

ÜBUNG 3 (LÖSUNG)

Ausgabedatum: 28. Oktober 2024
Abgabedatum: 4. November 2024

Hausaufgabe 3.1 (Stabilität der Q-Konvergenzordnungen bei Normwechsel) 4 + 8 = 12 Punkte

- Es seien zwei äquivalente Normen $\|\cdot\|_a, \|\cdot\|_b : \mathbb{R}^n \rightarrow \mathbb{R}$ gegeben. Zeigen Sie, dass aus der Q-superlinearen bzw. Q-quadratischen Konvergenz einer Folge bzgl. $\|\cdot\|_a$ die Q-superlineare bzw. Q-quadratische Konvergenz bzgl. $\|\cdot\|_b$ folgt.
- Erklären Sie, warum die Argumente für den Nachweis der Aussage in Punkt (a) für *Q-lineare* Konvergenz nicht funktionieren, und geben Sie ein Gegenbeispiel an, das zeigt, dass eine entsprechende Aussage für *Q-lineare* Konvergenz i. A. nicht gilt.

Lösung.

- (a) Gegeben Sei die Normäquivalenz $\underline{\alpha}\|\cdot\|_b \leq \|\cdot\|_a \leq \bar{\alpha}\|\cdot\|_b$ mit zwei Konstanten $\underline{\alpha}, \bar{\alpha} > 0$. Sei nun zusätzlich $(x^{(k)})_{k \in \mathbb{N}_0}, x^*$, aus \mathbb{R}^n .

- Wenn $x^{(k)} \rightarrow x^*$ superlinear in $\|\cdot\|_a$, dann existiert $\varepsilon^{(k)} \rightarrow 0$, so dass

$$\underline{\alpha}\|x^{(k+1)} - x^*\|_b \leq \|x^{(k+1)} - x^*\|_a \leq \varepsilon^{(k)}\|x^{(k)} - x^*\|_a \leq \bar{\alpha}\varepsilon^{(k)}\|x^{(k)} - x^*\|_b$$

und somit

$$\|x^{(k+1)} - x^*\|_b \leq \frac{\bar{\alpha}}{\underline{\alpha}}\varepsilon^{(k)}\|x^{(k)} - x^*\|_b$$

mit $\frac{\bar{\alpha}}{\underline{\alpha}}\varepsilon^{(k)} \rightarrow 0$. (2 Punkte)

- Wenn $x^{(k)} \rightarrow x^*$ quadratisch in $\|\cdot\|_a$, dann konvergiert die Folge erstmal überhaupt bzgl. $\|\cdot\|_a$ und somit auch bzgl. $\|\cdot\|_b$ gegen x^* . Außerdem existiert $C > 0$, so dass

$$\underline{\alpha}\|x^{(k+1)} - x^*\|_b \leq \|x^{(k+1)} - x^*\|_a \leq C\|x^{(k)} - x^*\|_a^2 \leq \bar{\alpha}^2 C\|x^{(k)} - x^*\|_b^2$$

und somit

$$\|x^{(k+1)} - x^*\|_b \leq \frac{\bar{\alpha}^2}{\underline{\alpha}} C \|x^{(k)} - x^*\|_b^2$$

mit $\frac{\bar{\alpha}^2}{\underline{\alpha}} C > 0$. (2 Punkte)

(b) Wenn $x^{(k)} \rightarrow x^*$ linear in $\|\cdot\|_a$, dann existiert $c \in (0, 1)$, so dass

$$\underline{\alpha} \|x^{(k+1)} - x^*\|_b \leq \|x^{(k+1)} - x^*\|_a \leq c \|x^{(k)} - x^*\|_a \leq \bar{\alpha} c \|x^{(k)} - x^*\|_b$$

und somit

$$\|x^{(k+1)} - x^*\|_b \leq \frac{\bar{\alpha}}{\underline{\alpha}} c \|x^{(k)} - x^*\|_b.$$

Die neue Konstante $\frac{\bar{\alpha}}{\underline{\alpha}} c$ ist zwar positiv, jedoch nicht mehr notwendigerweise kleiner als 1. (2 Punkte)

Als Gegenbeispiel betrachten wir den \mathbb{R}^2 mit der euklidischen Norm und einer Norm die durch Skalierung entlang einer der Achsen entsteht, also für einen Parameter $\bar{\alpha} > 1$ die Normen

$$\|x\|_a = \sqrt{x^\top \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} x} \quad \text{und} \quad \|x\|_b = \sqrt{x^\top \begin{pmatrix} \bar{\alpha}^2 & 0 \\ 0 & 1 \end{pmatrix} x}$$

Dann gilt

$$\|x\|_b \leq \|x\|_a \leq \bar{\alpha} \|x\|_b$$

für alle $x \in \mathbb{R}^2$. Wir betrachten nun eine Folge, die zwischen den beiden Achsen hin und her springt und damit das "ungünstige" Skalierungsverhältnis in der skalierten Norm immer wieder mitnimmt. Für $c := \frac{1}{\bar{\alpha}} \in (0, 1)$ definieren wir

$$x^{(k)} := \begin{cases} (c^k, 0) & \text{für } k \text{ gerade} \\ (0, c^k) & \text{für } k \text{ ungerade} \end{cases}$$

Dann gilt offensichtlich $\|x^{(k)}\|_a = c^k \rightarrow 0$ und damit wegen

$$\|x^{(k+1)}\|_a = c^{k+1} = c c^k = c \|x^{(k)}\|_a$$

lineare Konvergenz gegen 0 in der euklidischen Norm. Für ungerade $k \in \mathbb{N}_0$ ist aber

$$\|x^{(k+1)}\|_b = \bar{\alpha} c^{k+1} = \bar{\alpha} c c^k = \bar{\alpha} c \|x^{(k)}\|_b = \|x^{(k)}\|_b.$$

Der Sprung auf die hochskalierte x_1 -Achse (ungerades k auf gerades k) bringt also keinen Fortschritt im Fehler bzgl. der b -Norm und damit gilt natürlich keine lineare Konvergenz bzgl. der b -Norm. (6 Punkte)

Beachte: Die Frage, ob in unserem Beispiel oben eine in der a -Norm linear konvergente Folge in der b -Norm ebenfalls noch linear konvergiert, hängt von dem Verhältnis der Konstanten c und $\bar{\alpha}$ ab – genauer, ob die Zunahme im Fehler auf Grund von einem Skalierungswechsel bzgl. der Richtung des Fehlers den linearen Abstiegsfaktor (multiplikativ) übersteigt.

Es fällt übrigens auf, dass im obigen Beispiel

$$\|x^{(k+1)} - x^*\|_b \leq \bar{\alpha} c^{k+1} = c \bar{\alpha} c^k$$

für alle $k \in \mathbb{N}_0$ gilt. Es existiert also eine Folge mit Gliedern $\xi^{(k)} := \bar{\alpha} c^k$, die Q-linear konvergiert und $\|x^{(k+1)} - x^*\|_b$ majorisiert. Das ist genau die Definition von r-linearer Konvergenz. Im r-Konvergenz Begriff ist es einer Folge also möglich, nicht in jedem Schritt das Quotientenkriterium zu erfüllen, aber im großen und ganzen doch durch eine Q-konvergente Folge majorisiert zu konvergieren.

Hausaufgabe 3.2 (Langsame Konvergenz des Newton-Verfahrens zur Nullstellensuche) 6 Punkte

Berechnen Sie ausgehend vom Startpunkt $x^{(0)} = 1$ die Iterierten des Newton-Verfahrens zur Nullstellensuche ([Algorithmus 5.1](#)) für die Funktion

$$f(x) = x^2.$$

Zeigen Sie, dass die Iteriertenfolge konvergiert und entscheiden Sie, mit welcher Q-Konvergenzraten die Folge konvergiert. Warum ist das beobachtete Verhalten kein Widerspruch zu [Satz 5.7](#)?

Lösung.

Ausgehend von $x^{(0)} = 1$ bestimmen wir $x^{(k+1)} = x^{(k)} - \frac{f(x^{(k)})}{f'(x^{(k)})}$ und erhalten:

$$\begin{aligned} x^{(1)} &= 1 - \frac{1}{2} = \frac{1}{2} \\ x^{(2)} &= \frac{1}{2} - \frac{1/4}{1} = \frac{1}{4} \\ x^{(3)} &= \frac{1}{4} - \frac{1/16}{1/2} = \frac{1}{8} \\ &\vdots \\ x^{(k)} &= \frac{1}{2^k}. \end{aligned}$$

(2 Punkte)

Wir stellen fest, dass die Ungleichung

$$\underbrace{\|x^{(k+1)} - x^*\|}_{=1/2^{k+1}} \leq c \underbrace{\|x^{(k)} - x^*\|}_{=1/2^k}$$

(sogar mit Gleichheit) für $c = 1/2$ erfüllt ist. Hier liegt also offensichtlich lediglich Q-lineare Konvergenz vor. (2 Punkte)

Satz 5.7 besagt, dass wir unter den dort genannten Voraussetzungen Q-superlineare Konvergenz erwarten können. Allerdings ist bei unserer Funktion $f'(x^*) = 0$ nicht regulär und die Voraussetzungen sind damit nicht erfüllt. (2 Punkte)

Hausaufgabe 3.3 (Affine Invarianz des Newton-Verfahrens) 5 + 4 + 3 = 12 Punkte

Wir wollen die (affine) Invarianz des lokalen Newton-Verfahrens zur Lösung von Gleichungen der Form $F(x) = 0$ mit stetig differenzierbarem $F: \mathbb{R}^n \rightarrow \mathbb{R}^n$ ([Algorithmus 5.1](#)) untersuchen.

Es seien dafür $A \in \mathbb{R}^{n \times n}$ regulär und $b \in \mathbb{R}^n$ gegeben, und $(x^{(k)})_{k \in \mathbb{N}_0}$ eine vom Newton-Verfahren zum Startpunkt $x^{(0)} \in \mathbb{R}^n$ erzeugte Folge von Iterierten. Zeigen Sie:

- (a) Das Newton-Verfahren für die Funktion

$$G: \mathbb{R}^n \mapsto \mathbb{R}^n, \quad G(y) := F(Ay + b)$$

zum Startpunkt $y^{(0)} \in \mathbb{R}^n$ mit $x^{(0)} = Ay^{(0)} + b$ ist wohldefiniert und für die erzeugte Folge $(y^{(k)})_{k \in \mathbb{N}_0}$ von Iterierten gilt

$$x^{(k)} = Ay^{(k)} + b.$$

- (b) Das Newton-Verfahren für die Funktion

$$H: \mathbb{R}^n \mapsto \mathbb{R}^n, \quad H(y) := AF(y)$$

zum Startpunkt $y^{(0)} \in \mathbb{R}^n$ mit $x^{(0)} = y^{(0)}$ ist wohldefiniert und für die erzeugte Folge $(y^{(k)})_{k \in \mathbb{N}_0}$ von Iterierten gilt

$$x^{(k)} = y^{(k)}.$$

- (c) Erläutern Sie (intuitiv, nicht notwendigerweise mit Rechnungen oder Beispielen) was wir erwarten können, wenn wir die Transformation in [Punkt \(b\)](#) um eine konstante Verschiebung auf

$$H: \mathbb{R}^n \mapsto \mathbb{R}^n, \quad H(y) := AF(y) + b$$

erweitert wird.

Lösung.

(a) Das Newton-Verfahren liefert die durch

$$\begin{aligned}x^{(k+1)} &= x^{(k)} - F'(x^{(k)})^{-1}F(x^{(k)}) \\y^{(k+1)} &= y^{(k)} - G'(y^{(k)})^{-1}G(y^{(k)})\end{aligned}$$

gegebenen Folgen. Dabei besteht der Zusammenhang

$$\begin{aligned}G(y) &:= F(Ay + b) \\G'(y) &= F'(Ay + b)A\end{aligned}$$

(2 Punkte)

Aus $x^{(0)} = Ay^{(0)} + b$ ergibt sich

$$\begin{aligned}y^{(1)} &= y^{(0)} - G'(y^{(0)})^{-1}G(y^{(0)}) \\&= y^{(0)} - A^{-1}\underbrace{F'(Ay^{(0)} + b)^{-1}F(Ay^{(0)} + b)}_{=x^{(0)}} \\&= y^{(0)} - A^{-1}(x^{(0)} - x^{(1)}),\end{aligned}$$

so dass wegen

$$\begin{aligned}Ay^{(1)} + b &= \underbrace{Ay^{(0)} + b}_{=x^{(0)}} - (x^{(0)} - x^{(1)}) \\&= x^{(1)}\end{aligned}$$

per Induktion die Behauptung folgt.

(3 Punkte)

(b) Das Newton-Verfahren liefert die durch

$$\begin{aligned}x^{(k+1)} &= x^{(k)} - F'(x^{(k)})^{-1}F(x^{(k)}) \\y^{(k+1)} &= y^{(k)} - H'(y^{(k)})^{-1}H(y^{(k)})\end{aligned}$$

gegebenen Folgen. Dabei besteht der Zusammenhang

$$\begin{aligned}H(y) &:= AF(y) \\H'(y) &= AF'(y)\end{aligned}$$

(2 Punkte)

Aus $x^{(0)} = y^{(0)}$ ergibt sich

$$\begin{aligned} y^{(1)} &= y^{(0)} - H'(y^{(0)})^{-1}H(y^{(0)}) \\ &= y^{(0)} - F'\underbrace{(y^{(0)})}_{=x^{(0)}}^{-1}A^{-1}AF\underbrace{(y^{(0)})}_{=x^{(0)}} \\ &= x^{(1)} \end{aligned}$$

und per Induktion folgt die Behauptung. (2 Punkte)

- (c) Für affin-lineare Funktionen F können wir (sofern der lineare Teil invertierbar ist) davon ausgehen, dass ein vergleichbares Resultat gilt. Affin-lineare Transformationen im Wertebereich einer nichtlinearen Funktion haben allerdings i. A. einen nichtaffinen Einfluss auf die Lage der Nullstellen einer solchen Funktion. Das hat natürlich auch einen Einfluss auf die Newton Schritte, denn die Richtungen werden i. A. von der konstanten Verschiebung im Wertebereich abhängen. Das Update in jedem Schritt hängt affin-linear von der Verschiebung ab, die gesamte Folge aber i. A. nichtlinear. Wir werden also Beispiele finden können, in denen zwischen den Iterierten der Newtonverfahren für das ursprüngliche und das transformierte Problem kein affin linearer Zusammenhang (irgendeiner Art) besteht. (3 Punkte)

Die Funktion $F(x) = e^x$ auf \mathbb{R} mit vertikaler Verschiebung bietet sich hier als Beispiel an. Wir setzen dann $H(x) = e^x + b$ und untersuchen die ersten drei Iterierten für zwei verschiedene Startpunkte und Verschiebungen b (drei Iterierte, weil zwischen je 2 immer noch ein affin linearer Zusammenhang gefunden werden kann). Das ist aber nicht sonderlich illustrativ aufzuschreiben.

Beachte:

- Besonders gut lässt sich die Skalierungs(in)-varianz vom Newton- bzw. Gradientenverfahren in der Optimierung für den Fall, dass f eine quadratische Funktion ist veranschaulichen. In dem Fall ist auch g eine quadratische Funktion und das Newton-Verfahren hat nach einem Schritt die Lösung erreicht, während die Konvergenzgeschwindigkeit des Gradientenverfahrens von der Skalierungsmatrix A abhängt.
- Einige in der Vorlesung gemachten Aussagen zum Newton-Verfahren (z. B. Größe der Konvergenzumgebung) können durch die Transformation jedoch beliebig verändert werden.

Zusatzaufgabe 3.4 (Nichtkonvergenz des lokalen Newton-Verfahrens in der Optimierung) 2 + 3 = 5 Bonuspunkte

Das lokale Newton-Verfahren zur Nullstellensuche ([Algorithmus 5.1](#)) kann durch Anwendung auf die notwendigen Bedingungen erster Ordnung zur Minimierung der Funktion $f(x) = \cos(x)$ verwendet werden. Da die zweite Ableitung an den stationären Punkten nicht null ist, wissen wir, dass das

Verfahren q-superlinear gegen einen stationären Punkt konvergiert, wenn man nur nah genug an diesem startet. Was schiefgehen kann, wenn man das nicht tut, wollen wir hier untersuchen.

- Bestimmen Sie einen Startpunkt $x^{(0)}$, so dass die Folge der Iterationspunkte $x^{(k)}$ bestimmt gegen ∞ divergiert.
- Bestimmen Sie einen Startpunkt $x^{(0)}$, sodass die Folge der Iterationspunkte zwischen zwei verschiedenen (nicht optimalen) Punkten alterniert.

Lösung.

Der Newton-Schritt für $F(x) = f'(x)$ lautet

$$x^{(k+1)} = x^{(k)} + d^{(k)} = x^{(k)} - f''(x^{(k)})^{-1}f'(x^{(k)})$$

und konkret für unseren Fall von $f(x) = \cos(x)$ mit $f'(x) = -\sin(x)$, $f''(x) = -\cos(x)$:

$$x^{(k+1)} = x^{(k)} + d^{(k)} = x^{(k)} - \frac{-\sin(x^{(k)})}{-\cos(x^{(k)})} = x^{(k)} - \tan(x^{(k)}).$$

- Wir wollen uns die Periodizität der Funktion zu Nutzen machen und wollen eine Folge bauen, die in 2π Schritten gegen ∞ geht. Dafür setzen wir $x^{(k+1)} \stackrel{!}{=} x^{(k)} + 2\pi$ an. Dazu ist äquivalent, dass

$$x^{(k+1)} - x^{(k)} = d^{(k)} = -\tan(x^{(k)}) \stackrel{!}{=} 2\pi,$$

also dass

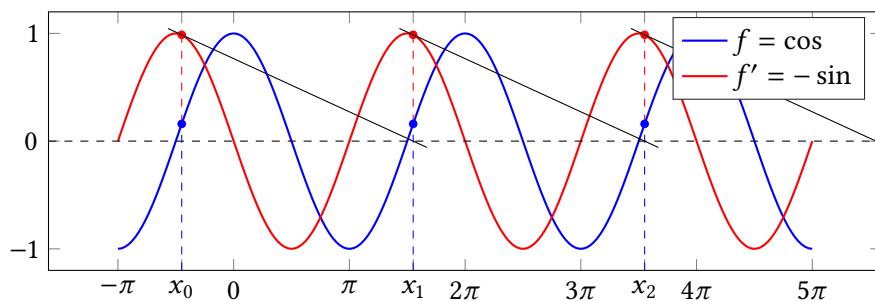
$$x^{(k)} = \arctan(-2\pi) + 2i(k)\pi$$

für alle $k \in \mathbb{N}_0$ und irgendein $i(k) \in \mathbb{Z}$. Wir können also beispielsweise den Startwert

$$x^{(0)} = \arctan(-2\pi) \approx -1.41$$

wählen.

(2 Punkte)



- (b) Wir wollen uns die Symmetrie der Funktion f und die Asymmetrie der Ableitung um den stationären Punkt (den **Maximierer**) $x^* = 0$ zu nutzen machen um eine Folge zu bauen, die immer zwischen $x^{(0)}$ und $-x^{(0)}$ alterniert. Dafür müssen wir nur die Bedingung $x^{(1)} = -x^{(0)}$ erfüllen, denn dann ist wegen der (A)symmetrie der Funktion bzw. ihrer Ableitung auch

$$x^{(2)} = x^{(1)} + d^{(1)} = x^{(0)} + d^{(0)} + d^{(1)} = x^{(0)} + d^{(0)} - d^{(0)} = x^{(0)}$$

und wir sind zurück am Ausgangspunkt, an dem sich das Ganze wiederholt.

Wir versuchen also einen Punkt $x^{(0)}$ zu finden, der

$$x^{(1)} = x^{(0)} - \tan(x^{(0)}) = -x^{(0)}$$

erfüllt, also die Gleichung

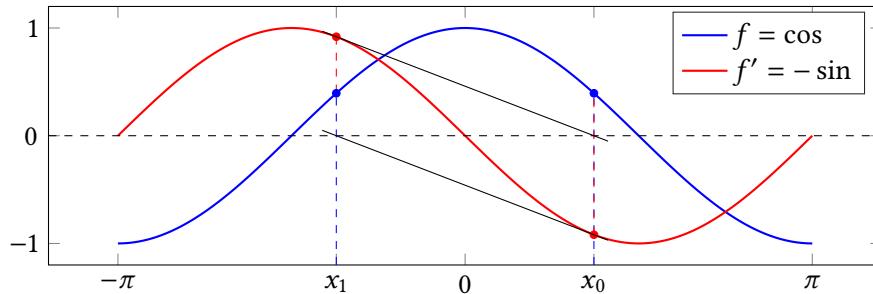
$$x^{(0)} = \frac{1}{2} \tan(x^{(0)})$$

löst. Diese Gleichung hat natürlich auch bspw. die Lösung $x^{(0)} = 0$ – das Verfahren startet also im Maximierer und terminiert sofort, bzw. das Verfahren produziert die konstante Folge, die immer den Startpunkt als jede Iterierte hat. Das gleiche Verhalten ist an jedem Optimalpunkt zu beobachten. Wir wollen aber eine Folge, die zwischen **verschiedenen** Punkten alterniert und zeigen also, dass die Gleichung auch eine Lösung zwischen 0 und dem nächsten, größeren Optimalwert (nämlich $\frac{\pi}{2}$) eine Lösung besitzt.

Dafür bemerken wir, dass für die stetige Funktion $g(x) := x - \frac{1}{2} \tan(x)$

$$g(0) = 0, \quad g'(0) = 1 - \frac{1}{\cos^2(0)} = \frac{1}{2}, \quad g(x) \xrightarrow{x \nearrow \frac{\pi}{2}} -\infty$$

gilt. Die Funktion hat also rechts der Null erst positive Werte und divergiert dann bestimmt gegen $-\infty$, womit eine weitere Nullstelle vor $\frac{\pi}{2}$ vorliegen muss, die die Gleichung löst. Diese liegt etwa bei $x^{(0)} \approx 1.1656$ (findet man z. B. numerisch). (3 Punkte)



Bitte reichen Sie Ihre Lösungen als ein PDF auf **Mampf** ein.