Hochbruck

Numerische Mathematik 2

Dauer: 90 min. Lösung: offiziell

Aufgabe 1: (weiß)

Gegeben sei die Funktion $f\colon \mathbb{R}^3 \to \mathbb{R}^3$ durch

$$f(x) = \frac{1}{4}\begin{bmatrix} \sin(x_1)+\cos(x_3)+2 \\ \sin(2x_1)-\sin(x_2)+\sin(x_3)-1 \\ \cos(x_2)\cos(x_3)-2 \end{bmatrix}, \qquad x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.$$

a) Berechnen Sie die Jacobi-Matrix J_f von f und zeigen Sie für alle $x \in \mathbb{R}^3$

$$\|J_f(x)\|_\infty \le 1 \quad \text{und} \quad \|J_f(x)\|_1 \le \tfrac{3}{4}.$$

b) Beweisen Sie, dass f einen eindeutig bestimmten Fixpunkt $\widehat{x} \in \Omega = [-1,1]^3$ besitzt.

c) Für welche Startwerte $x^0 \in \mathbb{R}^3$ konvergiert die Fixpunktiteration gegen $\widehat{x}$?

d) Wählen Sie eine geeignete Norm $\|\cdot\|$ und bestimmen Sie für $x^0 = 0$ und $p \in \mathbb{N}$ ein $k = k(p) \ge 0$, sodass

$$\|\widehat{x} - x^k\| \le 5 \cdot 10^{-p}$$

gilt, wobei $x^k \in \mathbb{R}^3$ die k-te Iterierte der Fixpunktiteration bezeichnet.

Aufgabe 2: (gelb)

a) Sei $A \in \mathbb{R}^{n,n}$ symmetrisch und positiv definit sowie $D = \operatorname{diag}(\sigma_i) \in \mathbb{R}^{n,n}$, $\sigma_i \in \{0,1\}$. Zeigen Sie: Für den Wertebereich von ADA gilt

$$\mathcal{F}(ADA) \subseteq [0, \lambda_{\max}(A)^2],$$

wobei mit $\lambda_{\max}(A)$ der größte Eigenwert von A bezeichnet wird.

b) Gegeben sei die Matrix

$$S = \begin{bmatrix} 6i & -2 & 0 \\ 2 & 0 & 1 \\ 0 & -1 & -5i \end{bmatrix}.$$

α) Berechnen und skizzieren Sie mit Hilfe des Satzes von Gershgorin und Symmetrieeigenschaften von S eine möglichst kleine Menge $\mathcal{M}$, die alle Eigenwerte von S enthält.

β) Was können Sie über die Vielfachheit der Eigenwerte sagen?

γ) Wir betrachten die Potenzenmethode mit einem Startvektor, der zu keinem Eigenvektor von S orthogonal ist. Können Sie aus der Menge $\mathcal{M}$ die Konvergenz der Potenzenmethode schließen? Begründen Sie Ihre Antwort.

Aufgabe 3: (rosa)

Es sei $A \in \mathbb{C}^{n,n}$ eine reguläre Matrix und $b \in \mathbb{C}^n$. Wir betrachten die näherungsweise Lösung des Gleichungssystems

$$Ax = b$$

mit einem vorkonditionierten GMRES-Verfahren, d. h., wir wenden das GMRES-Verfahren auf $BAx = Bb$ mit einer regulären Matrix $B \in \mathbb{C}^{n,n}$, $B \approx A^{-1}$, an.

Der folgende Pseudocode gibt das GMRES-Verfahren ohne Vorkonditionierung an:

1: Wähle $x_0 \in \mathbb{C}^n$, $\texttt{tol} > 0$ und setze $r_0 = b - Ax_0$
2: Initialisiere $m = 0$, $\beta = \rho_0 = \|r_0\|$, $v_1 = r_0/\beta$
3: **while** $\rho_m > \texttt{tol}$ **do**
4:     $m = m + 1$
5:     **for** $j = 1, \dots, m$ **do**
6:         $h_{j,m} = v_j^H A v_m$
7:     **end for**
8:     $\widetilde{v}_{m+1} = Av_m - \sum_{j=1}^m h_{j,m} v_j$
9:     $h_{m+1,m} = \|\widetilde{v}_{m+1}\|$
10:     $v_{m+1} = \widetilde{v}_{m+1}/h_{m+1,m}$
11:     Berechne y_m als Lösung von $\|\beta e_1 - \widetilde{H}_m y\| = \min$
12:     $\rho_m = \|\beta e_1 - \widetilde{H}_m y_m\|$
13: **end while**
14: Setze $x_m = x_0 + V_m y_m$

a) Modifizieren Sie den angegebenen Algorithmus so, dass das vorkonditionierte Problem gelöst wird. Geben Sie hierzu die veränderten Zeilen (vollständig) unter Angabe der Zeilennummer an.

b) Wie viele Matrix-Vektormultiplikationen mit A und mit B sind für k Schritte des Verfahrens bei effizienter Implementierung notwendig? Begründen Sie Ihr Ergebnis durch Angabe der Zeilennummern.

c) Sei x_m die Näherungslösung nach m Schritten des vorkonditionierten GMRES-Verfahrens mit $x_0 = 0$. Zeigen Sie, dass es einen Untervektorraum $\mathcal{K}_m$ gibt, sodass

$$\|B(b - Ax_m)\| \leq \|B(b - Ax)\| \qquad \text{für alle } x \in \mathcal{K}_m$$

gilt und geben Sie $\mathcal{K}_m$ an.

d) Zeigen Sie: Das vorkonditionierte GMRES-Verfahren mit Vorkonditionierungsmatrix $B = A^{-1}$ liefert die exakte Lösung nach einem Schritt.

Aufgabe 4: (grün)

Gegeben sei die Funktion $f\colon [0,2\pi] \to \mathbb{R}$, definiert durch

$$f(\xi) = 2\cos(24\,\xi) - \cos(\xi),$$

sowie für ein festes $N \in \mathbb{N}$ die Stützstellen $\xi_j = \frac{2\pi j}{N}$, $j = 0,1,\dots,N-1$.

a) Zeigen Sie, dass für die Fourierkoeffizienten von f gilt

$$\widehat{f}(\pm 1) = -\frac{1}{2}, \quad \widehat{f}(\pm 24) = 1, \quad \widehat{f}(k) = 0, \quad k \in \mathbb{Z} \setminus \{\pm 1, \pm 24\}.$$

b) Bestimmen Sie für $N = 20$ die diskreten Fourierkoeffizienten $\widehat{f}_N(k)$ des Vektors $\big(f(\xi_j)\big)_{j=0}^{N-1}$ für $k = -N/2, -N/2+1, \dots, N/2-1$.

c) Bestimmen Sie für $N = 20$ das *reelle* trigonometrische Interpolationspolynom t_N zu den Daten $\big(\xi_j, f(\xi_j)\big)$, $j = 0,1,\dots,N-1$.

d) Wie groß muss N gewählt werden, sodass das trigonometrische Interpolationspolynom t_N aus Aufgabenteil c) mit der Funktion f übereinstimmt?

Lösung von Aufgabe 1:

a) (3P; J_f, Normen) Die Jacobimatrix ist gegeben durch

$$J_f(x) = \frac{1}{4}\begin{pmatrix} \cos(x_1) & 0 & -\sin(x_3) \\ 2\cos(2x_1) & -\cos(x_2) & \cos(x_3) \\ 0 & -\sin(x_2)\cos(x_3) & -\cos(x_2)\sin(x_3) \end{pmatrix}.$$

Für die Zeilen- bzw. Spaltensummennorm erhält man mit $|\sin(\xi)|, |\cos(\xi)| \leq 1$ für $\xi \in \mathbb{R}$ die Abschätzungen

$$\|J_f(x)\|_\infty \leq \frac{1}{4}\max\{2,4,2\} = 1, \quad \|J_f(x)\|_1 \leq \frac{1}{4}\max\{3,2,3\} = \frac{3}{4}.$$

b) (3P; Banach: abgeschlossen, Selbstabbildung, kontraktiv in $\|\cdot\|_1$)

Wir verwenden den Banach'schen Fixpunktsatz und zeigen dafür

α) Selbstabbildung: Für die Komponenten f_i von f gilt

$$|f_i(x)| \leq 1, \qquad i = 1,2,3 \implies f(x) \in \Omega \text{ für alle } x \in \mathbb{R}^3,$$

β) Abgeschlossenheit von Ω: klar,

γ) Kontraktion bzgl. $\|\cdot\|_1$: Gilt nach Teil a).

Damit existiert ein eindeutiger Fixpunkt $\hat{x} \in \Omega$.

c) (1P) <u>Variante 1:</u> Wir können den Banach'schen Fixpunktsatz auch auf $\mathbb{R}^3$ anwenden und erhalten auch hier einen eindeutigen Fixpunkt und die Fixpunktiteration konvergiert für jeden Startwert. Da dieser nach Teil b) in Ω liegen muss, erhalten wir Konvergenz gegen $\hat{x}$ für alle $x \in \mathbb{R}^3$.

<u>Variante 2:</u> In Teil b) haben wir gesehen, dass für beliebiges x^0, der Wert $x^1 = f(x^0) \in \Omega$ erfüllt. Mit dem Startwert x^1 erhalten wir aber aus Teil b) die Konvergenz der Fixpunktiteration gegen $\hat{x}$, sodass wir $x^0 \in \mathbb{R}^3$ beliebig wählen können.

d) (3P) Wir verwenden die a-priori-Abschätzung des Banach'schen Fixpunktsatzes in der $\|\cdot\|_1$-Norm:

$$\left\|\hat{x} - x^k\right\|_1 \leq \frac{L^k}{1-L}\left\|x^1 - x^0\right\|_1.$$

Dazu berechnen wir die erste Iterierte

$$x_1 = f(x_0) = \frac{1}{4}\begin{bmatrix} 3 \\ -1 \\ -1 \end{bmatrix}$$

und damit

$$\left\|x^1 - x^0\right\|_1 = \left\|x^1\right\|_1 = \frac{5}{4},$$

sodass sich mit $L = \frac{3}{4}$ die Schranke

$$\left\|\hat{x} - x^k\right\|_1 \leq 4\left(\frac{3}{4}\right)^k\frac{5}{4} = 5\left(\frac{3}{4}\right)^k$$

ergibt. Damit ist

$$\left\|\hat{x} - x^k\right\|_1 \leq 5 \cdot 10^{-p}$$

erfüllt, falls

$$\left(\frac{3}{4}\right)^k \leq 10^{-p} \iff -k\log_{10}\frac{4}{3} \leq -p \iff k \geq \frac{p}{\log_{10}\frac{4}{3}}$$

gilt.

Lösung von Aufgabe 2:

a) (3P; symmetrisch, ≥ 0, $\|A\|^2$)

Da A und D symmetrisch sind, folgt $(ADA)^T = ADA$, also ist ADA symmetrisch. Damit folgt aus dem Satz von Rayleigh-Ritz

$$\mathcal{F}(ADA) = [\lambda_{\min}(ADA), \lambda_{\max}(ADA)] \subseteq \mathbb{R}.$$

Ferner gilt für $x \in \mathbb{C}^n$ mit $\|x\| = 1$ wegen $D^2 = D$ und $D = D^H$

$$\rho_{ADA}(x) = x^H ADAx = x^H ADDAx = (DAx)^H (DAx) = \|DAx\|^2 \geq 0.$$

Daher ist ADA positiv semidefinit und es gilt $\lambda_{\min}(ADA) \geq 0$.
Alternativer Beweis für ADA positiv semidefinit: Wegen $ADA = A^T DA$ und A nicht singulär (da A positiv definit ist), folgt aus dem Trägheitssatz, dass $\text{in}(ADA) = \text{in}(D)$. Damit hat ADA keine negativen Eigenwerte.

Weil ADA symmetrisch und positiv semidefinit und $\|D\| \leq 1$ ist, folgt

$$0 \leq \lambda_{\max}(ADA) = \|ADA\| \leq \|A\|^2 \|D\| \leq \|A\|^2 = \lambda_{\max}(A)^2$$

und daraus die Behauptung.

b) α) (4P; 0,5P für jeden Kreis, 1P für schiefsymmetrisch, 1,5P Skizze)
Wir definieren für $r > 0$ und $a \in \mathbb{C}$ $\quad B_r(a) = \{z \in \mathbb{C} \,:\, |z - a| \leq r\}$.

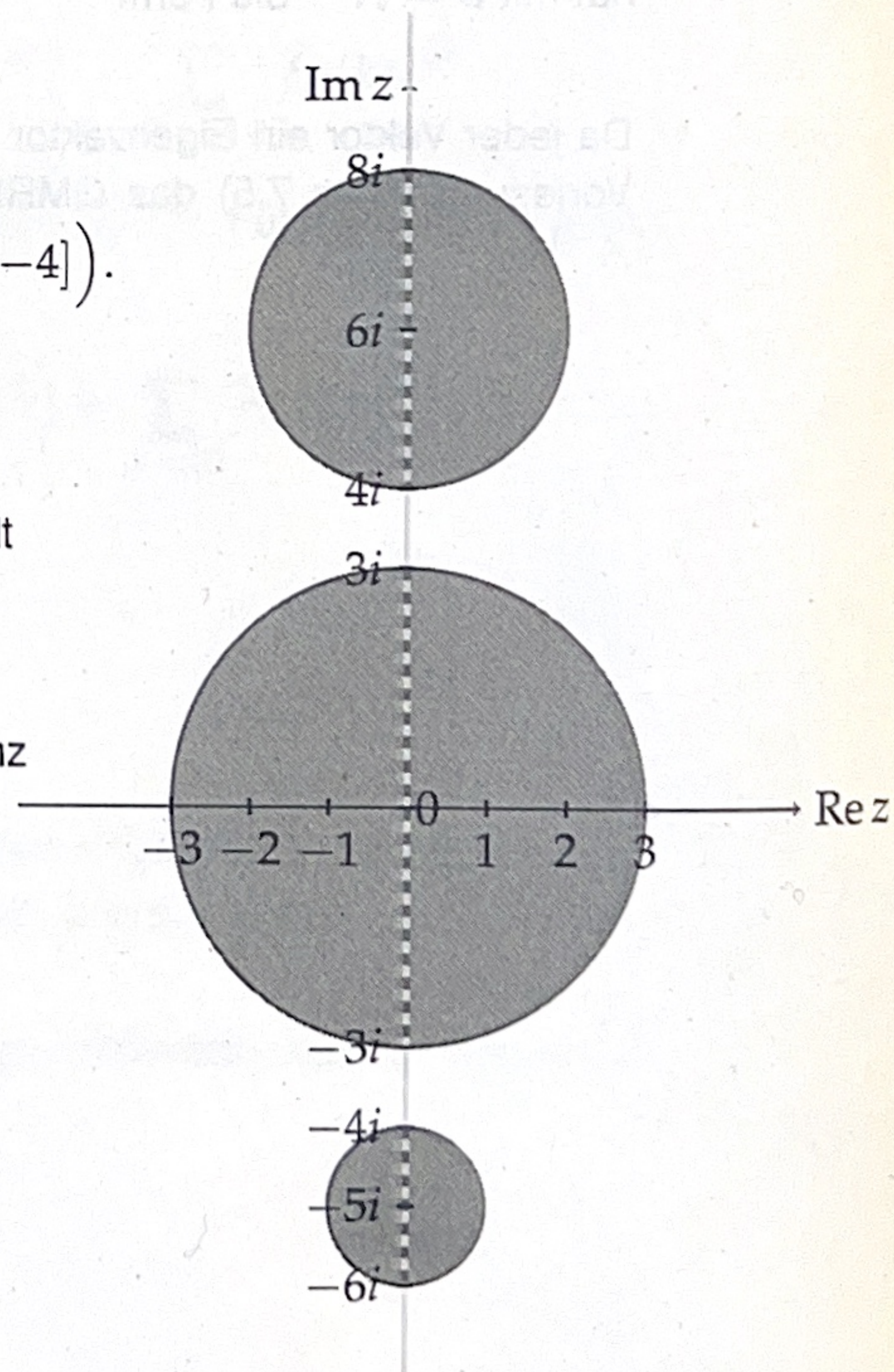
Mit dem Satz von Gershgorin gilt $\quad \lambda(S) \subseteq M_1 = B_2(6i) \cup B_3(0) \cup B_1(-5i)$.

Da außerdem $S = -S^H$ gilt, erhalten wir

$$\lambda(S) \subseteq i\mathbb{R}.$$

Durch den Schnitt der beiden Mengen ergibt sich
$\lambda(S) \subseteq M_3 = M_1 \cap i\mathbb{R} = i\left([4,8] \cup [-3,3] \cup [-6,-4]\right)$.

Skizze der Mengen M_1, $i\mathbb{R}$ und M_3 (gepunktet):



β) (1P) Da alle 3 Gershgorin-Kreise disjunkt sind, enhält jeder Kreis genau einen Eigenwert und somit sind alle Eigenwerte einfach.

γ) (2P) Aus der Menge $\mathcal{M}$ kann man keine Konvergenz der Potenzenmethode schließen, da beispielsweise $|\lambda_1| = |\lambda_2| \in [4,6]$ gelten könnte und man daraus nicht $\eta = \frac{|\lambda_2|}{|\lambda_1|} < 1$ folgern kann.

Lösung von Aufgabe 3:

a) (3P) Es müssen folgende Modifikationen vorgenommen werden:

- Zeile 1: Wähle $x_0 \in \mathbb{C}^n$, setze $r_0 = B(b - Ax_0)$,
- Zeile 6: $h_{j,m} = v_j^H B A v_m$,
- Zeile 8: $\widetilde{v}_{m+1} = BAv_m - \sum_{j=1}^{m} h_{j,m} v_j$.

b) (3P) Für die einzelnen Zeilen gilt:

- Zeile 1: Je eine Multiplikation mit A und B (da B ausgeklammert wurde).
- Zeile 6: Je eine Multiplikation mit A und B und Speichern von BAv_m in temporärem Vektor, dieser muss somit nur einmal vor der for-Schleife berechnet werden.
- Zeile 8: BAv_m aus temporärem Vektor kann wiederverwendet werden, somit ist keine weitere Multiplikation nötig.

Da die while-Schleife bei k Schritten k-mal durchlaufen wird, ergibt sich damit insgesamt ein Aufwand von jeweils $k+1$ Multiplikationen mit A und B (einmal vor der while-Schleife und danach einmal pro Iteration).

c) (2P) Nach Vorlesung (Satz 7.6) minimiert das GMRES-Verfahren angewendet auf $BAx = Bb$ die Norm des Residuums $\|Bb - BAx_m\| = \|B(b - Ax_m)\|$ über den Krylovraum $\mathcal{K}_m = \mathcal{K}_m(BA, Bb)$.

d) (2P) Die vorkonditionierte Gleichung

$$BAx = Bb$$

hat mit $B = A^{-1}$ die Form

$$Ix = A^{-1}b.$$

Da jeder Vektor ein Eigenvektor von I ist, ist $\mathcal{K}_1(I, B(b - Ax_0))$ invariant. Somit liefert nach Vorlesung (Satz 7.5) das GMRES-Verfahren nach einem Schritt die exakte Lösung $x_1 = A^{-1}b$.

Lösung von Aufgabe 4:

a) (2P) Da f stetig differenzierbar und 2π-periodisch, stimmt f mit seiner Fourierreihe überein. Wegen $\cos(t) = \frac{1}{2}(e^{it} + e^{-it})$ für $t \in \mathbb{R}$ gilt

$$f(\xi) = 2 \cdot \tfrac{1}{2}\left(e^{24i\xi} + e^{-24i\xi}\right) - \tfrac{1}{2}\left(e^{i\xi} + e^{-i\xi}\right) = e^{24i\xi} + e^{-24i\xi} - \tfrac{1}{2}e^{i\xi} - \tfrac{1}{2}e^{-i\xi},$$

woraus mit der Definition der Fourierreihe $\sum_{k\in\mathbb{Z}} \widehat{f}(k)e^{ik\xi}$ mittels Koeffizientenvergleich folgt, dass

$$\widehat{f}(\pm 1) = -\frac{1}{2}, \qquad \widehat{f}(\pm 24) = 1, \qquad \widehat{f}(k) = 0, \quad k \neq \pm 1, \pm 24.$$

b) (4P) Die Berechnung der diskreten Fourierkoeffizienten erfolgt mithilfe des Aliasing-Theorems:

$$\widehat{f}_N(k) = \sum_{j\in\mathbb{Z}} \widehat{f}(k + jN) \qquad \text{für } k \in \{-N/2, -N/2+1, \dots, N/2-1\}.$$

Für $N = 20$ gilt dann mit Aufgabenteil a)

$$\widehat{f}_{20}(\pm 1) = \widehat{f}(\pm 1) = -\frac{1}{2} \qquad \text{und} \qquad \widehat{f}_{20}(\pm 4) = \widehat{f}(\pm 24) = 1,$$

da $\widehat{f}(k) = 0$ für alle $k \notin \{\pm 1, \pm 24\}$. Mit der gleichen Argumentation folgt $\widehat{f}_{20}(k) = 0$ für $k \notin \{-4, -1, 1, 4\}$.

c) (3P) Für N gerade ist ein trigonometrisches Interpolationspolynom gegeben durch

$$t_N(\xi) = \frac{1}{2}\left(\widehat{f}_N(-N/2)e^{-iN\xi/2} + \widehat{f}_N(N/2)e^{iN\xi/2}\right) + \sum_{|k|<N/2} \widehat{f}_N(k)e^{ik\xi},$$

wobei $\widehat{f}_N(k)$ periodisch fortgesetzt wird ($\widehat{f}_N(N/2) = \widehat{f}_N(-N/2)$). Für $N = 20$ und aus Aufgabenteil b) ergibt sich damit

$$\begin{aligned} t_{20}(\xi) &= \frac{1}{2}\left(\widehat{f}_{20}(-10)e^{-20i\xi/2} + \widehat{f}_{20}(10)e^{20i\xi/2}\right) + \sum_{|k|<10} \widehat{f}_{20}(k)e^{ik\xi} \\ &= -\frac{1}{2}(e^{i\xi} + e^{-i\xi}) + (e^{4i\xi} + e^{-4i\xi}) \\ &= -\cos(\xi) + 2\cos(4\xi). \end{aligned}$$

Insbesondere folgt hieraus direkt $f([0, 2\pi]) \subset \mathbb{R}$.

d) (1P) Nach der Vorlesung stimmt eine beliebige 2π-periodische Funktionen $f\colon [0, 2\pi] \to \mathbb{C}$ mit $\widehat{f}(k) = 0$ für $|k| > K$, $K \in \mathbb{N}$, mit dem trigonometrischen Interpolationspolynom überein, wenn $N > 2K$. Da hier $K = 24$, gilt also $N > 48$.

Anmerkung: Für die hier gegebene Funktion stimmt das oben definierte trigonometrische Interpolationspolynom bereits für $N = 48$ überein.