NUM-Prüfung 31.1.2018

1. (a) Beeinflussen die folgenden Größen die Maschinengenauigkeit eines Gleitkommazahlensystems? Kreuzen Sie die jeweils richtige Antwort an (2 Punkte).

   Die Maschinengenauigkeit ist mit Runden auf die nächstgelegene Nachkommastelle $\frac{\beta^{1-p}}{2}$ und mit Runden durch Abschneiden β^{1-p}, wobei β die Basis und p die Mantissenlänge ist. Daraus ergeben sich die Antworten.

   - Basis des Gleitkommazahlensystems: Ja
   - minimaler Exponent des Gleitkommazahlensystems: Nein
   - maximaler Exponent des Gleitkommasystems: Nein
   - Mantissenlänge: Ja
   - Art der Rundung: Ja

   (b) Die Maschinengenauigkeit eines Gleitkommazahlensystems ist eine obere Schranke für eine wichtige Eigenschaft des Gleitkommazahlensystems. Für welche (2 Punkte)?

   Beim Runden auf eine Maschinenzahl ist der relative Fehler höchstens die Maschinengenauigkeit.

   (c) Im Standard IEEE 754-2008 ist half precision (offiziell binary 16) durch folgende Festlegungen spezifiziert:

   - Basis: 2
   - minimaler Exponent: -14
   - maximaler Exponent: 15
   - Mantissenlänge: 11 (wovon nur 10 Stellen explizit gespeichert werden müssen)

   Welchen Wert hat die Maschinengenauigkeit in half precision, wenn Rundung auf die nächstgelegene Maschinenzahl verwendet wird (1 Punkt)?

   $\beta = 2, p = 11$ und es wird auf die nächstgelegene Maschinenzahl gerundet. Wir erhalten $\epsilon_{mach} = \frac{\beta^{1-p}}{2} = \frac{2^{1-11}}{2} = 2^{-11} = \frac{1}{2048}$.

(d) Betrachten Sie die beiden Dezimalzahlen $a = 11.38$ und $b = 0.01372 \cdot 10^2$.

i. Stellen Sie diese beiden Dezimalzahlen in einem Gleitkommazahlensystem mit Basis 10 und Mantissenlänge 3 dar. Verwenden Sie diejenige Normalisierung, bei der die führende Ziffer d_0 ungleich 0 ist. Falls notwendig, runden Sie auf die nächstgelegene Maschinenzahl.

Bei a müssen wir die Manstisse um 1 nach rechts verschieben und erhalten $1.138 \cdot 10^1$. Da unsere Mantissenlänge 3 ist, müssen wir die letzte Stelle wegrunden und erhalten $1.14 \cdot 10^1$.

Bei b geht das analog: $0.01372 \cdot 10^2 = 1.372 \cdot 10^0 \rightarrow 1.37 \cdot 10^0$

ii. Berechnen Sie die Differenz $a - b$ in diesem Gleitkommazahlensystem. Geben Sie jeden Schritt explizit an und beachten Sie, dass das Ergebnis normalisiert werden muss.
Bestimmen Sie den resultierensen relativen Fehler (3 Punkte). (Darstellung als Bruch ausreichend!)

Um die Exponenten auszugleichen, verschieben wir die Mantisse von b um 1 nach rechts und erhalten $1.14 \cdot 10^1 - 0.137 \cdot 10^1 = 1.003 \cdot 10^1$. Wenn wir die letzte Stelle wegrunden, erhalten wir $1.00 \cdot 10^1$.

Für das exakte Ergebnis berechnen wir $a - b$ ohne eine einzige Rundung: $a - b = 11.38 - 1.372 = 10.008$.

Der relative Fehler ist dann $\left|\frac{10-10.008}{10.008}\right| = \frac{0.008}{10.008} = \frac{8}{10008}$.

2. (a) Gegeben sei die Funktion $f(x) = e^x + x^2 - 2$.

   i. Formuliere folgendes Nullstellenproblem in ein Fixpunktproblem um (2 Punkte): $f(x) = 0$

$$\begin{aligned} e^x + x^2 - 2 &= 0 && | + \tfrac{x}{10} \\ e^x + x^2 + \tfrac{x}{10} - 2 &= \tfrac{x}{10} && | \cdot 10 \\ 10e^x + 10x^2 + x - 20 &= x \\ g(x) &= x \end{aligned}$$

   $g(x) = 10e^x + 10x^2 + x - 20 = x$

   Warum habe ich nicht einfach auf beiden Seiten mit x addiert? Für diese Teilaufgabe könnte man das machen, jedoch ist es schwierig, herauszufinden, ob die Fixpunktiteration konvergiert. Wenn man auf beiden Seiten $\frac{x}{n}$ addiert, wobei n eine "große" Zahl ist, und anschließend auf beiden Seiten mit n multipliziert, dann ist es ziemlich wahrscheinlich, dass die Fixpunktiteration nicht konvergiert.

   ii. Beginnt man die Fixpunktiteration bei $x_0 = 0$, welchen Punkt x_1 würde man im nächsten Schritt erhalten (2 Punkte)?

   $x_1 = g(x_0) = g(0) = 10e^0 + 10 \cdot 0^2 + 0 - 20 = -10$

   iii. Erwarten Sie, dass die Fixpunktiteration konvergiert (1 Punkt)?

   Sie wird nicht konvergieren. Schauen wir uns $x_2 = g(-10) = 10e^{-10} + 10 \cdot (-10)^2 - 10 - 20 \geq 970$ an. Man sieht relativ leicht, dass die Fixpunktiteration gegen unendlich geht, den e^x und x^2 beschleunigen das Wachstum.

(b) Gegeben sei die Funktion $g(x) = x^3$

   i. Formuliere folgendes Fixpunktproblem in ein Nullstellenproblem um (1 Punkte): $g(x) = x$

$$\begin{aligned} x^3 &= x && | - x \\ x^3 - x &= 0 \\ f(x) &= 0 \end{aligned}$$

   $f(x) = x^3 - x = 0$

   ii. Wendet man dann die Newton Nullstellensuche für diese Aufgabe an und startet bei $x_0 = 1/2$, welchen Punkt x_1 erreicht man im nächsten Schritt (3 Punkte)?

   $f'(x) = 3x^2 - 1, \quad x_1 = f(1/2) = 1/2 - \frac{f(1/2)}{f'(1/2)} = 1/2 - \frac{-3/8}{-1/4} = -1$

   iii. Bestimme die 3 Nullstellen für die in Aufgabe (b) i. als Antwort erhaltene Funktion (zur Nullstellenbestimmung) exakt (1 Punkt).

   $x^3 - x = x(x^2 - 1) = 0 \Rightarrow x_1 = 0, x^2 - 1 = 0 \Rightarrow x^2 = 1 \Rightarrow x_{2,3} = \pm 1 \Rightarrow x_2 = -1, x_3 = 1$

3. Gegeben seien folgende Messwerte: $f(1) = 4$, $f(2) = 10$.

   (a) Bestimme mittels linearer Interpolation den Wert an der Stelle 1.5 (3 Punkte).

   Wir berechnen das Interpolationspolynom mithilfe der Monomialen Basis:

$$\left(\begin{array}{cc|c} 1 & 1 & 4 \\ 1 & 2 & 10 \end{array}\right) \to \left(\begin{array}{cc|c} 1 & 1 & 4 \\ 0 & 1 & 6 \end{array}\right) \to \left(\begin{array}{cc|c} 1 & 0 & -2 \\ 0 & 1 & 6 \end{array}\right)$$

   Somit erhalten wir das lineare Interpolationspolynom $p(x) = -2 + 6x$.

   $p(1.5) = -2 + 6 \cdot 1.5 = 7$

   Man kann natürlich auch Lagrange oder Newton verwenden, da es nur um das Ergebnis an der Stelle 1.5 geht.

   (b) Falls die zugrunde liegende Funktion eigentlich $f(x) = x^2 + 3x$ ist, bestimme für die Rechnung aus (a) den absoluten und den relativen Fehler (2 Punkte).

   $f(1.5) = 1.5^2 + 3 \cdot 1.5 = 6.75$

   Absoluter Fehler $= |7 - 6.75| = 0.25$
   Relativer Fehler $= |\frac{0.25}{6.75}| = \frac{1}{27}$

   (c) Man erhält als weiteren Messwert $f(3) = 20$. Welches Polynom erhält man in der Monomialen Basis? Tragen Sie die Koeffizienten der Monome ein (3 Punkte):

   Wir verwenden wieder die Vandermonde-Matrix um das Interpolationspolynom zu bestimmen.

$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & 4 \\ 1 & 2 & 4 & 10 \\ 1 & 3 & 9 & 20 \end{array}\right) \to \left(\begin{array}{ccc|c} 1 & 1 & 1 & 4 \\ 0 & 1 & 3 & 6 \\ 0 & 2 & 8 & 16 \end{array}\right) \to \left(\begin{array}{ccc|c} 1 & 1 & 1 & 4 \\ 0 & 1 & 3 & 6 \\ 0 & 0 & 2 & 4 \end{array}\right) \to \left(\begin{array}{ccc|c} 1 & 1 & 1 & 4 \\ 0 & 1 & 3 & 6 \\ 0 & 0 & 1 & 2 \end{array}\right) \to$$
$$\left(\begin{array}{ccc|c} 1 & 1 & 0 & 2 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 2 \end{array}\right) \to \left(\begin{array}{ccc|c} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 2 \end{array}\right)$$

   Koeffizient für das Basiselement 1: 2
   Koeffizient für das Basiselement t: 0
   Koeffizient für das Basiselement t^2: 2

   (d) Bestimme jetzt den Wert der interpolierenden Funktion an der Stelle 1.5 (2 Punkte).
   $p(1.5) = 2 + 2 \cdot 1.5^2 = 6.5$

4. Für die Matrix A ist die Zerlegung gegeben:

$$A = \begin{pmatrix} 12 & 4 & 4 & 4 \\ 0 & 12 & 0 & -8 \\ 3 & 7 & -3 & 5 \\ 6 & 5 & 3 & -10 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1/4 & 1/2 & 1 & 0 \\ 1/2 & 1/4 & -1/4 & 1 \end{pmatrix} \begin{pmatrix} 12 & 4 & 4 & 4 \\ 0 & 12 & 0 & -8 \\ 0 & 0 & -4 & 8 \\ 0 & 0 & 0 & -8 \end{pmatrix} = LU$$

(a) Bestimme die Determinante von A (1 Punkt).

$\det(A) = \det(LU) = \det(L)\det(U) = (1 \cdot 1 \cdot 1 \cdot 1)(12 \cdot 12 \cdot (-4) \cdot (-8)) = 4608$

(b) Bestimme die Eigenwerte der Matrix U (2 Punkte).

Einfach auf der Hauptdiagonale ablesen.

Eigenwert 1: 12
Eigenwert 2: 12
Eigenwert 3: -4
Eigenwert 4: -8

Man kann auch λ von der Hauptdiagonale abziehen und die Determinante berechnen. Das charakteristische Polynom ist dann $(12-\lambda)^2(-4-\lambda)(-8-\lambda)$, wodurch man auf die Eigenwerte kommt.

(c) Es sei Q eine orthogonale Matrix und es gelte

$$Q \begin{pmatrix} 3 \\ 0 \\ 4 \end{pmatrix} = \begin{pmatrix} 0 \\ c \\ 0 \end{pmatrix}$$

Geben Sie $|c|$ an (2 Punkte):
Wendet man eine orthogonale Matix auf einen Vektor an, so ändert sich die 2-Norm des Vektors nicht. Es gilt also:

$$\left\| \begin{pmatrix} 3 \\ 0 \\ 4 \end{pmatrix} \right\|_2 = \left\| \begin{pmatrix} 0 \\ c \\ 0 \end{pmatrix} \right\|_2 \Rightarrow \sqrt{3^2 + 0^2 + 4^2} = \sqrt{0^2 + c^2 + 0^2} \Rightarrow 5 = |c|$$

(d) Kreuzen Sie die wahren Aussagen an (5 Punkte):

- Jede reelle symmetrische $n \times n$ Matrix hat n unterschiedliche Eigenwerte

  Nein. Die Einheitsmatrix hat den Eigenwert 1 n-mal.

- Jede reelle symmetrische Matrix hat nur reelle Eigenwerte

  Ja.

- Für reelle symmetrische Matrizen sind die Eigenvektoren unterschiedlicher Eigenwerte orthogonal

  Ja.

- Die Matrix L der LU-Zerlegung ist orthogonal

  Nein. Ein Gegenbeispiel ist die Matrix L aus Beispiel 4.

- Das Produkt zweier Givens Rotationen ist orthogonal

  Ja. Eine Givens Rotation ist orthogonal und das Produkt von zwei orthogonalen Matrizen ist orthogonal.