# Wissenschaftliches Rechnen - Großübung 1.2

**Themen: Gleitkommazahlen, Kondition**

Ugo & Gabriel

8. November 2022

## Aufgabe 1: Gleitkommazahlen

1. Wie ist der absolute Fehler durch eine fehlerbehaftete Funktion $G$ definiert?

2. Wie ist der relative Fehler durch eine fehlerbehaftete Funktion $G$ definiert?

3. Gegeben sei das dezimale Gleitkommazahlenformat $\mathbb{G}(10, 3)$ mit 3 Ziffern und das dezimale Festkommazahlenformat $\mathbb{F}(10, 2, 2)$ mit zwei Stellen vor und zwei Stellen nach dem Komma, sowie die Funktionen $G : \mathbb{R} \to \mathbb{G}(10, 3)$ und $F : \mathbb{R} \to \mathbb{F}(10, 2, 2)$, die jeweils auf die nächste darstellbare Zahl **abrunden**.
   a) Geben Sie den Abstand zwischen zwei Zahlen in den jeweiligen Formaten im Intervall $[0{,}1, 1[$ sowie $[10, 100[$ an
   b) Geben Sie die obere Grenze des absoluten Fehlers an, der sich durch $G$ sowie $F$ auf dem Intervallen $[0{,}1, 1[$ sowie $[10, 100[$ ergibt.
   c) Geben Sie die obere Grenze des relativen Fehlers an, der sich durch $G$ sowie $F$ auf dem Intervallen $[0{,}1, 1[$ sowie $[10, 100[$ ergibt.

4. Welche der folgenden Gesetze gelten für Festkommazahlen?
   a) Assoziativgesetz für die Addition: $a + (b + c) = (a + b) + c$
   b) Distributivgesetz: $a(b + c) = ab + ac$
   c) Transitivität bzgl. Kleiner: $a > b \wedge b > c \Rightarrow a > c$
   d) Transitivität bzgl. Gleich: $a = b \wedge b = c \Rightarrow a = c$
   e) Antisymmetrie $a \leq b \wedge b \leq a \Rightarrow a = b$

5. Geben Sie die zwei in der Vorlesung/Skript vorgestellten Definitionen der Maschinengenauigkeit an.

6. Die zwei Definitionen sind äquivalent für den Fall $G(x) = \text{floor}(x)$, wobei floor auf die nächste Gleitkommazahl des gegebenen Gleitkommzahlenformates abrundet. Überprüfen Sie ob diese Definition für unterschiedliche Funktionen übereinstimmen, indem Sie die Werte der jeweiligen Definition berechnen:
   a) $G_\mathsf{f}(x) = \text{floor}(x)$, wobei floor auf die nächste Gleitkommazahl des gegebenen Gleitkommzahlenformates abrundet.
   b) $G_\mathsf{c}(x) = \text{ceil}(x)$, wobei ceil auf die nächste Gleitkommazahl des gegebenen Gleitkommzahlenformates aufrundet.

c) $G_{\mathsf{r}}(x) = \text{round}(x)$, wobei round auf die nächste Gleitkommazahl des gegebenen Gleitkommzahlenformates kaufmännisch rundet.

7. Geben Sie eine sinnvolle Obergrenze für den Fehler, der bei der Division zweier Gleitkommazahlen $x, y \in \mathbb{G}(b, n_m)$ entstehen kann, in Abhängigkeit der Mantissenstellen $n_m$ und Basis $b$ an (relativer Fehler von $\frac{G(x)}{G(y)}$).

8. Gegeben sei das dezimale Gleitkommazahlenformat $\mathbb{G}(10, 3)$ mit 3 Ziffern sowie eine beliebige Zahl $k$. Geben Sie eine Subtraktion $x - y$ an, die einen größeren oder gleich großen relativen Fehler hat als/wie $k$.

# Aufgabe 2: Kondition

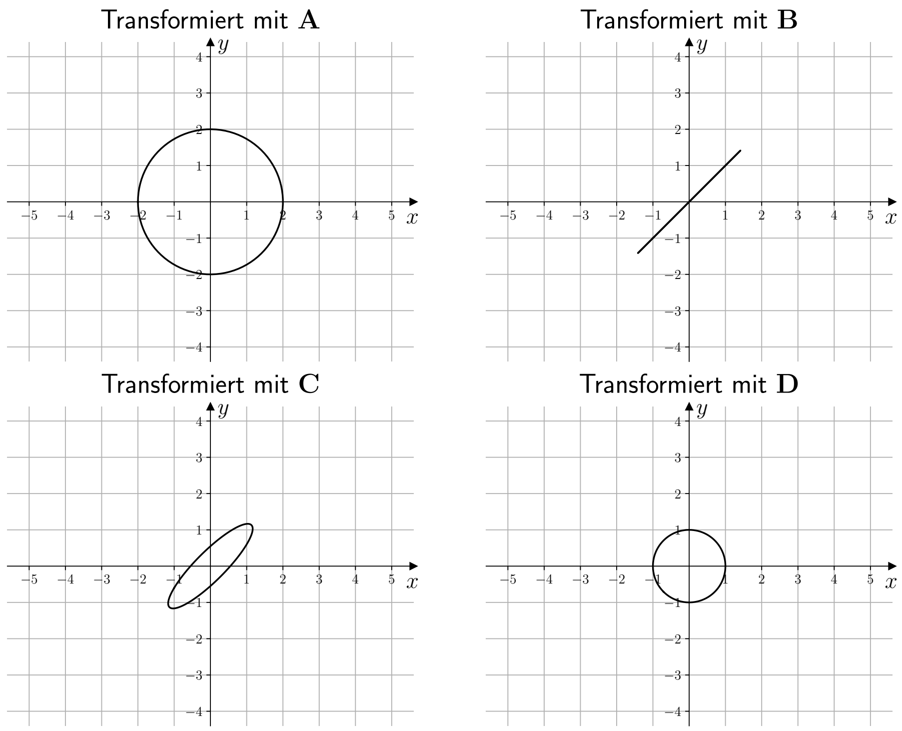
Die Kondition[1] einer Matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ ist definiert als

$$\kappa(\mathbf{A}) = \frac{\max_{\|\mathbf{x}\|=1}\|\mathbf{Ax}\|}{\min_{\|\mathbf{x}\|=1}\|\mathbf{Ax}\|}$$

und charakterisiert den potentiellen numerischen Genauigkeitsverlust jener Matrix. Zunächst kann die Norm $\|\cdot\|$ beliebig gewählt werden. Wie (fast) überall sonst im Kurs wählen wir im Folgenden die euklidische/$\ell^2$-Norm.

1. Wie sieht die Menge aus, die durch $\|\mathbf{x}\|_2 = 1$ beschrieben wird?

2. Gegeben seien vier lineare Transformationen $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D} \in \mathbb{R}^{2\times 2}$. Im Folgenden ist die Transformation des Einheitskreises unter diesen vier Transformationen zu sehen.




Entscheiden Sie, ob die folgenden Aussagen gelten oder nicht.

a) $\mathbf{A}$ ist orthogonal.

b) $\mathbf{B}$ ist singulär.

c) $\mathbf{C}$ ist regulär.

d) $\mathbf{D}$ ist orthogonal.

e) $\mathbf{A}$ hat eine Kondition von $1$.

f) $\mathbf{A}$ hat eine größere Kondition als $\mathbf{D}$.

g) $\mathbf{C}$ hat eine größere Kondition als $\mathbf{B}$.

h) $\mathbf{C}$ hat eine größere Kondition als $\mathbf{D}$.

[1] Falls der Nenner zu Null wird, gilt per Konvention $\kappa(\mathbf{A}) = \infty$.

3. Geben Sie, unter Zuhilfenahme der Erkenntnisse der vorherigen Aufgabe, eine geometrische Interpretation für die Kondition an.

4. Berechnen Sie die Kondition der folgenden Matrizen:

$$\mathbf{A} = \begin{bmatrix} 12 & 0 \\ 0 & 3 \end{bmatrix}, \qquad \mathbf{B} = \begin{bmatrix} 1/8 & 0 \\ 0 & 8 \end{bmatrix}$$

5. Matrizen mit schlechter Kondition müssen nicht unbedingt einen hohen Genauigkeitsverlust aufweisen. Geben Sie eine Matrix $\mathbf{A} \in \mathbb{G}(10,3)^{3\times 3}$ mit einer endlichen Kondition von größer oder gleich $100$ an, welche einen relativen Fehler von $0$ für alle Berechnungen $\mathbf{Ax}$ mit $\mathbf{x} \in \mathbb{G}(10,3)^3$ aufweist.