



Persönlicher Sticker



S0496

**Bestätigung der Verhaltensregeln**

Hiermit versichere ich, dass ich diese Klausur ausschließlich unter Verwendung der unten aufgeführten Hilfsmittel selbst löse und unter meinem Namen abgabe.

Unterschrift oder vollständiger Name, falls keine Stifteingabe verfügbar

## Numerisches Programmieren

**Klausur:** IN0019 / Endterm

**Datum:** Mittwoch, 2. März 2022

**Prüfer:** Prof. Dr. Hans-Joachim Bungartz

**Uhrzeit:** 14:15 – 15:45

### Bearbeitungshinweise

- Diese Klausur umfasst **14 Seiten** mit insgesamt **5 Aufgaben**.  
Bitte kontrollieren Sie jetzt, dass Sie eine vollständige Angabe erhalten haben.
- Die Gesamtpunktzahl in dieser Klausur beträgt 55 Punkte.
- Das Heraustrennen von Seiten aus der Prüfung ist untersagt.
- Als Hilfsmittel sind zugelassen:
  - alle Hilfsmittel, insbesondere Bücher, persönliche Notizen, Internetsuchmaschinen, selbst erstellte Skripte und Programmcodes.
  - nicht erlaubt sind Hilfestellungen von Dritten oder Kommunikation mit Dritten.
  - nicht erlaubt sind Plagiate jeder Art.
- Mit \* gekennzeichnete Teilaufgaben sind ohne Kenntnis der Ergebnisse vorheriger Teilaufgaben lösbar.
- **Es werden nur solche Ergebnisse gewertet, bei denen der Lösungsweg erkennbar ist.** Auch Textaufgaben sind **grundsätzlich zu begründen**, sofern es in der jeweiligen Teilaufgabe nicht ausdrücklich anders vermerkt ist.
- Schreiben Sie weder mit roter / grüner Farbe noch mit Bleistift.
- Beachten Sie die TUMExam Empfehlungen zur Abgabe von PDFs.



Klausur leer





## Aufgabe 1 Gleitkommazahlen und Kondition (10 Punkte)

Wir betrachten Gleitkommazahlen mit geringer Präzision. Eine Gleitkommazahl wird mit 16 Bits dargestellt, wobei ein Bit das Vorzeichen (S), acht Bit den Exponenten (E) und sieben Bits die Mantisse kodieren:

S	E	E	E	E	E	E	E	M	M	M	M	M	M	M	M
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Zahldarstellung und Rundung wird wie für IEEE754 double und single precision Gleitkommazahlen definiziert. Der Bias für den Exponenten ist 127.

0

0	1	2
0	1	2

a) Stellen Sie die Zahl  $\frac{19}{6}$  als Zahl im Binärsystem dar.

0

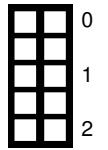
0	1	2	3
0	1	2	3

b) Runden Sie die Zahl die Zahl  $\frac{19}{6}$  korrekt in unser Gleitkommazahlsystem und geben Sie die Bitdarstellung an.

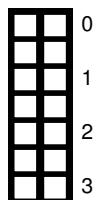




c)\* Finden Sie die kleinste positive Ganzahl, die nicht exakt von unserem Gleitkommazahlsystem dargestellt werden kann.



d)\* Gegeben sei die Funktion  $f(x) = \frac{1-x}{1+x}$ ,  $x > 0$ . Bestimmen Sie die relative Konditionszahl  $\text{cond}(f, x)$  der Funktion  $f$ .





## Aufgabe 2 Interpolation (13 Punkte)

Wir möchten eine Funktion mit kubischen Splines approximieren. Gegeben sind die Stützstellen  $x_0 = 0, x_1 = 1, x_2 = 2$  und die Funktionswerte  $y_0 = 4, y_1 = 5, y_2 = 8$ . Außerdem sind die Randbedingungen  $s'_0 = -1$  und  $s'_2 = 1$  gegeben.

- a) Berechnen Sie die fehlende Ableitung  $s'_1$ .

0			
1			
2			

- b)\* Beschreiben Sie allgemein, wie man die Spline Funktion an einer Stelle  $x_0 < \hat{x} < x_2$  auswertet.

0			
1			
2			

- c)\* Werten Sie den Spline Funktion mit den oben gegebenen Stützstellen an der Stelle  $\hat{x} = 1.5$  aus. Falls Sie die Aufgabe a) nicht gelöst haben, verwenden Sie  $s'_1 = 2$ .

0			
1			
2			



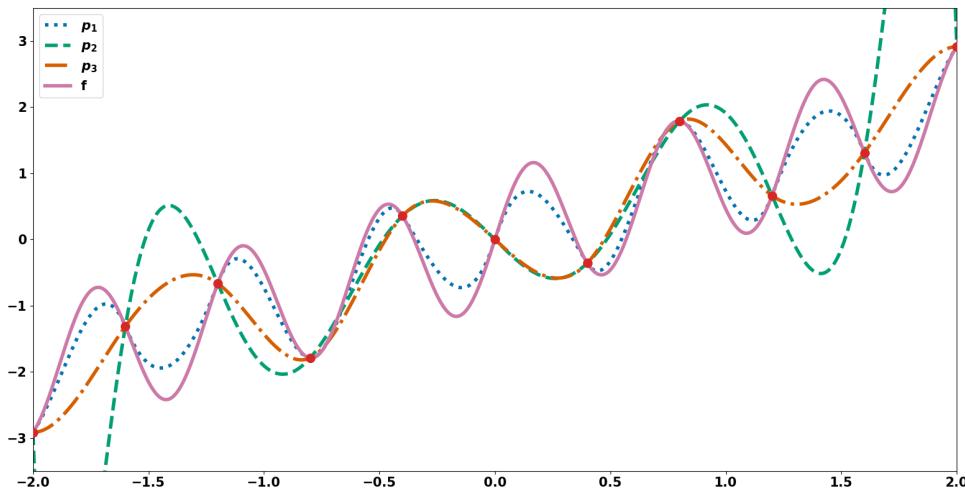
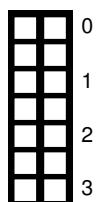


Abbildung 2.1: Die Funktion  $f$  und drei verschiedene Interpolationsfunktionen, sowie die Interpolationspunkte

d)\* Gegeben Sei die Funktion  $f(x) = x + \sin(10x)$ . Gegeben sind außerdem die Stützstellen  $x_i = -2, -1.6, -1.2, \dots, 1.6, 2.0$ . Wir interpolieren diese Funktion mit drei verschiedenen Methoden:

- Polynominterpolation auf den Stützstellen  $x_i$
- Kubische Splines mit  $s'(x_0) = s'(x_n) = 0$
- Stückweise Hermite Interpolation

In Abbildung 2.1 sind die Funktion  $f$ , sowie die drei Interpolationsfunktionen  $p_1$ ,  $p_2$  und  $p_3$  abgebildet. Geben Sie an, welcher Plot zu welcher Methode gehört und begründen Sie Ihre Wahl kurz.





0  
1  
2  
3  
4

e)\* Wir wollen nun den Rechenaufwand für die Interpolation mit kubischen Splines und für die stückweise Hermite Interpolation analysieren. Nehmen Sie an, es sind  $n$  paarweise verschiedene Stützstellen  $x_0, \dots x_{n-1}$  gegeben, sowie die Funktionswerte  $y_i$  und die Werte der Ableitungen  $y'_i$ .

Analysieren Sie wie viele Gleitkommaoperationen ( $+, -, \cdot, \div$ ) Sie benötigen um die Interpolationsfunktionen aufzustellen. Geben Sie darüber hinaus an, wie viele Gleitkommaoperationen Sie benötigen um die Interpolationsfunktion an einer Stelle auszuwerten.

Sie müssen keine exakte Anzahl angeben, eine Abschätzung mit  $\mathcal{O}$  genügt. Begründen Sie Ihre Antwort kurz. Sie können davon ausgehen, dass die Stützstellen aufsteigend geordnet sind.



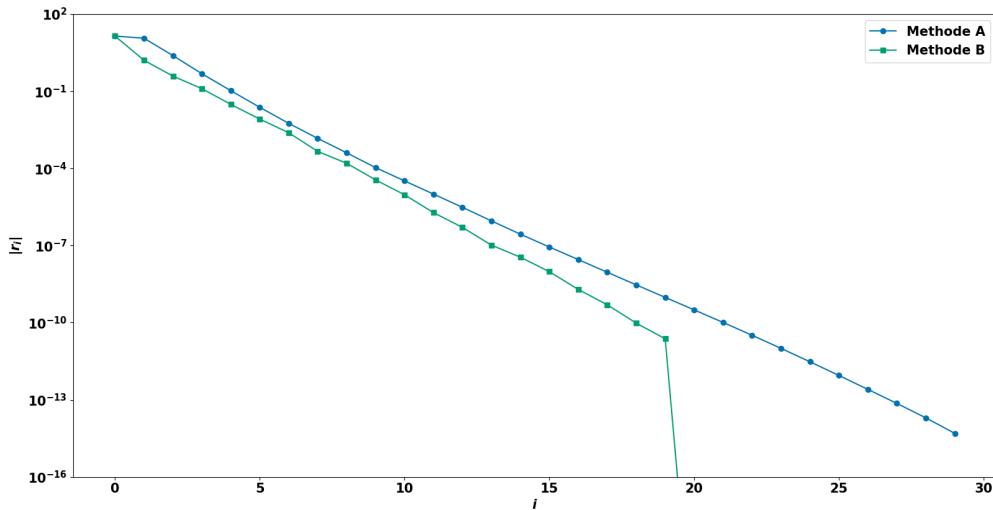


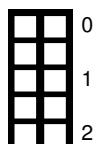
Abbildung 3.1: Residuen der verschiedenen Methoden

### Aufgabe 3 Iterative Lösung linearer Gleichungssysteme (7 Punkte)

Wir betrachten das lineare Gleichungssystem  $Ax = b$  mit  $A \in \mathbb{R}^{20 \times 20}$  und  $b \in \mathbb{R}^{20}$ :

$$A = \begin{pmatrix} 4 & 1 & 0 & 0 & \cdots & 0 \\ 1 & 4 & 1 & 0 & \cdots & 0 \\ 0 & 1 & 4 & 1 & \cdots & 0 \\ 0 & 0 & 1 & 4 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 4 \end{pmatrix}$$

- a) Wir lösen das Gleichungssystem  $Ax = b$  mit dem CG Verfahren und dem Gauss-Seidel Verfahren. In Abbildung 3.1 sind die Residuen der zwei Methoden nach der  $i$  ten Iteration angegeben. Ordnen Sie die Verfahren dem jeweiligen Plot zu und begründen Sie ihre Wahl kurz.





0  
1  
2  
3

b)\* Finden Sie eine obere Schranke für die Kondition  $\kappa_2(A)$

0  
1  
2

c)\* Die Methoden in Teil a) sind alle (semi-)iterative Verfahren. Wäre es sinnvoll, ein direktes Verfahren anzuwenden? Begründen Sie Ihre Wahl kurz mit einem Vorteil (bzw. einen Nachteil).





## Aufgabe 4 Vektoriteration und Nullstellen von Polynomen (16 Punkte)

Für jedes Polynom  $p(x) = x^n + a_{n-1}x^{n-1} + a_1x + a_0$  existiert die Begleitmatrix  $A_p \in \mathbb{R}^{n \times n}$  mit

$$A_p = \begin{pmatrix} 0 & 0 & \dots & 0 & -a_0 \\ 1 & 0 & \dots & 0 & -a_1 \\ 0 & 1 & \dots & 0 & -a_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & -a_{n-1} \end{pmatrix}. \quad (4.1)$$

Die Eigenwerte dieser Matrix sind genau die Nullstellen von  $p$ . Um die betragsmäßig größte Nullstelle zu bestimmen, verwenden wir die Vektoriteration:

```
double groessteNullstelle = vektorIteration(begleitmatrix(polynomialKoeffizienten), x_0);
```

Hinweis: Die Vektoriteration funktioniert auch für nicht-symmetrische Matrizen. Dies können Sie als gegeben annehmen und müssen es nicht weiter begründen.

Matrizen speichern wir als Arrays double[][], wobei  $A_{ij}$  als  $a[i][j]$  gespeichert wird. Vektoren sind als Arrays double[] realisiert. Sie können die Funktionen

- public static double[] matrixVektorMultiplikation(double[][] matrix, double[] vektor)
- public static double skalarprodukt(double[] x, double[] y)
- public static double norm(double[] x)

verwenden ohne diese selbst implementieren zu müssen.

a) Füllen Sie das untenstehende Codegerüst aus, um die Matrix  $A_p$  zu bestimmen.

```
public static double[][] begleitmatrix(double[] polynomKoeffizienten) {
```

}

	0
	1
	2
	3





0  
1  
2  
3

b)\* Als nächstes implementieren wir das Abbruchkriterium  $\|w^{(k)} - \lambda^{(k)}x^{(k)}\| \leq \epsilon \|w^{(k)}\|$ . Füllen Sie das untenstehende Codegerüst aus, um zu überprüfen, ob  $w^{(k)}, x^{(k)}, \lambda^{(k)}$  das Abbruchkriterium erfüllen. Geben Sie true zurück, wenn die Approximation des Eigenwertes gut genug ist. Verwenden Sie  $\epsilon = 10^{-8}$ .

```
public static boolean abbruchKriterium(double[] w, double[] x, double lambda) {
```

0  
1  
2  
3  
4  
5

c)\* Füllen Sie das untenstehende Codegerüst aus, um eine Vektoriteration zur Bestimmung des betragsmäßig größten Eigenwerts einer Matrix durchzuführen. Es sollen dabei maximal 10.000 Iterationen ausgeführt werden. Verwenden Sie das Abbruchkriterium aus Aufgabe b).

```
public static double vektorIteration(double[][] matrix, double[] x_0) {
```

}





d)\* Für die Vektoriteration benötigen wir keinen expliziten Zugriff auf die Elemente der Matrix  $A_p$ . Anstatt die Matrix  $A_p$  aufzustellen und dann in `vektorIteration` die generische Matrix-Vektor-Multiplikation zu verwenden, können wir eine Funktion verwenden, die direkt die Multiplikation  $A_p v$  auswertet: Füllen Sie das untenstehende Codegerüst aus, um das Ergebnis der Matrix-Vektor-Multiplikation  $A_p \cdot v$  zu bestimmen. Stellen Sie die Matrix  $A_p$  nicht explizit auf.

```
public static double[] multipliziereMitBegleitmatrix(double[] polynomKoeffizienten, double[] vektor) {
```

```
}
```

e)\* Analysieren Sie die Anzahl an Gleitkommaoperationen ( $+, -, \cdot, \div$ ), die für `multipliziereMitBegleitmatrix` nötig sind und vergleichen Sie Ihr Ergebnis mit der generischen Matrix-Vektor-Multiplikation. Eine Abschätzung mit  $\mathcal{O}$  genügt.

0
1
2
3





## Aufgabe 5 Quadratur in der Wahrscheinlichkeitstheorie (9 Punkte)

Ein großer Anwendungsschwerpunkt der Quadratur ist die Wahrscheinlichkeitstheorie. Hier ist die Berechnung des Erwartungswertes eines der zentralen Elemente. Häufig wird betrachtet wie sich das Ergebnis einer Funktion verhält, wenn die Eingabewerte  $x$  einer Wahrscheinlichkeitsdichte  $p(x)$  folgen. Der Erwartungswert  $E[f]$  kann dann wie folgt berechnet werden:

$$E[f] = \int_D f(x) \cdot p(x) dx. \quad (5.1)$$

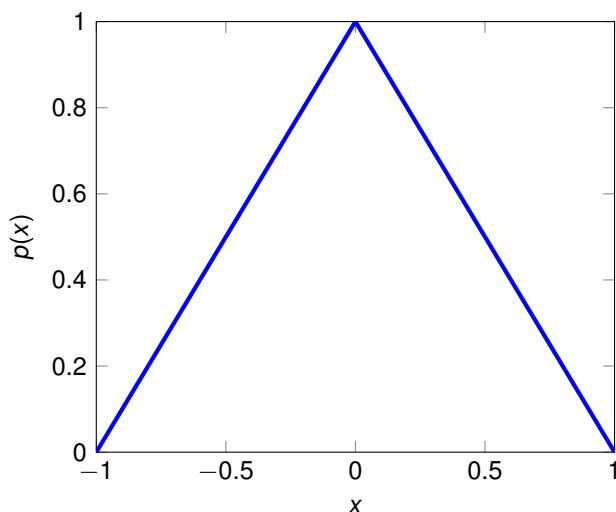
Folglich ist die Berechnung des Erwartungswertes nur eine spezielle Art der Quadratur der Funktion  $f(x)$  bei der die Wahrscheinlichkeit der jeweiligen Werte berücksichtigt wird. Numerisch approximieren wir dieses Integral wieder mithilfe von Gewichten und Stützpunkten, d.h.

$$E[f] = \int_D f(x) \cdot p(x) dx \approx \sum_{i=0}^{n-1} w_i f(x_i). \quad (5.2)$$

In dieser Aufgabe betrachten wir Integrale dieser Art. Als Wahrscheinlichkeitsdichte verwenden wir die Dreieksdichte

$$p(x) = -|x| + 1, \quad (5.3)$$

welche auch hier zu sehen ist:



Im Folgenden betrachten wir verschiedene Verfahren, welche für die Berechnung des Integrals verwendet werden könnten. Als Integrationsgebiet verwenden wir  $D = [a, b] = [-1, 1]$ .

0
1
2

- a) Erläutern Sie warum die Verwendung der Trapezregel (also nur mit einem Trapez!) für die Berechnung des Integrals nicht geeignet ist.





b)\* In der Praxis werden für spezielle Wahrscheinlichkeitsdichten eigene Quadraturregeln berechnet, welche das Integral effizienter berechnen. Hierfür kann, wie auf dem Übungsblatt, die Methode der unbestimmten Koeffizienten verwendet werden. Hierfür muss lediglich die rechte Seite angepasst werden:

$$f_k(x) := x^k$$

$$\sum_{i=0}^{n-1} f_k(x_i) \cdot w_i = \int_a^b f_k(x) p(x) dx = M_k.$$

	0
	1
	2
	3
	4

Die Terme  $M_k$  werden auch als Momente der Wahrscheinlichkeitsdichte  $p(x)$  bezeichnet.

Wir wollen nun eine Gaussquadatur mit 2 Punkten speziell für die Dreiecksdichte berechnen. Verwenden Sie hierfür  $M_0 = 1$ ,  $M_1 = 0$ ,  $M_2 = 1/6$  und  $M_3 = 0$ . Zusätzlich kann zur Vereinfachung der Berechnung die Beziehung  $x_0 = -x_1$  verwendet werden. Stellen sie alle nötigen Gleichungen auf und berechnen Sie die Gewichte  $w_0$  und  $w_1$  und die Punkte  $x_0$  und  $x_1$ . Geben Sie auch den Rechenweg an!

c)\* Berechnen Sie nun für die Funktion  $f(x) = x^2 + 5$  und die Dreiecksdichte  $p(x)$  den Erwartungswert  $E[f]$  mithilfe der von Ihnen hergeleiteten Gaussquadatur. Falls Sie die Aufgabe b) nicht gelöst haben, verwenden Sie  $x_0 = -\frac{1}{3}$ ,  $x_1 = \frac{1}{3}$  und  $w_0 = \frac{1}{3}$ ,  $w_1 = \frac{2}{3}$ .

	0
	1
	2

d)\* Ist der berechnete Erwartungswert aus der vorherigen Teilaufgabe exakt? Begründen Sie Ihre Antwort! Berechnen Sie dabei den Erwartungswert  $E[f]$  nicht analytisch.

	0
	1





Zusätzlicher Platz für Lösungen. Markieren Sie deutlich die Zuordnung zur jeweiligen Teilaufgabe. Vergessen Sie nicht, ungültige Lösungen zu streichen.

