§3: Banach-Räume

Die reellen Zahlen unterscheiden sich vor allem dadurch von den rationalen Zahlen, daß viele Folgen rationaler Zahlen, die keinen Grenzwert in $\mathbb{Q}$ haben, doch gegen einen Grenzwert aus $\mathbb{R}$ konvergieren. Insbesondere hat in $\mathbb{R}$ jede CAUCHY-Folge einen Grenzwert. Diese Eigenschaft der reellen Zahlen wollen wir in diesem Paragraphen verallgemeinern und dabei auch sehen, daß sich beispielsweise das aus Kapitel I bekannte HERON-Verfahren zur Berechnung der Quadratwurzel einordnet in eine Gruppe viel allgemeinerer Techniken.

a) Vollständigkeit

CAUCHY-Folgen und der Begriff der Vollständigkeit lassen sich für beliebige metrische Räume definieren; da wir die Begriffe nicht in dieser Allgemeinheit benötigen, beschränken wir uns auf normierte Vektorräume; die Verallgemeinerung auf metrische Räume sollte für jeden interessierten Leser offensichtlich sein.

Definition: V sei ein normierter Vektorraum mit Norm $\|\cdot\|$.
a) Eine Folge $(x_n)_{n\in\mathbb{N}}$ von Elementen $x_n \in V$ heißt CAUCHY-*Folge*, wenn es zu jedem $\varepsilon > 0$ ein $N \in \mathbb{N}$ gibt, so daß $\|x_n - x_m\| < \varepsilon$ für alle $n, m \geq N$.
b) V heißt ein *vollständiger normierter Vektorraum* oder BANACH-*Raum*, wenn jede CAUCHY-Folge aus V gegen ein Element von V konvergiert.

STEFAN BANACH (1892–1945) wurde in Krakau geboren und ausgebildet, promovierte und arbeitete dann aber an der Universität von Lvov in der Ukraine, wo er unter schwierigen Bedingungen unter deutscher Besatzung den zweiten Weltkrieg verbrachte. Durch seine Arbeiten über lineare Operatoren und über Vektorräume von Funktionen wurde er zum Begründer der modernen Funktionalanalysis. Nach dem Krieg wollte er auf einen Lehrstuhl an der Universität Krakau wechseln, starb aber 1945 an Lungenkrebs. Das wichtigste mathematische Forschungsinstitut Polens, das Banach-Zentrum in Warschau, ist nach ihm benannt.

Einfachstes Beispiel eines BANACH-Raums ist natürlich $\mathbb{R}$ selbst mit der Betragsfunktion als Norm; hier ist die Vollständigkeitsaussage gerade

das CAUCHYsche Konvergenzkriterium. Da es in diesem Semester vor allem um Funktionen mehrerer Veränderlicher geht, sollten wir uns als nächstes überlegen, ob auch $\mathbb{R}^n$ ein BANACH-Raum ist. Während wir in $\mathbb{R}$ immer mit dem Betrag arbeiten, haben wir im Mehrdimensionalen allerdings verschiedene Normen, und müssen uns, bevor wir von Vollständigkeit reden können, auf eine davon festlegen.

Zumindest für die Konvergenz von Folgen kommt es im $\mathbb{R}^n$ *nicht* darauf an, mit welcher Norm wir arbeiten: Wie wir in Abschnitt *a)* gesehen haben, sind alle Normen auf $\mathbb{R}^n$ äquivalent, und wie wir bereits aus §1*b)* wissen, führen äquivalente Normen zum gleichen Konvergenzbegriff. Dasselbe gilt auch für CAUCHY-Folgen: Sind $\|\cdot\|_1$ und $\|\cdot\|_2$ zwei äquivalente Normen auf einem $\mathbb{R}$-Vektorraum V und ist $(x_n)_{n\in\mathbb{N}}$ eine CAUCHY-Folge bezüglich $\|\cdot\|_1$, so gibt es zu jedem $\varepsilon > 0$ ein $N \in \mathbb{N}$, so daß $\|x_n - x_m\|_1 < \varepsilon$ für alle $n, m \geq N$. Wegen der Äquivalenz der beiden Normen gibt es außerdem eine positive reelle Zahl c, so daß $\|x\|_2 \leq c\,\|x\|_1$ ist für alle $x \in V$. Wählen wir daher ein M, so daß $\|x_n - x_m\|_1 < \varepsilon/c$ ist für alle $n, m \geq M$, so ist

$$\|x_n - x_m\|_2 \leq c\,\|x_n - x_m\|_1 < c \cdot \frac{\varepsilon}{c} = \varepsilon$$

für alle $n, m \geq M$; die Folge ist also auch bezüglich $\|\cdot\|_2$ eine CAUCHY-Folge. Damit folgt insbesondere, daß V genau dann ein BANACH-Raum bezüglich der Norm $\|\cdot\|_2$ ist, wenn V ein BANACH-Raum bezüglich der dazu äquivalenten Norm $\|\cdot\|_1$ ist.

Speziell für $V = \mathbb{R}^n$, wo *alle* Normen äquivalent sind, reicht es also, die Vollständigkeit bezüglich irgendeiner beliebigen Norm zu beweisen; sie folgt dann automatisch auch für alle anderen Normen.

Lemma: $\mathbb{R}^n$ ist ein BANACH-Raum bezüglich der Maximumsnorm und damit bezüglich jeder beliebigen Norm.

Beweis: $(x_k)_{k\in\mathbb{N}}$ sei eine CAUCHY-Folge von Elementen aus $\mathbb{R}^n$; wir schreiben das n-Tupel $x_k \in \mathbb{R}^n$ als $x_k = (x_{k1}, \ldots, x_{kn})$. Zu jedem $\varepsilon > 0$ gibt es ein $N \in \mathbb{N}$, so daß

$$\|x_k - x_\ell\|_\infty = \max\bigl\{|x_{kj} - x_{\ell j}| \bigm| j = 1, \ldots, n\bigr\} < \varepsilon$$

ist für alle $k, \ell \geq N$. Damit ist insbesondere $\left|x_{kj} - x_{\ell j}\right| < \varepsilon$ für jeden Index j, d.h. die Folgen $(x_{kj})_{k\in\mathbb{N}}$ sind CAUCHY-Folgen reeller Zahlen. Nach dem CAUCHYschen Konvergenzkriterium konvergiert daher jede dieser Folgen gegen einen Grenzwert $y_j \in \mathbb{R}$. Damit konvergiert die Folge $(x_k)_{k\in\mathbb{N}}$ in $\mathbb{R}^n$ gegen den Punkt $(y_1, \dots, y_n)$, denn wie wir bereits in §1*b)* gesehen haben, ist Konvergenz bezüglich der Maximumsnorm einfach Konvergenz in jeder Komponente. Somit konvergiert jede CAUCHY-Folge in $\mathbb{R}^n$, und damit ist $\mathbb{R}^n$ vollständig, d.h. ein BANACH-Raum. ■

Da jeder endlichdimensionale $\mathbb{R}$-Vektorraum V isomorph ist zu einem $\mathbb{R}^n$, sind somit auch alle diese Räume vollständig. Um Vektorräume zu finden, die keine BANACH-Räume sind, müssen wir entweder $\mathbb{Q}$ oder allgemeiner $\mathbb{Q}^n$ betrachten oder aber unendlichdimensionale $\mathbb{R}$-Vektorräume. Im letzten Abschnitt dieses Paragraphen werden wir erste Beispiele von Funktionenräumen betrachten, die teils BANACH-Räume sind, teils auch nicht.

b) Fixpunkte von Abbildungen

Betrachten wir noch einmal das HERON-Verfahren zur näherungsweisen Berechnung von $\sqrt{2}$: Wir starten mit irgendeiner positiven Zahl x_0 und berechnen sukzessive neue Werte

$$x_n = f(x_{n-1}) \quad \text{mit} \quad f(x) = \frac{1}{2}\left(x + \frac{2}{x}\right).$$

Für $x = \sqrt{2}$ ist

$$f(\sqrt{2}) = \frac{1}{2}\left(\sqrt{2} + \frac{2}{\sqrt{2}}\right) = \frac{1}{2}\left(\sqrt{2} + \sqrt{2}\right) = \sqrt{2}\,;$$

ist umgekehrt x eine positive reelle Zahl mit $f(x) = x$, so ist $f(x) = x$ äquivalent zur Gleichung

$$x = \frac{1}{2}\left(x + \frac{2}{x}\right) \quad \text{oder} \quad \frac{1}{2}\left(\frac{2}{x} - x\right) = 0\,,$$

also ist $x = 2/x$ und somit $x^2 = 2$, was im Positiven nur die Lösung $x = \sqrt{2}$ hat. HERON hat also die Gleichung $x^2 = 2$ umgeschrieben in eine Gleichung $f(x) = x$, und löst sie näherungsweise, indem er auf einen beliebigen Startwert immer wieder die Funktion f anwendet. Lösungen von Gleichungen der Form $f(x) = x$ beschreiben *Fixpunkte* im Sinne der folgenden

Definition: M sei eine Menge und $f\colon M \to M$ eine Abbildung. Ein *Fixpunkt* von f ist ein Element $x \in M$ mit $f(x) = x$.

HERONs iterativer Ansatz funktioniert nicht für jede Funktion: Die Gleichung $x^2 = 2$ ist beispielsweise auch äquivalent zur Gleichung $x = g(x) = 2/x$; wenn wir aber ausgehend von $x_0 = 1$ immer wieder die Funktion g anwenden, pendeln wir nur zwischen den beiden Werten 1 und 2 hin und her, ohne der Wurzel je näher zu kommen.

Ein wichtiges Thema dieses Paragraphen ist die Frage, unter welchen Bedingungen ein iteratives Verfahren wie das von HERON zum Erfolg führt. Wir wollen dieses Problem nicht unter den schwächstmöglichen Voraussetzungen lösen, sondern suchen stattdessen nach einfach zu überprüfenden *hinreichenden* Kriterien. So ist auch die folgende Definition für den eindimensionalen Fall nicht die allgemeinstmögliche:

Definition: $f\colon D \to \mathbb{R}$ sei eine differenzierbare Abbildung auf der offenen Teilmenge $D \subseteq \mathbb{R}$. Ein Punkt $x \in D$ mit $f(x) = x$ heißt *stabiler* oder *anziehender* Fixpunkt von f, wenn $|f'(x)| < 1$ ist; er heißt *instabiler* oder *abstoßender* Fixpunkt, wenn $|f'(x)| > 1$ ist.

(Den Fall $|f'(x)| = 1$ betrachten wir nicht, da er im allgemeinen erheblich schwieriger zu behandeln ist.)

Lemma: Ist x ein anziehender Fixpunkt von f, so gibt es ein $\varepsilon > 0$, so daß für alle $x_0 \in D$ mit $|x - x_0| < \varepsilon$ die durch $x_k = f(x_{k-1})$ definierte Folge gegen x konvergiert. Für einen abstoßenden Fixpunkt dagegen konvergiert diese Folge nur dann gegen x, wenn sie bereits nach endlich vielen Iterationen den Wert$x$ erreicht.

Beweis: Sei zunächst x ein anziehender Fixpunkt. Für alle $h \in \mathbb{R}$ mit $y = x + h \in D$ ist dann

$$f(x+h) = f(x) + f'(x)h + o(h) = x + f'(x)h + o(h)\,.$$

Wir schreiben $|f'(x)| = 1 - 2c$; da $|f'(x)|$ kleiner ist als eins, ist die so definierte Zahl c positiv. Der Fehlerterm $o(h)$ geht schneller gegen Null geht als h; daher gibt es ein $\varepsilon_1 > 0$, so daß $|o(h)| < ch$ für alle h mit $|h| < \varepsilon_1$. Für solche h ist daher

$$\begin{aligned} |f(y) - f(x)| &= |f(x+h) - x| = |f'(x)h + o(h)| \le |f'(x)h| + |o(h)| \\ &< (1-2c)\,|h| + c\,|h| = (1-c)\,|h| \le |h| = |y - x|\,. \end{aligned}$$

Das allein reicht allerdings noch nicht, denn wir wissen nicht, ob $f(y)$ im Definitionsbereich D liegt, so daß wir f auch iterieren können.

Da D offen ist, gibt es aber ein $\varepsilon_2 > 0$, so daß alle $z \in \mathbb{R}^n$ mit $|z - x| < \varepsilon_2$ in D liegen. Nehmen wir nun als ε das Minimum von ε_1 und ε_2, so ist wieder für jedes $y \in D$ mit $|y - x| < \varepsilon$

$$|f(y) - x| \le (1 - c)\,|y - x| < \varepsilon\,,$$

und da $\varepsilon \le \varepsilon_2$ ist, liegt $f(y)$ in$D$.

Starten wir also mit einem x_0, für das $|x - x_0| < \varepsilon$ ist, so folgt induktiv, daß auch alle x_k mit $k \ge 1$ in D liegen, so daß wir die Folge $(x_k)_{k\in\mathbb{N}}$ definieren können; außerdem ist

$$|x - x_k| \le (1 - c)\,\big|x - x_{k-1}\big| \le \cdots \le (1 - c)^k\,|x - x_0| < (1 - c)^k \varepsilon\,.$$

Dies zeigt, daß die Folge der x_k gegen x konvergiert.

Gehen wir allerdings aus von einem abstoßenden Fixpunkt x, so ist $|f'(x)| > 1$; wir schreiben dies als $1 + 2c$ mit einer reellen Zahl $c > 0$. Wieder gibt es ein $\varepsilon > 0$, so daß in der Formel

$$f(x + h) = f(x) + f'(x)h + o(h) = x + f'(x) + o(h)$$

der Betrag von $o(h)$ kleiner ist als c für alle h mit $|h| < \varepsilon$. Für ein $y = x + h$ mit $|h| < \varepsilon$ ist daher

$$\begin{aligned}|f(y) - f(x)| &= |f(x + h) - x| = |f'(x)h + o(h)| \ge |f'(x)h| - |o(h)| \\ &> (1 - 2c)\,|h| + c\,|h| = (1 - c)\,|h| \ge |h| = |y - x|\;.\end{aligned}$$

Nehmen wir nun an, für irgendein x_0 aus D lasse sich die Folge $(x_k)_{k\in\mathbb{N}}$ definieren, und sie konvergiere gegen x. Dann gäbe es ein $N \in \mathbb{N}$, so daß $|x_n - x| < \varepsilon$ wäre für alle $n \ge N$. Für jedes $n \ge N$ wäre daher $|x_n - x| \ge |x_{n+1} - x|$, was für eine gegen x konvergierende Folge nur dann möglich ist, wenn alle $x_n = x$ sind.

■

Im Mehrdimensionalen ist die Situation *im Prinzip* genauso; der Beweis erfordert allerdings Sätze aus der Linearen Algebra, die nicht allen Hörern bekannt sind. Daher sei das Ergebnis nur kurz skizziert:

Für eine differenzierbare Abbildung $f\colon D \to \mathbb{R}^n$ auf einer offenen Teilmenge $D \subseteq \mathbb{R}^n$ mit Fixpunkt x ist, solange $x + h$ in D liegt,

$$f(x + h) = f(x) + J_f(h)\cdot h + o\big(\|h\|\big) = x + J_f(h)\cdot h + o(\|h\|)$$

und damit $\|f(x + h) - x\| \le \big\|J_f(h)\cdot h\big\| + \big\|o\big(\|h\|\big)\big\|$.

Den Fehlerterm können wir wieder für hinreichend kleine Norm von h unter jede gewünschte positive Schranke bringen; wir brauchen also in erster Linie eine Schranke für die Norm von $J_f(h)\cdot h$.

Falls $J_f(h)$ eine Diagonalmatrix ist, können wir vorgehen wie im eindimensionalen Fall: Die i-te Komponente des Vektors h wird mit dem i-ten Diagonaleintrag der Matrix multipliziert; falls dieser einen Betrag kleiner eins hat, konvergiert die Folge der iterierten Produkte zumindest in der i-ten Komponente gegen Null. Die Folge der x_k mit $x_k = f(x_{k-1}$ konvergiert also für einen Anfangswert x_0 hinreichend nahe bei x genau dann gegen x, wenn alle Diagonaleinträge Beträge kleiner als eins haben.

Leider ist $J_f(h)$ nur in den seltensten Fällen eine Diagonalmatrix. Für die Frage, ob eine Folge von Vektoren x_k gegen einen Vektor x konvergieren, ist es aber gleichgültig, in welcher Basis wir die Vektoren ausdrücken; falls es also eine Basis gibt, bezüglich derer $J_f(h)$ eine Diagonalmatrix ist, können wir wie oben argumentieren. Dabei spielt es keine Rolle, ob wir eine solche Basis für $\mathbb{R}^n$ oder nur für $\mathbb{C}^n$ finden können.

Die Einträge der Diagonalmatrix sind bekanntlich gerade die Eigenwerte von $J_f(h)$; die Folge der Iterierten konvergiert also, falls die Matrix $J_f(h)$ diagonalisierbar ist und alle ihre Eigenwerte Beträge kleiner eins haben.

Falls $J_f(h)$ nicht diagonalisierbar ist, läßt sich die Matrix als eine Summe $J_f(h) = D + N$ schreiben mit einer diagonalisierbaren Matrix D und einer Matrix N, deren Potenzen ab einem gewissen Exponenten $r \le n$ gleich der Nullmatrix sind. Außerdem kommutieren D und N, d.h. $DN = ND$. Deshalb können wir auf diese speziellen Matrizen den binomischen Lehrsatz anwenden und erhalten

$$J_f(h)^m = (D + N)^m = \sum_{k=0}^{m}\binom{m}{k} D^{m-k}N^k\,.$$

Für $k \ge r$ verschwindet N^k; für $m \ge r$ ist also

$$J_f(h)^m = (D+N)^m = \sum_{k=0}^{r-1}\binom{m}{k} D^{m-k}N^k = D^{m-r}\sum_{k=0}^{r-1}\binom{m}{k} D^{r-k}N^k\,.$$

Da r eine feste, von m unabhängige Zahl ist, wird das Verhalten von $J_f(h)^m \cdot h$ für $m \to \infty$ daher wieder von den Eigenwerten von $J_f(x)$ kontrolliert; auch hier konvergiert also die Folge der Iterierten, falls alle Eigenwerte von $J_f(h)$ einen Betrag kleiner eins haben. Deshalb definieren wir

Definition: $f\colon D \to \mathbb{R}^n$ sei eine differenzierbare Abbildung auf der offenen Teilmenge $D \subseteq \mathbb{R}^n$. Ein Punkt $x \in D$ mit $f(x) = x$ heißt *stabiler* oder *anziehender* Fixpunkt von f, wenn alle Eigenwerte der JACOBI-Matrix $J_f(x)$ einen Betrag kleiner eins haben; er heißt *instabiler* oder *abstoßender* Fixpunkt, wenn mindestens ein Eigenwert einen größeren Betrag als eins hat.

Damit läßt sich, mit praktisch demselben Beweis, das obige Lemma auch für höhere Dimensionen beweisen.

c) Das Newton-Verfahren

Nicht nur bei Extremwertproblemen, egal ob mit oder ohne Nebenbedingungen, ist es oft notwendig, die Nullstellen einer nichtlinearen Gleichung oder eines nichtlinearen Gleichungssystems zu bestimmen. Nur in sehr speziellen Fällen können diese Nullstellen exakt berechnet werden; meist muß man sich mit Näherungslösungen zufrieden geben.

Die numerische Mathematik kennt daher zahlreiche Methoden zur näherungsweisen Berechnung von Nullstellen; alle haben sowohl Stärken als auch Schwächen.

Das hier betrachtete NEWTON-Verfahren wird gerne verwendet bei differenzierbaren Funktionen, deren Ableitung sich einfach berechnen läßt, also beispielsweise bei Polynomen. Es wurde 1669 von ISAAC NEWTON vorgeschlagen und unabhängig davon 1690 von JOSEPH RAPHSON neu entdeckt und in der heute gebräuchlichen Form publiziert. Man bezeichnet es daher oft auch als Verfahren von NEWTON-RAPHSON. Wie viele numerische Verfahren ist es ein Iterationsverfahren; solche Verfahren haben den Vorteil, daß sie Rundungsfehler, die in einem Iterationsschritt entstehen, in den Folgeschritten im allgemeinen nicht vergrößern, sondern verkleinern.

SIR ISAAC NEWTON wurde gemäß dem damals noch in England geltenden Julianischen Kalender am 25. Dezember 1642 geboren. Nach dem in den meisten katholischen Staaten bereits eingeführten Gregorianischen Kalender war dies der 4. Januar 1643. Er studierte ab 1661 an der Universität Cambridge, wo er 1669 Professor wurde. Dort entwickelte er die Infinitesimalrechnung, die er 1671 in seinem Buch *De Methodis Serierum et Fluxionum* beschrieb, arbeitete über Optik, wo er unter anderem dünne Schichten und Beugungsphänomene untersuchte (NEWTONsche Ringe), entdeckte seine Bewegungsgesetze und das Gravitationsgesetz, veröffentlicht 1687 in seinem Buch *Philosophiae naturalis principia mathematica,* das von vielen als bedeutendstes wissenschaftliches Buch aller Zeiten angesehen wird. Nach zwei Nervenzusammenbrüchen ging er 1693 nach London, wo er die königliche Münze leitete. Er starb am 31. März 1727.

Über die Biographie von JOSEPH RAPHSON (1648–1715) ist sehr viel weniger bekannt. Auch er studierte in Cambridge, wo er 1692 seinen M.A. erhielt; bereits 1690 veröffentlichte er sein Buch *Analysis Aequationum universalis,* das seine Version des NEWTON-Verfahrens enthält, und wurde 1691 Mitglied der Royal Society. Spätere Bücher beschäftigen sich außer mit Mathematik auch mit theologischen und naturphilosophischen Fragen.

Gegeben sei eine differenzierbare Funktion $f\colon \mathbb{R} \to \mathbb{R}$; wir suchen eine Nullstelle x von f. Dem Thema dieses Paragraphen entsprechend wollen wir x als Fixpunkt einer Abbildung interpretieren und iterativ berechnen.

Der einfachste denkbare Ansatz besteht darin, daß wir die Gleichung $f(x) = 0$ umschreiben als $x = x + f(x)$. Testet man diesen Ansatz mit einfachen Funktionen $f(x)$, so merkt man schnell, daß er nur selten zu einem brauchbaren Ergebnis führt.

Das NEWTON-Verfahren geht aus von folgender Beobachtung: Ist x_0 eine *einfache* Nullstelle von f, d.h. $f'(x) \neq 0$, so schneidet die Tangente an die Kurve $y = f(x)$ im Punkt mit x-Koordinate x_0 die x-Achse an der Stelle $x = x_0$.

Für ein beliebiges x_0 aus dem Definitionsbereich von f, in dem $f'(x_0)$ nicht verschwindet, hat die Tangente im Punkt $\bigl(x_0, f(x_0)\bigr)$ die Gleichung $y = f(x_0) + f'(x_0)(x - x_0)$ und schneidet daher die x-Achse im Punkt

$$x = x_0 - \frac{f(x_0}{f'(x_0)}\,,$$

der in der Tat genau dann mit x_0 übereinstimmt, wenn $f(x_0)$ verschwindet. Zur iterativen Bestimmung einer Nullstelle können wir also versuchen, mit irgendeinem Startwert $x_0 \in D$ anzufangen und weitere Näherungswerte zu bestimmen durch die Vorschrift

$$x_n = x_{n-1} - \frac{f(x_{n-1})}{f'(x_{n-1})} \quad \text{für alle } n \in \mathbb{N}\,.$$

Als erstes Beispiel betrachten wir die Funktion $f(x) = x^2 - a$, deren Nullstellen die Quadratwurzeln von a sind. Hier ist $f'(x) = 2x$, für $x_i \neq 0$ haben wir also die Iterationsvorschrift

$$x_n = x_{n-1} - \frac{x_{n-1}^2 - a}{2x_{n-1}} = x_{n-1} - \frac{x_{n-1}}{2} + \frac{a}{2x_{n-1}} = \frac{1}{2}\left(x_{n-1} + \frac{a}{x_{n-1}}\right),$$

hier erhalten wir also die aus der Analysis I bekannte Formel, die HERON bereits rund sechzehn Jahrhunderte vor NEWTON benutzte, und von der wir gesehen haben, daß sie schnell gute Ergebnisse liefert.

Auch bei komplizierteren Polynomen hat sich das NEWTON-Verfahren in der Praxis sehr bewährt. Um zu verstehen, warum das so ist, betrachten wir die Funktion

$$\varphi(x) = x - \frac{f(x)}{f'(x)}\,,$$

die für alle x mit $f'(x) \neq 0$ definiert ist und die uns zu einem Iterationswert x_n den Folgewert x_{n+1} liefert. Offensichtlich ist z genau dann eine einfache Nullstelle von f, wenn $\varphi(z) = z$ ist; die einfachen Nullstellen von f sind also genau die Fixpunkte von φ.

Angenommen, wir haben uns einem solchen Fixpunkt bis auf die Distanz h genähert, d.h. wir haben ein $x_n = z + h$. Wir wollen abschätzen, wie weit dann $x_{n+1} = \varphi(x_i)$ von z entfernt ist.

Falls h klein ist, können wir auch φ ohne großen Fehler durch seine Linearisierung ersetzen:

$$x_{n+1} = \varphi(z+h) \approx \varphi(z) + h\varphi'(z) = z + h\varphi'(z)\,.$$

Die Ableitung von φ können wir leicht nach der Quotientenregel berechnen:

$$\varphi(x) = x - \frac{f(x)}{f'(x)} \Longrightarrow \varphi'(x) = 1 - \frac{f'(x)^2 - f(x)f''(x)}{f'(x)^2} = 1 - 1 + \frac{f(x)f''(x)}{f'(x)^2} = \frac{f(x)f''(x)}{f'(x)^2}\,.$$

Somit ist

$$x_{n+1} \approx z + h\frac{f(z)f''(z)}{f'(z)^2} = z\,,$$

da $f(z)$ verschwindet.

Dieses Ergebnis war, wenn man ein bißchen nachdenkt, natürlich zu erwarten; es ist aber völlig nutzlos, um den Abstand zwischen x_{i+1} und z zu bestimmen. Wenn wir ein nützliches Ergebnis erhalten wollen, dürfen wir uns daher nicht auf eine lineare Approximation beschränken, sondern müssen zumindest auch noch den quadratischen Term berücksichtigen.

Wir gehen daher aus von der Approximation

$$\varphi(x+h) \approx \varphi(x) + h\varphi'(x) + \tfrac{1}{2}h^2\varphi''(x)\,,$$

$\varphi''(x)$ ist die Ableitung von $\varphi'(x) = \frac{f(x)f''(x)}{f'(x)^2}$, also ist nach der Quotientenregel

$$\varphi''(x) = \frac{f'(x)^2\big(f'(x)f''(x) + f(x)f'''(x)\big)}{f'(x)^4}\,.$$

Speziell für $x = z$, wo $f(z)$ verschwindet und $\varphi(z) = z$ ist, erhalten wir die Abschätzung

$$\varphi''(z) = \frac{f''(z)}{f'(z)} \quad \text{und} \quad \varphi(z+h) \approx z + \frac{h^2}{2}\frac{f''(z)}{f'(z)}\,.$$

Der Abstand des neuen Iterationswert zur Nullstelle z ist also für kleine Werte von h bis auf einen nur von z abhängigen Vorfaktor gleich dem Quadrat des alten und verkleinert sich somit bei kleinen Werten von h sehr schnell.

Leider ist diese Aussage nicht so konkret, daß wir für irgendeinen vorgegebenen Startwert entscheiden können, ob und gegebenenfalls wohin das NEWTON-Verfahren konvergiert, denn wir wissen nicht, wann der

Abstand h „klein“ ist oder wird. Betrachten wir dazu als Beispiel das Polynom $f(x) = x^3 - 5x = x(x^2 - 5)$; seine Nullstellen sind offensichtlich $x = 0$ und $x = \pm\sqrt{5}$. Die zu iterierende Funktion ist hier

$$\varphi(x) = x - \frac{f(x)}{f'(x)} = x - \frac{x^3 - 5x}{3x^2 - 5} .$$

Für $x_0 = 1$ ist daher

$$x_1 = \varphi(1) = 1 - \frac{-4}{-2} = -1 \quad \text{und} \quad x_2 = \varphi(-1) = -1 - \frac{4}{-2} = 1 .$$

Damit ist klar, daß x_n für alle geraden n gleich eins ist und für die ungeraden -1; mit Startwert $x_0 = 1$ (oder -1) bekommen wir also nie ein nützliches Ergebnis.

Die folgende Tabelle zeigt, was passiert, wenn wir x_0 leicht vergrößern. (Die Werte der x_i sind zwar nur mit fünf geltenden Ziffern angegeben, die Iterationen wurden aber mit hundertstelliger Genauigkeit gerechnet.)

x_0	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}
$1+10^{-1}$	-1,9431	-2,3191	-2,2403	-2,2361	-2,2361	-2,2361	-2,2361	-2,2361	-2,2361	-2,2361	-2,2361	-2,2361
$1+10^{-2}$	-1,0623	1,4853	4,0498	3,0053	2,4569	2,2627	2,2365	2,2361	2,2361	2,2361	2,2361	2,2361
$1+10^{-3}$	-1,0060	1,0370	-1,2570	15,310	10,280	6,9631	4,8073	3,4540	2,6766	2,3254	2,2410	2,2361
$1+10^{-4}$	-1,0006	1,0036	-1,0220	1,1435	-2,7749	-2,3610	-2,2453	-2,2361	-2,2361	-2,2361	-2,2361	-2,2361
$1+10^{-5}$	-1,0001	1,0004	-1,0022	1,0131	-1,0826	1,7096	2,6521	2,3171	2,2401	2,2361	2,2361	2,2361
$1+10^{-6}$	-1,0000	1,0000	-1,0002	1,0013	-1,0078	1,0483	-1,3532	-10,050	-6,8125	-4,7108	-3,3956	-2,6462
$1+10^{-7}$	-1,0000	1,0000	-1,0000	1,0001	-1,0008	1,0047	-1,0286	1,1920	-4,5915	-3,3238	-2,6095	-2,3035
$1+10^{-8}$	-1,0000	1,0000	-1,0000	1,0000	-1,0001	1,0005	-1,0028	1,0170	-1,1090	2,0814	2,2552	2,2363

Das Verfahren konvergiert offensichtlich gegen eine der beiden Nullstellen $\sqrt{5}$ oder $-\sqrt{5}$, wobei keine Regel erkennbar ist, wann es gegen welche der beiden konvergiert. Wenn wir dieselbe Rechnung ausführen für die Startwerte $1 - 10^{-i}$, erhalten wir ein langweiligeres Ergebnis: Nun konvergiert das Verfahren stets gegen die dritte Nullstelle Null.

Der Startwert $x_0 = 1$ liegt also in der Nähe der Einzugsbereiche aller drei Nullstellen, was verständlich macht, daß dort Probleme auftreten.

Ein ähnliches Problem bekommen wir, wenn wir das NEWTON-Verfahren anwenden zur Nullstellenbestimmung des Polynoms $f(x) = x^2 + 2$: Hier ist

$$x_{n+1} = x_n - \frac{x_n^2 + 2}{2x_n} = \frac{1}{2}\left(x_n - \frac{2}{x_n}\right)$$

für reelles x_n natürlich auch wieder reell, die Folge der x_n kann also für keinen reellen Startwert x_0 gegen eine der Nullstellen $\pm\sqrt{-2}$ konvergieren. Für $x_0 = 1$ beispielsweise erhalten wir die Folge

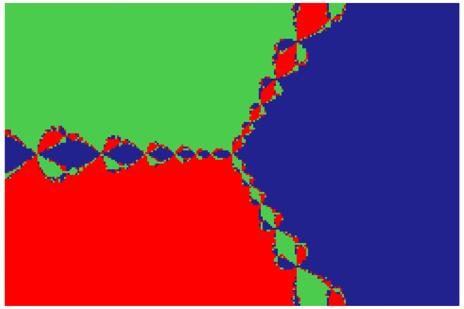
$$-\tfrac{1}{2},\ 1\tfrac{3}{4},\ 0{,}30357,\ -3{,}14233,\ -1{,}25293,\ 0{,}17166,\ -5{,}73953,\ \ldots .$$

Wenn wir allerdings mit $x_0 = i$ oder z.B. mit $x_0 = 1 + 2i$ beginnen, haben wir bereits nach wenigen Iterationen ein Ergebnis mit zehn korrekten Nachkommastellen sowohl im Realteil als auch in Imaginärteil.

Das folgende Bild zeigt das Verhalten des NEWTON-Verfahrens im Komplexen anhand des Polynoms $f(x) = x^3 - 1$, von dem wir aus der Analysis I wissen, daß es die drei komplexen Nullstellen

$$1, \quad \rho = -\frac{1}{2} + \frac{\sqrt{3}}{2}i \quad \text{und} \quad \overline{\rho} = -\frac{1}{2} - \frac{\sqrt{3}}{2}i$$

hat. Bei allen blau gezeichneten Startwerten konvergiert das Verfahren gegen eins, für die grünen gegen ρ und für die roten gegen $\overline{\rho}$. In der Umgebung jeder Nullstelle haben alle Punkte deren Farbe, an den Grenzen gibt es eine komplizierte (fraktale) Struktur. Treffpunkt der drei Grenzgebiete ist der Nullpunkt, der nicht als Startpunkt genommen werden kann, da $f'(x) = 3x^2$ dort verschwindet.



Hauptanwendung des NEWTON-Verfahrens ist allerdings die Bestimmung reeller Nullstellen. Wie wir gesehen haben, gibt es auch hier keine Garantie, daß es mit einem vorgegebenen Startwert wirklich gegen eine Nullstelle konvergiert; es gibt allerdings eine ganze Reihe von Untersuchungen, die auf hinreichende Kriterien führen. Ein einfaches Beispiel ist das folgende:

f sei ein Polynom und $a \le b$ seien zwei reelle Zahlen, für die $f(a) \cdot f(b) < 0$ ist. Dann haben $f(a)$ und $f(b)$ verschiedene Vorzeichen, also muß es zwischen a und b mindestens eine Nullstelle von f geben. Weiterhin sei $f'(x)$ entweder positiv für alle x mit $a \le x \le b$ oder aber negativ für alle solche x. Damit steigt oder fällt der Graph von $f(x)$ zwischen a und b monoton, so daß es dort nur *eine* Nullstelle geben kann. Außerdem folgt, daß das Maximum von $|f(x)|$ im Intervall $[a, b]$ in einem der beiden Endpunkte angenommen wird; diesen Endpunkt bezeichnen wir mit c. Um zu verhindern, daß die Tangentensteigungen zu sehr schwanken und uns die Iterationen aus dem Intervall $[a, b]$ hinausführen, verlangen wir außerdem noch, daß $f''(x)$ dort entweder überall nichtnegativ ist (d.h. der Graph ist konkav) oder überall nichtpositiv (Graph konvex), und wir fordern, daß

$$\left|\frac{f(c)}{f'(c)}\right| \le b - a$$

ist; daraus folgt, daß die Tangente im Punkt $\bigl(c, f(x)\bigr)$ die x-Achse im Intervall $[a, b]$ schneidet. Dann zeigt eine nicht sehr aufwendige Rechnung, daß das NEWTON-Verfahren für jeden Startwert x_0 zwischen a und b gegen die eindeutig bestimmte Nullstelle in $[a, b]$ konvergiert.

Ähnliche und auch sehr viel allgemeinere Aussagen findet man (mit Beweisen) in praktisch jedem Lehrbuch der Numerischen Mathematik; allgemein läßt sich sagen, daß sich das NEWTON-Verfahren in der Praxis fast immer sehr gut verhält, daß sich aber wirklich allgemeine theoretische Aussagen nur schwer beweisen lassen.

Selbstverständlich läßt sich das NEWTON-Verfahren auch auf mehrdimensionale Probleme anwenden: Wenn wir eine differenzierbare Funktion $f\colon D \to \mathbb{R}^n$ haben mit $D \subseteq \mathbb{R}^n$, können wir sie in der Nähe eines Punktes $x_0 \in D$ annähern durch die lineare Funktion

$$\ell(x) = f(x_0) + J_f(x_0) \cdot (x - x_0)\,.$$

Diese Funktion verschwindet genau dann, wenn x eine Lösung des linearen Gleichungssystems

$$J_f(x_0) \cdot x = J_f(x_0) \cdot x_0 - f(x_0)$$

ist. Falls die JACOBI-Matrix $J_f(x_0)$ invertierbar ist, hat dieses Gleichungssystem genau eine Lösung, nämlich $x = x_0 - J_f(x_0)^{-1} f(x_0)$.

Daher können wir auch hier eine Iteration definieren durch

$$x_n = x_{n-1} - J_f(x_n)^{-1} f(x_{n-1})$$

– immer vorausgesetzt, die Matrizen $J_f(x_n)$ werden nie singulär.

d) Der Banachsche Fixpunktsatz

In den beiden letzten Abschnitten hatten wir jeweils Aussagen darüber, daß eine Folge $(x_k)_{k\in\mathbb{N}}$ mit $x_k = f(x_{k-1})$ unter gewissen Voraussetzungen gegen einen vorgegebenen Fixpunkt x von f konvergiert, wenn der Startwert x_0 hinreichend nahe bei x liegt. Es sagt uns aber weder, *wie* nahe das in einem konkreten Fall sein muß, noch sagt es uns, ob es überhaupt einen Fixpunkt gibt. In diesem Abschnitt geht es um einen Satz, der sowohl die Existenz eines Fixpunkts als auch die Konvergenz dorthin unabhängig von Startwert garantiert. Die Voraussetzungen sind natürlich deutlich stärker, was die Anwendbarkeit etwas einschränkt; trotzdem ist der Satz von zentraler Bedeutung sowohl für die reine als auch die angewandte Mathematik. Wir werden ihn daher auch relativ allgemein formulieren und beweisen; auch wenn er uns im Augenblick vor allem für Funktionen auf Teilmengen des$\mathbb{R}^n$ interessiert.

Definition: V sei ein normierter Vektorraum und $X \subseteq V$. Eine Abbildung $f\colon X \to V$ heißt *kontrahierend,* wenn es eine reelle Zahl $q < 1$ gibt, so daß für alle $x, y \in X$ gilt: $\|f(y) - f(x)\| \le q\,\|y - x\|$.

Banachscher Fixpunktsatz: V sei ein normierter Vektorraum, $X \subseteq V$ eine abgeschlossene Teilmenge, und $f\colon X \to V$ eine kontrahierende Abbildung mit $f(X) \subseteq X$. Dann hat f genau einen Fixpunkt $x^* \in X$, und für jedes $x_0 \in X$ konvergiert die Folge $(x_n)_{n\in\mathbb{N}}$ mit $x_n = f(x_{n-1})$ gegen x^*.

Beweis: $q < 1$ sei die Konstante, für die $\|f(y) - f(x)\| \le q\,\|y - x\|$ ist für alle $x, y \in X$. Wir zeigen als erstes, daß es *höchstens* einen Fixpunkt gibt: Ist $x^* = f(x^*)$ und $y^* = f(y^*)$, so ist

$$\|y^* - x^*\| = \|f(x^*) - f(y^*)\| \le q\,\|x^* - y^*\|\ .$$

Das ist aber nur möglich, wenn $\|y^* - x^*\| = 0$ ist, also $x^* = y^*$.

Als nächstes wollen wir sehen, daß die Folge der x_n für jeden Startwert eine CAUCHY-Folge ist. Falls $f(x_0) = x_0$ ist, haben wir eine konstante

Folge, und alles ist klar. Andernfalls müssen wir Differenzen von Folgengliedern abschätzen; sei also n eine natürliche Zahl und $m = n + k$ für ein $k \in \mathbb{N}$. Dann ist nach der Dreiecksungleichung

$$\|x_m - x_n\| = \|x_{n+k} - x_n\| = \left\| \sum_{j=0}^{k-1} \big((x_{n+j+1} - x_{n+j}\big) \right\| \le \sum_{j=1}^{k} \left\|x_{n+j+1} - x_{n+j}\right\| .$$

Da f kontrahierend ist, folgt weiter, daß für jedes $\ell \in \mathbb{N}$ gilt

$$\|x_{\ell+1} - x_\ell\| \le q \left\|x_\ell - x_{\ell-1}\right\| \le q^2 \left\|x_{\ell-1} - x_{\ell-2}\right\| \le \cdots \le q^\ell \left\|x_1 - x_0\right\| .$$

Somit ist nach der Summenformel für die geometrische Reihe

$$\|x_m - x_n\| \le \sum_{j=0}^{k-1} q^{n+j} \|x_1 - x_0\| = \frac{q^n - q^m}{1-q} \|x_1 - x_0\| \le q^n \frac{\|x_1 - x_0\|}{1-q} .$$

Wenn wir uns ein $\varepsilon > 0$ vorgeben, ist somit $\|x_m - x_n\| < \varepsilon$, falls

$$q^n \frac{\|x_1 - x_0\|}{1-q} < \varepsilon \quad \text{oder} \quad q^n < \frac{(1-q)\varepsilon}{\|x_1 - x_0\|} .$$

Da rechts eine positive Zahl steht und die Folge der Potenzen von q eine Nullfolge ist, gibt es ein $N \in \mathbb{N}$, so daß diese Ungleichung für alle $n \ge N$ erfüllt ist; $(x_n)_{n\in\mathbb{N}}$ ist also eine CAUCHY-Folge.

Da V als BANACH-Raum vollständig ist, konvergiert diese Folge gegen einen Grenzwert $x^* \in V$. Da alle x_n in der abgeschlossenen Menge X liegen, liegt auch der Grenzwert dort, d.h. $x^* \in V$. Wir wollen uns überlegen, daß x^* ein Fixpunkt von f ist:

Für jedes $\varepsilon > 0$ gibt es ein $M \in \mathbb{N}$, so daß $\|x^* - x_n\| < \frac{1}{3}\varepsilon$ für alle $n \ge M$. Damit ist auch

$$\|f(x^*) - f(x_n)\| \le q \|x^* - x_n\| \le \frac{q\varepsilon}{3} < \frac{\varepsilon}{3} .$$

Außerdem gibt es, da wir eine CAUCHY-Folge haben, ein $N \ge M$, so daß $\left\|x_n - x_{n-1}\right\| < \frac{1}{3}\varepsilon$ für alle $n, m \ge N$. Damit ist

$$\begin{aligned}\|f(x^*) - x^*\| &= \left\| \big(f(x^*) - f(x_n)\big) + \big(f(x_n) - x_n\big) + \big(x_n - x^*\big) \right\| \\ &\le \|f(x^*) - x_{n+1}\| + \|x_{n+1} - x_n\| + \|x_n - x^*\| \\ &\le \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon .\end{aligned}$$

Da dies für jedes $\varepsilon > 0$ gilt, muß $\|f(x^*) - x^*\| = 0$ sein, und das gilt nur, falls $f(x^*) = x^*$ ist. Damit haben wir einen Fixpunkt in X gefunden; es gibt also genau diesen einen Fixpunkt, und für jeden Startwert x_0 konvergiert die Folge der $(x_n)_{n\in\mathbb{N}}$ gegen x^*.

■

e) Konvergenz von Funktionenfolgen

Im nächsten Kapitel werden wir zur Definition mehrdimensionaler Integrale auch Folgen von Funktionen betrachten. Zur Vorbereitung wollen wir hier bereits einige allgemeinere Tatsachen betrachten, für die kein Integralbegriff notwendig ist.

Seien zunächst $f_n: D \to \mathbb{R}$ irgendwelche Funktionen, die auf einer Teilmenge $D \subseteq \mathbb{R}^n$ definiert sind, und sei $f: D \to \mathbb{R}^m$ eine weitere Funktion. Für jedes $x \in D$ haben wir dann eine Folge $\big(f(x_n)\big)_{n\in\mathbb{N}}$ von Elementen aus $\mathbb{R}^n$ und können fragen, ob und gegebenenfalls wohin diese Folge konvergiert.

Definition: Eine Folge $(f_n)_{n\in\mathbb{N}}$ von Funktionen $f_n: D \to \mathbb{R}$ konvergiert auf der Teilmenge $A \subseteq D$ *punktweise* gegen die Funktion $f: D \to \mathbb{R}$, wenn für alle $x \in A$ gilt $\lim_{n\to\infty} f_n(x) = f(x)$.

Als Beispiel betrachten wir die Funktionen $f_n: \mathbb{R} \to \mathbb{R}$ mit $f_n(x) = \cos^n x$. Für jedes $x \in \mathbb{R}$ haben wir dann die Folge $(cos^n x)_{n\in\mathbb{N}}$. Für $|\cos| x < 1$ ist das bekanntlich eine Nullfolge, für $\sin x = 1$ haben wir die konstante Folge $(1^n)_{n\in\mathbb{N}}$, die gegen 1 konvergiert, und falls $\cos x = -1$ divergiert die Folge. Bekanntlich ist $\cos x = 1$ genau dann, wenn x ein ganzzahliges Vielfaches von 2π ist, und $\cos x = -1$ für alle ungeradzahligen Vielfachen von π. Setzen wir also

$$A = \mathbb{R} \setminus \left\{ (2k+1)\pi \;\middle|\; k \in \mathbb{Z} \right\} ,$$

so konvergiert die Folge der f_n auf A punktweise gegen die Funktion

$$f\colon \begin{cases} A \to \mathbb{R} \\ x \mapsto \begin{cases} 1 & \text{falls } x = 2k\pi \text{ für ein } k \in \mathbb{Z} \\ 0 & \text{sonst} \end{cases} \end{cases} .$$

Obwohl die Funktionen f_n allesamt stetig sind, ist die Grenzfunktion unstetig bei allen ganzzahligen Vielfachen von 2π; bei den ungeradzahligen Vielfachen von π ist sie sogar nicht einmal definiert.

Wir können bei der Definition der Konvergenz aber auch anders vorgehen: Wir betrachten für eine beliebige Teilmenge $D \subset \mathbb{R}^n$ die Menge $C_b(D, \mathbb{R})$ aller beschränkter Funktionen $f\colon D \to \mathbb{R}$, d.h. also die Menge aller Funktionen, für die es eine reelle Zahl M gibt, so daß $|f(x)| \le M$ ist für alle $x \in D$.

Offensichtlich ist $C_b(D, \mathbb{R})$ ein Vektorraum, denn die Nullfunktion ist beschränkt, und für zwei beschränkte Funktionen f, g mit Schranken M, N und zwei reelle Zahlen a, b ist

$$|af(x) + bg(x)| \le |a| \cdot |f(x)| + |b|\,|g(x)| \le M\,|a| + N\,|b| \ .$$

Für jedes $f \in C_b(D, \mathbb{R})$ existiert

$$\|f\|_\infty \underset{\text{def}}{=} \sup\bigl\{|f(x)| \bigm| x \in D\bigr\} ,$$

da die rechtsstehende Menge beschränkt ist. Wie bei der Maximumsnorm auf $\mathbb{R}^n$ rechnet man leicht nach, daß dies eine Norm auf $C_b(D, \mathbb{R})$ definiert, die sogenannte Supremumsnorm.

Wenn eine Folge $(f_n)_{n\in\mathbb{N}}$ von Funktionen aus $C_b(D, \mathbb{R})$ bezüglich dieser Norm gegen eine Funktion $f \in C_b(D, \mathbb{R})$ konvergiert, gibt es zu jedem $\varepsilon > 0$ ein $N \in \mathbb{N}$, so daß

$$\|f - f_n\|_\infty = \sup\bigl\{|f(x)| \bigm| x \in D\bigr\} < \varepsilon \quad \text{für alle } n \ge N ;$$

insbesondere ist also für *jedes* $x \in D$ der Betrag von $f(x) - f_n(x)$ kleiner als ε für alle $n \ge N$. Das ist eine stärkere Eigenschaft als bei der punktweisen Konvergenz: Dort reicht es, wenn für jedes $\varepsilon > 0$ und jedes $x \in D$ ein $N \in \mathbb{N}$ existiert, so daß $|f(x) - f_n(x)| < \varepsilon$ für alle $n \ge N$; das N darf also von x abhängen.

Definition: Eine Folge $(f_n)_{n\in\mathbb{N}}$ von Funktionen $f_n\colon D \to \mathbb{R}$ konvergiert *gleichmäßig* gegen die Funktion $f\colon D \to \mathbb{R}$ auf der Menge $A \subseteq D$, wenn es zu jedem $\varepsilon > 0$ ein $N \in \mathbb{N}$ gibt, so daß $|f(x) - f_n(x)| < \varepsilon$ für alle $x \in A$ und alle $n \ge N$.

Die obige Folge der Funktionen $\cos^n x$ konvergiert nicht gleichmäßig gegen ihre Grenzfunktion: Andernfalls müßte sie insbesondere auf $\mathbb{R} \setminus \mathbb{Z}\pi$ gleichmäßig gegen die Nullfunktion konvergieren, d.h. zu jedem $\varepsilon > 0$ müßte es ein $N \in \mathbb{N}$ geben, so daß $|\cos^n x| < \varepsilon$ ist für alle $n \ge N$. Angenommen, es gäbe so ein N zu $\varepsilon = \frac{1}{2}$. Für $n \ge N$ wäre dann $|\cos^n x| < \frac{1}{2}$ für alle x, die keine ganzzahligen Vielfachen von π sind. Andererseits ist aber $|\sin^n x|$ eine auf ganz $\mathbb{R}$ stetige Funktion, die bei allen ganzzahligen Vielfachen von π den Wert eins annimmt und somit in einer gewissen Umgebung dieser Punkte nur Werte annimmt, die größer sind als $\frac{1}{2}$.

Bei einer gleichmäßig konvergenten Folge kann es nicht passieren, daß die Grenzfunktion einer Folge stetiger Funktionen unstetig wird:

Lemma: $D \subseteq \mathbb{R}^n$ sei eine offene Menge und $f_n\colon D \to \mathbb{R}$ seien stetige Funktionen. Falls die Folge $(f_n)_{n\in\mathbb{N}}$ gleichmäßig gegen eine Funktion $f\colon D \to \mathbb{R}$ konvergiert, ist auch f stetig.

Beweis: Wir müssen zeigen, daß es zu jedem $x \in D$ und jedem $\varepsilon > 0$ ein $\delta > 0$ gibt, so daß $|f(y) - f(x)| < \varepsilon$ ist für alle $y \in D$ mit $\|y - x\| < \delta$, wobei $\|\cdot\|$ irgendeine Norm auf $\mathbb{R}^n$ bezeichnet.

Wegen der gleichmäßigen Konvergenz der Folge $(f_n)_{n\in\mathbb{N}}$ gibt es zunächst ein $N \in \mathbb{N}$, so daß $|f(y) - f_n(y)| < \frac{1}{3}\varepsilon$ ist für alle $n \ge N$ und alle $y \in D$. Wir wählen irgendein solches n. Die Funktion f_n ist Voraussetzung stetig; daher gibt es ein $\delta > 0$, so daß $|f_n(y) - f_n(x)| < \frac{1}{3}\varepsilon$ ist für alle $y \in D$ mit $|y - x| < \delta$. Für diese y ist dann auch

$$\begin{aligned} |f(y) - f(x)| &= \bigl|\bigl(f(y) - f_n(y)\bigr) + \bigl(f_n(y) - f_n(x)\bigr) + \bigl(f_n(x) - f(x)\bigr)\bigr| \\ &\le |f(y) - f_n(y)| + |f_n(y) - f_n(x)| + |f_n(x) - f(x)| \\ &< \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon \,. \end{aligned}$$

Damit ist die Stetigkeit von f bewiesen. ■