterativen Lösers (CG) für grosses LGS (im Vergleich zu direkter Elimination)

(G1) Nur Matrix × Vektor ('evalA') verfügbar.

(G2) $A \times$ Vektor billig, z.B. bei dünnbesetzten Matrizen, aber LU-Faktor (praktisch) vollbesetzt ("fill-in")

(G3) $A \times$ Vektor billig und $\underbrace{\kappa(A) \approx 1}$

→ schnelle Kvgz. von CG

(G4) Gute Anfangsnäherung verfügbar oder geringe Genauigkeitsanforderung

CG in MATLAB:

```
x=pcg(A,b,tol,maxit,[],[],x0)

x=pcg(Afun,b,tol,maxit,[],[],x0)

[x,flag,relr,it,resv] = pcg(...)
```

Allgemeines A: gmres, bicg, qmr

Aus Spektralsatz: $M = Q D Q^T \Rightarrow W = Q \sqrt{D} Q^T$

Ähnlichkeit: $\kappa(\widehat{A}) = \kappa(B^{-1/2} A B^{-1/2}) = \kappa(B^{-1}A)$

$\rightarrow$ Satz 6.2.2.B: Schnelle Konvg. von CG für $\widehat{A}\widehat{x} = \widehat{b}$ (bzgl. $\widehat{A}$-Norm)

Aber: Berechnung von $B^{-1/2}$ i.d.R. sehr teuer

6.3. Vorkonditionierung

S.p.d. LGS $Ax = b$, $\kappa(A) \gg 1$, $A \in \mathbb{R}^{n,n}$

Idee: Finde s.p.d. $B \in \mathbb{R}^{n,n}$ mit

(i) $B^{-1}\times$ Vektor so teuer wie $A\times$ Vektor
(ii) $\kappa(B^{-1}A) = \frac{\lambda_n(B^{-1}A)}{\lambda_1(B^{-1}A)} \ll \kappa(A)$

$\rightarrow$ B heisst Vorkonditionierer

und wende CG-Iteration an auf transformiertes LGS

$\widetilde{A}\widetilde{x} = \widetilde{b}$, $\widehat{A} = B^{-1/2} A B^{-1/2}$ s.p.d.,
$\widehat{b} = B^{-1/2} b$, $\widehat{x} = B^{1/2} x$

Satz 6.3.A (Wurzel einer s.p.d. Matrix)
Zu jeder s.p.d. Matrix $\mathbf{M} \in \mathbb{R}^{n\times n}$ gibt es eine eindeutige s.p.d. Matrix $\mathbf{W} \in \mathbb{R}^{n\times n}$ so, dass $\mathbf{W}^2 = \mathbf{M}$. $\mathbf{W}$ heisst die Wurzel von $\mathbf{M}$. : Notation $W = M^{1/2}$

CG für $\widetilde{\mathbf{A}}\underline{\widetilde{x}} = \underline{\widetilde{b}}$

Eingabe : Anfangsnäherung $\underline{\widetilde{x}}^{(0)} \in \mathbb{R}^n$
Ausgabe : Näherungslösung $\underline{\widetilde{x}}^{(l)} \in \mathbb{R}^n$

$\underline{\widetilde{p}}_1 := \underline{\widetilde{r}}_0 := \underline{\widetilde{b}} - \mathbf{B}^{-1/2}\mathbf{A}\mathbf{B}^{-1/2}\underline{\widetilde{x}}^{(0)}$;
for $j = 1$ **to** l **do** {

$$\alpha := \frac{\underline{\widetilde{p}}_j^T \underline{\widetilde{r}}_{j-1}}{\underline{\widetilde{p}}_j^T \mathbf{B}^{-1/2}\mathbf{A}\mathbf{B}^{-1/2}\underline{\widetilde{p}}_j}$$

(f): $\underline{\widetilde{x}}^{(j)} := \underline{\widetilde{x}}^{(j-1)} + \alpha \underline{\widetilde{p}}_j$;

(h): $\underline{\widetilde{r}}_j = \underline{\widetilde{r}}_{j-1} - \alpha \mathbf{B}^{-1/2}\mathbf{A}\mathbf{B}^{1/2}\underline{\widetilde{p}}_j$;

(g): $\underline{\widetilde{p}}_{j+1} = \underline{\widetilde{r}}_j - \frac{(\mathbf{B}^{-1/2}\mathbf{A}\mathbf{B}^{-1/2}\underline{\widetilde{p}}_j)^T \underline{\widetilde{r}}_j}{\underline{\widetilde{p}}_j^T \mathbf{B}^{-1/2}\mathbf{A}\mathbf{B}^{-1/2}\underline{\widetilde{p}}_j}\underline{\widetilde{p}}_j$;

}

Äquivalentes CG in transformierten Variablen

Eingabe : Anfangsnäherung $\underline{x}^{(0)} \in \mathbb{R}^n$
Ausgabe : Näherungslösung $\underline{x}^{(l)} \in \mathbb{R}^n$

$\mathbf{B}^{1/2}\underline{\widetilde{r}}_0 := \mathbf{B}^{1/2}\underline{\widetilde{b}} - \mathbf{A}\mathbf{B}^{-1/2}\underline{\widetilde{x}}^{(0)}$;
$\mathbf{B}^{-1/2}\underline{\widetilde{p}}_1 := \mathbf{B}^{-1}(\mathbf{B}^{1/2}\underline{\widetilde{r}}_0)$;
for $j = 1$ **to** l **do** {

$$\alpha := \frac{(\mathbf{B}^{-1/2}\underline{\widetilde{p}}_j)^T \mathbf{B}^{1/2}\underline{\widetilde{r}}_{j-1}}{(\mathbf{B}^{-1/2}\underline{\widetilde{p}}_j)^T \mathbf{A}\mathbf{B}^{-1/2}\underline{\widetilde{p}}_j}$$

$\mathbf{B}^{-1/2}\underline{\widetilde{x}}^{(j)} := \mathbf{B}^{-1/2}\underline{\widetilde{x}}^{(j-1)} + \alpha \mathbf{B}^{-1/2}\underline{\widetilde{p}}_j$;

$\mathbf{B}^{1/2}\underline{\widetilde{r}}_j = \mathbf{B}^{1/2}\underline{\widetilde{r}}_{j-1} - \alpha \mathbf{A}\mathbf{B}^{-1/2}\underline{\widetilde{p}}_j$;

$$\mathbf{B}^{-1/2}\underline{\widetilde{p}}_{j+1} = \mathbf{B}^{-1}(\mathbf{B}^{-1/2}\underline{\widetilde{r}}_j) - \frac{(\mathbf{B}^{-1/2}\underline{\widetilde{p}}_j)^T \mathbf{A}\mathbf{B}^{-1}(\mathbf{B}^{1/2}\underline{\widetilde{r}}_j)}{(\mathbf{B}^{-1/2}\underline{\widetilde{p}}_j)^T \mathbf{A}\mathbf{B}^{-1/2}\underline{\widetilde{p}}_j}\mathbf{B}^{-1/2}\underline{\widetilde{p}}_j;$$

}

$\widehat{r}_k = B^{-1/2} r_k$ $\widehat{p}_k = B^{-1/2} p_k$ $\widehat{x}^{(k)} = B^{1/2} x^{(k)}$

Eingabe: Anfangsnäherung $\underline{x} \in \mathbb{R}^n \hat{=} \underline{x}^{(0)} \in \mathbb{R}^n$, Toleranz $\tau > 0$
Ausgabe: Näherungslösung $\underline{x} \hat{=} \underline{x}^{(l)}$

$\underline{p} := \underline{r} := \underline{b} - \mathbf{A}\underline{x}$; $\underline{p} := \mathbf{B}^{-1}\underline{r}$; $\underline{q} := \underline{p}$; $\tau_0 := \underline{p}^\top \underline{r}$;
for $l = 1$ **to** $l_{\max}$ **do** {
$\beta := \underline{r}^\top \underline{q}$; $\underline{h} := \mathbf{A}\underline{p}$; $\alpha := \frac{\beta}{\underline{p}^\top \underline{h}}$;
$\underline{x} := \underline{x} + \alpha \underline{p}$;
$\underline{r} := \underline{r} - \alpha \underline{h}$;
$\underline{q} := \mathbf{B}^{-1}\underline{r}$; $\beta := \frac{\underline{r}^\top \underline{q}}{\beta}$;
if $|\underline{q}^\top \underline{r}| \leq \tau \cdot \tau_0$ **then stop**;
$\underline{p} := \underline{q} + \beta \underline{p}$;
}

Wurzelfrei !

Es gilt: $\|\underline{x}^{(k)} - \underline{x}^*\|_A \leq 2 \left(\frac{\sqrt{\mathrm{cond}(A)} - 1}{\sqrt{\mathrm{cond}(A)} + 1} \right)^k \|\underline{x}^{(0)} - \underline{x}^*\|_A$

Beachte: $\|\hat{x}\|_{\hat{A}}^2 = (B^{1/2}x)^T (B^{-1/2} A B^{-1/2})(B^{1/2}x) = \|x\|_A^2$

Bsp: Bandvorkonditionierer

$$(B)_{ij} = \begin{cases} (A)_{ij} & , |i-j| \leq m \\ 0 & , \text{sonst} \end{cases}$$

▽ B nicht s.p.d möglich für $m > 0$

Nun $m = 1$:
$n \gg 1$
gerade

$$A = \begin{pmatrix} 2+1/n & -1 & & & 1/n \\ -1 & & & & 1/n \\ & & & & \\ 1/n & & & & -1 \\ 1/n & & & -1 & 2+1/n \end{pmatrix}$$

▷ Aufwand Matrix × Vektor $= O(n)$

$A = T + N$, $B = T$

Tridiagonalanteil ↑ ↖ "klein": $\|N\| \ll \|T\|$

$B^{-1}A = 1 + T^{-1}N$

↳ klein $\Rightarrow \kappa(B^{-1}A) \approx 1$

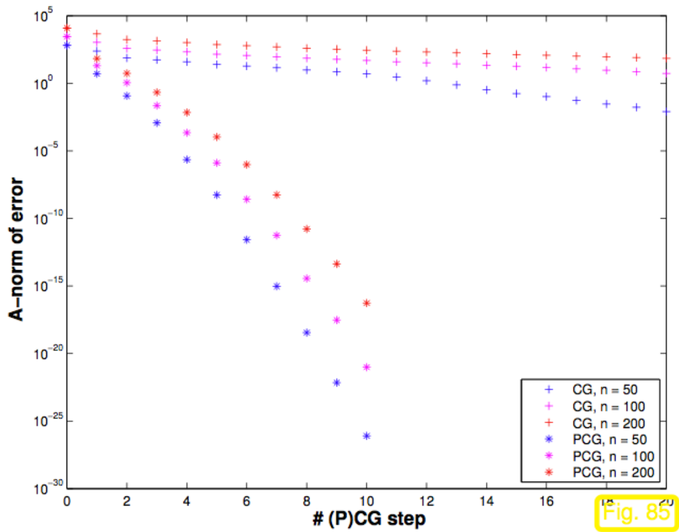


Fig. 85

Beachte: Hier "$B^{-1} \times$ Vektor" mit Aufwand $O(n)$

▷ $B = T$ ist guter Vorkonditionierer