Numerische Mathematik für StudentInnen
mit Ziel Diplom in Mathematik

Gerhard Opfer WS 1997/1998 (1. Teil) SS 1998 (2. Teil)

Bogen 18: Übungen am 19. Mai 1998

Lösungshinweise

Aufgabe P52. Ist $\mathbf{A}$ eine HPD-Matrix, und ist $\mathbf{x}^*\mathbf{A}\mathbf{y} = 0$ für $\mathbf{x}, \mathbf{y} \neq 0$, so heißen $\mathbf{x}, \mathbf{y}$ *konjugiert in bezug auf* $\mathbf{A}$. Unter einer HPD-Matrix versteht man eine komplexe, *h*ermitesche, *p*ositiv *d*efinite Matrix $\mathbf{A}$. Eine reelle, hermitesche Matrix ist symmetrisch, eine reelle HPD-Matrix ist also eine SPD-Matrix. Die konjugierten Richtungen gestatten die Aufspaltung des Minimierungsproblems, das in $H_{j+1} := \langle \mathbf{h}_0, \mathbf{h}_1, \ldots, \mathbf{h}_j \rangle$ gestellt ist auf den Raum in H_j und auf den eindimensionalen Raum $\langle \mathbf{h}_j \rangle$.

Aufgabe P53. Die konjugierte Richtung $\mathbf{h}_{j+1}$ läßt sich linear kombinieren aus dem Residuum (= Richtung des steilsten Abstiegs) $\mathbf{r}_{j+1}$ und aus der Richtung $\mathbf{h}_j$. Das ergibt eine einfache Berechnungsformel (5.57), (5.58) für die neue Richtung.

Aufgabe P54. Es ist hier $\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & a \end{pmatrix}$. Sucht man z. B. eine zu $\mathbf{h}_0 = (1,1)^{\mathrm{T}}$ konjugierte Richtung $\mathbf{h}_1 = (h_1, h_2)^{\mathrm{T}}$, so gilt dafür die Gleichung $\mathbf{h}_0^{\mathrm{T}}\mathbf{A}\mathbf{h}_1 = h_1 + ah_2 = 0$ mit Lösung $h_1 = 1, h_2 = -1/a$.

Aufgabe 35. Seien $\mathbf{z} = \begin{pmatrix} x \\ y \end{pmatrix}$ und $\mathbf{h} = \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} \neq 0$ gegeben. Das Minimum von $g(\alpha) := f(\mathbf{z}+\alpha\mathbf{h}), \alpha \in \mathbb{R}$ liegt dann für das gegebene $f(\mathbf{z}) := \frac{1}{2}(x^2 + ay^2)$, $a > 0$ bei $\hat{\alpha} = -\frac{h_1 x + a h_2 y}{h_1^2 + a h_2^2}$, und $\hat{\mathbf{z}} := \mathbf{z} + \hat{\alpha}\mathbf{h} = \begin{pmatrix} x + \hat{\alpha}h_1 \\ y + \hat{\alpha}h_2 \end{pmatrix}$. Die Richtung des steilsten Abstiegs ist $\mathbf{h} := -f'(\mathbf{z}) = \begin{pmatrix} -x \\ -ay \end{pmatrix}$. Gehen wir also von $\mathbf{z}_j = \begin{pmatrix} x_j \\ y_j \end{pmatrix}, \mathbf{h}_j = \begin{pmatrix} -x_j \\ -ay_j \end{pmatrix}$ aus, so erhalten wir

$$\alpha_j = \frac{x_j^2 + a^2 y_j^2}{x_j^2 + a^3 y_j^2}, \quad \mathbf{z}_{j+1} := \mathbf{z}_j + \alpha_j \mathbf{h}_j = \begin{pmatrix} (1-\alpha_j)x_j \\ (1-a\alpha_j)y_j \end{pmatrix}.$$

Ist $y_j = 0$ aber $x_j \neq 0$, so ist $\alpha_j = 1$ und somit $x_{j+1} = 0$, und daher ist $\mathbf{z}_{j+1} = \mathbf{0}$, die Lösung. Für $y_j \neq 0$ setzen wir $u_j = x_j/y_j$, $j = 0, 1, \ldots$ Dann ist

$$\alpha_j = \frac{u_j^2 + a^2}{u_j^2 + a^3}, \quad u_{j+1} = \frac{1-\alpha_j}{1-a\alpha_j} u_j =: q_j u_j. \tag{L18.1}$$

Daraus folgt

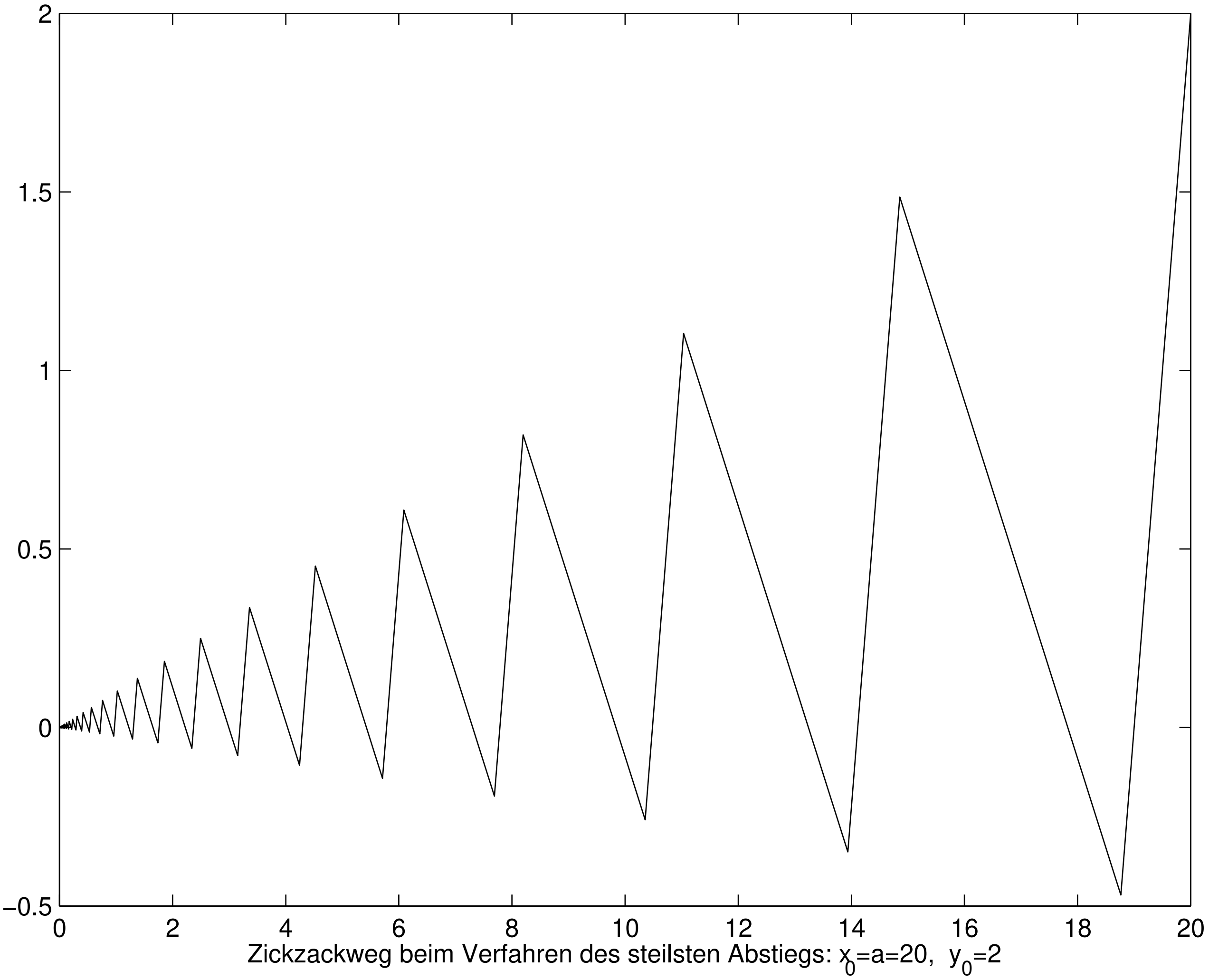
$$\alpha_{j+1} = \frac{u_{j+1}^2 + a^2}{u_{j+1}^2 + a^3} = \frac{q_j^2 u_j^2 + a^2}{q_j^2 u_j^2 + a^3}, \quad \alpha_{j+2} = \frac{u_{j+2}^2 + a^2}{u_{j+2}^2 + a^3} = \frac{q_{j+1}^2 u_{j+1}^2 + a^2}{q_{j+1}^2 u_{j+1}^2 + a^3} = \frac{q_j^2 q_{j+1}^2 u_j^2 + a^2}{q_j^2 q_{j+1}^2 u_j^2 + a^3}.$$

Nun ist (gibt es einen kürzeren Beweis?)

$$q_j q_{j+1} = q_j \frac{1-\alpha_{j+1}}{1-a\alpha_{j+1}} = q_j \frac{1 - \frac{q_j^2 u_j^2 + a^2}{q_j^2 u_j^2 + a^3}}{1 - a\frac{q_j^2 u_j^2 + a^2}{q_j^2 u_j^2 + a^3}} = q_j \frac{q_j^2 u_j^2 + a^3 - q_j^2 u_j^2 - a^2}{q_j^2 u_j^2 + a^3 - a q_j^2 u_j^2 - a^3} = q_j \frac{a^2(a-1)}{q_j^2 u_j^2 (1-a)} = -\frac{a^2}{q_j u_j^2}$$

(L18.2)

$$= -\frac{a^2(1-a\alpha_j)}{u_j^2(1-\alpha_j)} = -\frac{a^2(1 - a\frac{u_j^2+a^2}{u_j^2+a^3})}{u_j^2(1 - \frac{u_j^2+a^2}{u_j^2+a^3})} = -\frac{a^2(u_j^2 + a^3 - a(u_j^2 + a^2))}{u_j^2(u_j^2 + a^3 - u_j^2 - a^2)} = -\frac{a^2(1-a)u_j^2}{u_j^2 a^2 (a-1)} = 1.$$



Figur L18.1: Zickzackweg beim Verfahren des steilsten Abstiegs

Damit haben wir $\alpha_{j+2} = \alpha_j$ gezeigt, insbesondere ist also $u_{j+2} = u_j$ und

$$\alpha_{2j} = \alpha_0 = \frac{u_0^2 + a^2}{u_0^2 + a^3}, \quad \alpha_{2j+1} = \alpha_1 = \frac{q_0^2 u_0^2 + a^2}{q_0^2 u_0^2 + a^3}.$$

Damit ist

$$x_{j+1} = (1-\alpha_j)x_j = (1-\alpha_0)^{\lceil (j+1)/2 \rceil}(1-\alpha_1)^{\lfloor (j+1)/2 \rfloor} x_0,$$

(L18.3)

$$y_{j+1} = (1-a\alpha_j)y_j = (1-a\alpha_0)^{\lceil (j+1)/2 \rceil}(1-a\alpha_1)^{\lfloor (j+1)/2 \rfloor} y_0.$$

Wegen $q_0 q_1 = 1$ ist

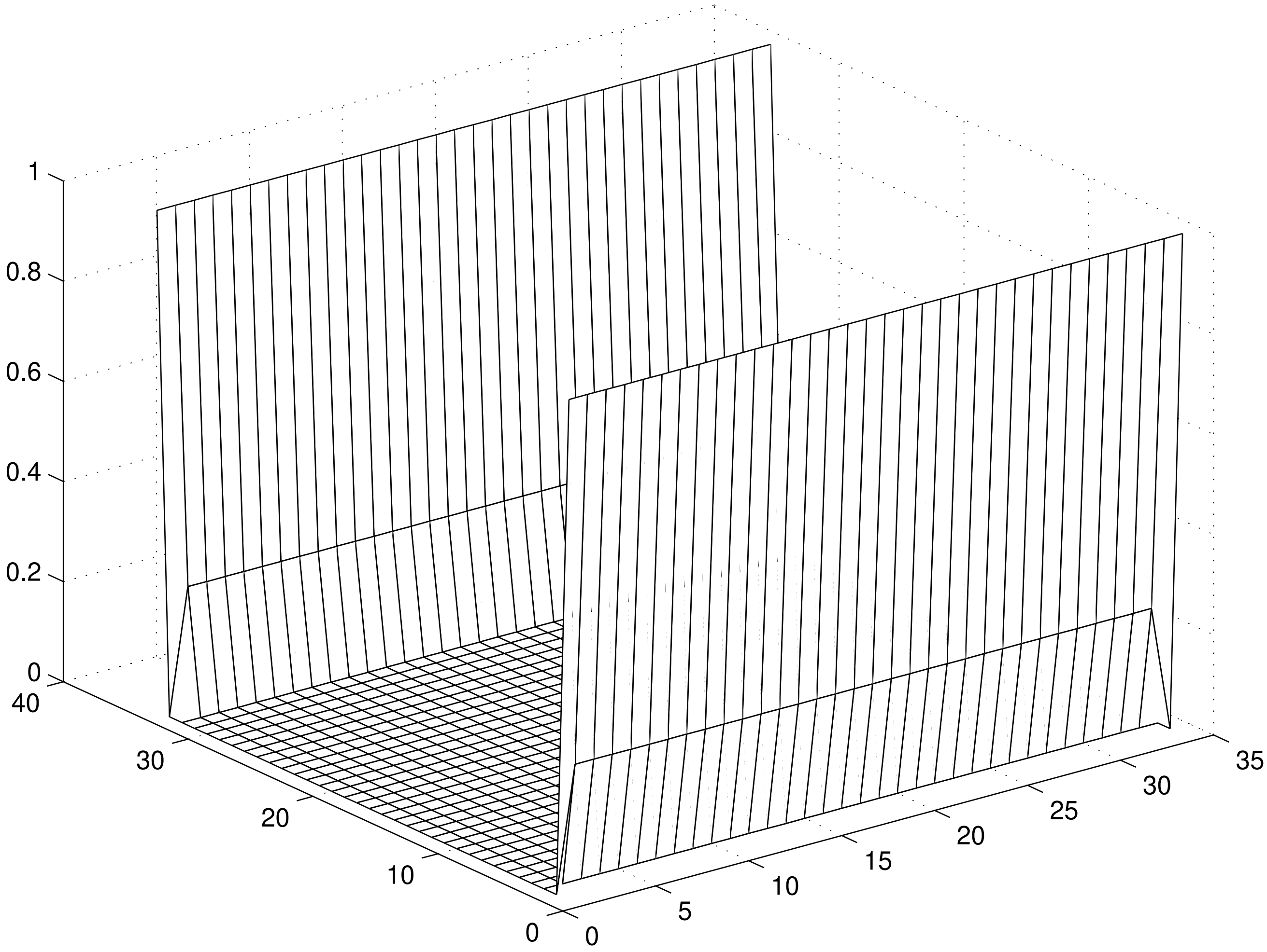
(L18.4) $$c^2 := (1-\alpha_0)(1-\alpha_1) = (1-a\alpha_0)(1-a\alpha_1).$$

Sei $a > 1$. Dann ist nach der Definition (L18.1) $0 < \alpha_j < 1$, und daher $0 < (1-\alpha_0)(1-\alpha_1) < (1-\frac{1}{a})^2 < 1$. Sei $0 < a < 1$. Dann ist wieder nach (L18.1) $0 < a\alpha_j = \frac{a u_j^2 + a^3}{u_j^2 + a^3} < 1$ und somit $0 < (1-a\alpha_0)(1-a\alpha_1) < (1-a)^2 < 1$. Insgesamt ist also

(L18.5) $$0 < c < \begin{cases} 1 - \frac{1}{a} & \text{für } a > 1, \\ 1 - a & \text{für } 0 < a < 1. \end{cases}$$

Insgesamt haben wir $\quad x_j = (1-\alpha_0)^{2\lceil j/2\rceil - j} c^j x_0, \quad y_j = (1 - a\alpha_0)^{2\lceil j/2\rceil - j} c^j y_0,$
also langsame Konvergenz für große und für kleine a. Schnellere Konvergenz liegt für a in der Nähe von Eins vor. Das Verhalten des Verfahrens des steilsten Abstiegs kann gut aus der Figur L18.1 entnommen werden. Im Anhang 1 ist ein Konvergenzbeweis für den allgemeinen Fall wiedergegeben.

Aufgabe 36. Das gesuchte Bild L18.2 für $n = 33$ wurde mit dem folgenden matlab-Programm hergestellt:



Figur L18.2: Vektor $\mathbf{y} = \mathbf{A}\mathbf{u}$.

Programm L18.3: matlab-Programme zur Herstellung der Figur L18.2

```
%Es soll das Produkt A*u ausgerechnet werden, wobei A i.w. den Fnfpunkte-
%differenzenstern darstellt, ohne da A irgendwo vorkommt.
%u ist mit Null und mit Randwerten vorbesetzt.
%Die eigentliche Produktbildung geschieht mit der Prozedur y=Amalu(u),
%die separat, in einer anderen Datei, formuliert wurde.
n=33;  u=zeros(n+1,n+1); %Totale Besetzung mit Nullen
u(1,:)=ones(size(u(1,:)));u(n+1,:)=ones(size(u(1,:))); %Besetzung der Randwerte
y=Amalh(u);   mesh(y); %Bildschirmanzeige
print -dps Amalu.ps  %bertragung in eine Datei "Amalu.ps"
```

Das obige Programm ruft eine Prozedur `Amalu` auf, die im unteren Programm bereitgestellt wird.

```
%Es soll das Produkt A*u ausgerechnet werden, wobei A i.w. den Fnfpunkte-
%differenzenstern darstellt. A wird nicht gebraucht, die Randwerte von u
%werden ungendert auf y bertragen.
function y=Amalh(u);
[z s]=size(u);    y=u;
hhh=u(1:z-2,2:s-1)+u(3:z,2:s-1)+u(2:z-1,1:s-2)+u(2:z-1,3:s);
y(2:z-1,2:s-1)=0.25*hhh-u(2:z-1,2:s-1); %Rand(y)=Rand(u)
return
```

Anhang 1: Über die langsame Konvergenz des Verfahrens des steilsten Abstiegs

Wir minimieren

$$f(\mathbf{x}) := \frac{1}{2}\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} - \mathbf{x}^{\mathrm{T}}\mathbf{b}, \quad \mathbf{A} \text{ eine SPD-Matrix}$$

mit dem beschriebenen Richtungsverfahren und wählen die Richtung des steilsten Abstiegs $\mathbf{h} := \mathbf{r} := \mathbf{b} - \mathbf{A}\mathbf{x}$. Sei $\hat{\mathbf{x}}$ die Lösung von $\mathbf{A}\mathbf{x} = \mathbf{b}$ und damit gleichzeitig die Minimalstelle von f. Wir definieren den *Fehler* von $\mathbf{x}$ bzw. von $\mathbf{x}_j$ durch

$$\mathbf{e} := \hat{\mathbf{x}} - \mathbf{x}, \text{ bzw. } \mathbf{e}_j := \hat{\mathbf{x}} - \mathbf{x}_j,\ j = 0, 1 \ldots$$

Wegen

$$g(\mathbf{x}) := \frac{1}{2}\mathbf{e}^{\mathrm{T}}\mathbf{A}\mathbf{e} =: \frac{1}{2}||\mathbf{e}||_{\mathbf{A}}^2 = f(\mathbf{x}) + \hat{\mathbf{x}}^{\mathrm{T}}\mathbf{b}$$

spielt es keine Rolle, ob wir f oder g minimieren, da der Unterschied nur die Konstante $\hat{\mathbf{x}}^{\mathrm{T}}\mathbf{b}$ ist. Nach den angegebenen Formeln (5.40)–(5.42), Buch, S. 165 ist dann

$$\text{(L18.2)} \qquad \mathbf{x}_{j+1} = \mathbf{x}_j + \alpha_j \mathbf{r}_j \text{ mit } \alpha_j = \frac{||\,\mathbf{r}_j\,||^2}{||\,\mathbf{r}_j\,||_{\mathbf{A}}^2} \text{ und } \mathbf{r}_j = \mathbf{b} - \mathbf{A}\mathbf{x}_j.$$

Definieren wir

$$\mathbf{x}(\alpha) := \mathbf{x}_j + \alpha \mathbf{r}_j \text{ mit } \alpha \in \mathbb{R},$$

so folgt wegen der Minimalität von g an der Stelle $\mathbf{x}_{j+1}$

$$\text{(L18.3)} \qquad ||\mathbf{e}_{j+1}||_{\mathbf{A}} := ||\hat{\mathbf{x}} - \mathbf{x}_{j+1}||_{\mathbf{A}} \le ||\hat{\mathbf{x}} - \mathbf{x}(\alpha)||_{\mathbf{A}} \text{ für alle } \alpha \in \mathbb{R}.$$

Nun ist

$$\text{(L18.4)} \qquad (\mathbf{I} - \alpha\mathbf{A})\mathbf{e}_j = \hat{\mathbf{x}} - \mathbf{x}(\alpha) \quad \overset{\text{wg. (L19.3)}}{\Rightarrow} \quad ||\mathbf{e}_{j+1}||_{\mathbf{A}} \le ||(\mathbf{I} - \alpha\mathbf{A})\mathbf{e}_j||_{\mathbf{A}} \le ||(\mathbf{I} - \alpha\mathbf{A})||_{\mathbf{A}}||\mathbf{e}_j||_{\mathbf{A}}.$$

Bei der letzten Ungleichung haben wir den Abschnitt 7.1 über Normen von linearen Abbildungen, insbesondere Formel (7.6), S. 185 bemüht. Ist p irgendein Polynom, $\sigma(\mathbf{A})$ das Spektrum der SPD-Matrix $\mathbf{A}$, so folgt aus dem Beweis von Satz 7.7, Teil (ii) (wir wiederholen den Beweis im Anhang 2)

$$\text{(L18.5)} \qquad ||p(\mathbf{A})||_{\mathbf{A}} = \max_{\lambda \in \sigma(\mathbf{A})} |p(\lambda)|.$$

Eine SPD-Matrix hat ein Spektrum, das nur aus positiven Zahlen besteht: $\sigma(\mathbf{A}) = \{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ mit $0 < \lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$. Wir wählen jetzt α so, daß $||(\mathbf{I} - \alpha\mathbf{A})||_{\mathbf{A}}$ möglichst klein ausfällt, also mit (L18.5)

$$\mu := \min_{\alpha \in \mathbb{R}} \max_{\lambda_1, \lambda_2, \ldots, \lambda_n} |1 - \alpha\lambda|.$$

Auf dem Intervall $[\lambda_1, \lambda_n]$ wird das Maximum von $|p(\lambda)| = |1 - \alpha\lambda|$ (als Funktion von λ) am kleinsten, wenn $|p(\lambda)|$ genau in der Intervallmitte, also bei $\lambda = 1/\alpha = (\lambda_1 + \lambda_2)/2$ eine Nullstelle besitzt. In diesem Fall ist der minimale Wert des Maximums

$$\text{(L18.6)} \qquad \mu = |p(\lambda_1)| = |p(\lambda_n)| = \frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1} \le \frac{\lambda_n - \lambda_1}{\lambda_n} = 1 - \frac{1}{\kappa(\mathbf{A})} < 1 \text{ mit } \kappa(\mathbf{A}) := \frac{\lambda_n}{\lambda_1}.$$

Die hier eingeführte Zahl $\kappa(\mathbf{A})$ ist die *Kondition* von $\mathbf{A}$, eine Zahl die man als Multiplikator der Rundungsfehler identifizieren kann. Aus (L18.4) folgt damit schließlich die Konvergenz, denn

$$\text{(L18.7)} \qquad ||\mathbf{e}_j||_{\mathbf{A}} \le \left(\frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1}\right)^j ||\mathbf{e}_0||_{\mathbf{A}} \le \left(1 - \frac{1}{\kappa(\mathbf{A})}\right)^j ||\mathbf{e}_0||_{\mathbf{A}}.$$

Ist also die Kondition $\kappa(\mathbf{A})$ groß, wird die Konvergenz sehr langsam sein, bei $\kappa(\mathbf{A}) = 1$ andererseit liegt Konvergenz bereits nach einem Schritt vor.

Anhang 2: Die Berechnung von $||p(\mathbf{A})||_{\mathbf{A}}$

Die Bestimmung von $||p(\mathbf{A})||_{\mathbf{A}}$ ist eine gute Übung zur Bestimmung einer Norm einer linearen Abbildung und zugleich ein Blick in die Welt der Eigenwerte. Sei also wieder $\mathbf{A}$ eine SPD-Matrix und p ein beliebiges Polynom. Wir wollen uns mit dem Matrixpolynom $p(\mathbf{A})$ beschäftigen. Ist beispielsweise $p(x) = 2 - x + 3x^2$, so ist $p(\mathbf{A}) = 2\mathbf{I} - \mathbf{A} + 3\mathbf{A}^2$, wobei $\mathbf{I}$ die Einheitsmatrix bedeutet. Wesentlich ist hier, daß

(L18.8) $$||\mathbf{x}||_{\mathbf{A}} := \sqrt{\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}}$$

eine Norm auf dem $\mathbb{R}^n$ definiert. (Ist $\mathbf{A}$ eine komplexe HPD-Matrix, so geht auch alles für den $\mathbb{C}^n$ durch, man vgl. Aufgabe P52). Die sogenannte Operatornorm von $||p(\mathbf{A})||_{\mathbf{A}}$ ist definiert durch

(L18.9) $$||p(\mathbf{A})||_{\mathbf{A}} := \sup_{\mathbf{x}\neq\mathbf{0}} \frac{||p(\mathbf{A})\mathbf{x}||_{\mathbf{A}}}{||\mathbf{x}||_{\mathbf{A}}} =: \sup_{\mathbf{x}\neq\mathbf{0}} Q(\mathbf{x}).$$

Unser Ziel ist der Beweis der unten angegebenen Formel L18.12. (i) Wir beweisen zuerst, daß die unten angegebene Größe eine untere Schranke von $||p(\mathbf{A})||_{\mathbf{A}}$ ist. Nach der Definition gilt $||p(\mathbf{A})||_{\mathbf{A}} \geq Q(\mathbf{x})$ für jedes $\mathbf{x} \neq 0$. Unter einem *Eigenvektor* von $\mathbf{A}$ versteht man einen Vektor $\mathbf{x} \neq 0$ mit $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$. Die Zahl λ heißt der zu $\mathbf{x}$ gehörende *Eigenwert* von $\mathbf{A}$. Ist also $\mathbf{x}$ ein Eigenvektor von $\mathbf{A}$ zum Eigenwert λ, so ist leicht einzusehen, daß $p(\mathbf{A})\mathbf{x} = p(\lambda)\mathbf{x}$. Die Menge aller Eigenwerte von $\mathbf{A}$ nennt man *Spektrum* von $\mathbf{A}$ in der Bezeichnung $\sigma(\mathbf{A})$. Daher ist

$$||p(\mathbf{A})||_{\mathbf{A}} \geq \frac{||p(\mathbf{A})\mathbf{x}||_{\mathbf{A}}}{||\mathbf{x}||_{\mathbf{A}}} = \frac{||p(\lambda)\mathbf{x}||_{\mathbf{A}}}{||\mathbf{x}||_{\mathbf{A}}} = |p(\lambda)|.$$

Da diese Ungleichung für jeden Eigenwert λ gilt, folgt

(L18.10) $$||p(\mathbf{A})||_{\mathbf{A}} \geq \max_{\lambda\in\sigma(\mathbf{A})} |p(\lambda)|.$$

(ii) Um die Umkehrung zu beweisen, verwenden wir, daß die SPD-Matrix $\mathbf{A}$ n paarweise orthonormale Eigenvektoren $\mathbf{u}_j$ zu Eigenwerten $\lambda_j > 0$ besitzt. Jedes $\mathbf{x} \neq \mathbf{0}$ hat dann die eindeutige Darstellung $\mathbf{x} = \sum_{j=1}^n \alpha_j \mathbf{u}_j$ mit $\sum_{j=1}^n |\alpha_j| > 0$. Setzen wir in L18.9 ein, so ist

(L18.11) $$||p(\mathbf{A})||_{\mathbf{A}}^2 = \sup_{(\alpha_1,\alpha_2,\ldots,\alpha_n)\neq 0} \frac{\sum_{j=1}^n |p(\lambda_j)|^2 \lambda_j \alpha_j^2}{\sum_{j=1}^n \lambda_j \alpha_j^2} \leq \max_{\lambda\in\sigma(\mathbf{A})} |p(\lambda)|^2.$$

Da rechts die α_j nicht mehr vorkommen, haben wir $||p(\mathbf{A})||_{\mathbf{A}} \leq \max_{\lambda\in\sigma(\mathbf{A})} |p(\lambda)|$ bewiesen und zusammen mit L18.10 schließlich

(L18.12) $$||p(\mathbf{A})||_{\mathbf{A}} = \max_{\lambda\in\sigma(\mathbf{A})} |p(\lambda)|.$$