Kapitel 3

Der Satz von Kantorovich

> Ich kann die Bewegungen der Himmelskörper
> berechnen, aber nicht die Verrücktheit der Menschen
>
> Isaac Newton

Wir betrachten hier das Newton-Verfahren im unendlichdimensionaleen Kontext. Nach einer Einführung bereiten wir den Beweis des Satzes von Kantorovich vor durch die Diskussion eines quadratischen Polynoms. Es dient für den Beweis als majorisierendes Polynom für die Abschätzungen des Fehlers der Iterierten. Im Anhang stellen wir eine Variante des Satzes bereit, die mit etwas abgeschwächten Voraussetzungen auskommt.

3.1 Nullstellensuche nach Newton

Iterative Verfahren lassen sich in die Antike zurückverfolgen: einschlägige Namen hierzu sind: Euklid, Archimedes, Theon von Alexandrien. Iterative Verfahren schreibt man „meist“ in Form eines Algorithmus auf[1]. Der erste *europäische Mathematiker*, der iterative Verfahren nutzte, scheint Vieta (1540-1603) gewesen zu sein. Er nutzte die Störungsrechnung für die Lösung skalarer Gleichungen. Isaac Newton entdeckte Vieta's Methode, beschreibt[2] ein Rechenverfahren zum Lösen einer polynomialen Gleichung und begründet damit ein Verfahren, das heutzutage als **Newton-Verfahren** bezeichnet wird. Er tut dies am Beispiel des Polynoms $p(x) := x^3 - 2x - 5 = 0$. Eine leicht zu erratende Näherung ist $x_0 = 2$, denn $p(2) = -1$ ist „klein“. Newton machte den Ansatz $x = x_0 + u = 2 + u$ mit einem als „klein“ angenommenen u und setzte diesen Ansatz in die Gleichung ein. Es gilt:

$$x^3 = (2+u)^3 = 8 + 12u + 6u^2 + u^3\,,\ 2x = 2(2+u) = 4 + 2u\,.$$

Also folgt

$$x^3 - 2x - 5 \;=\; -1 + 10u + 6u^2 + u^3 \stackrel{!}{=} 0\,.$$

Da u als „klein“ angenommen wurde, können die Terme höherer Ordnung gegen den linearen und konstanten Anteil vernachlässigt werden, womit $10u - 1 = 0$ bzw. $u = 0.1$ übrig bleibt. Als Näherung x_1 resultiert $x_1 = 2.1$.

Wir können nun dieses Vorgehen wiederholen: wir setzen $u = 0.1 + v$ an, betrachten die Gleichung $p(2 + 0.1 + v) = 0$, berücksichtigen wiederum nur den linearen Anteil und erhalten so $v = -0.061/11.23 = -0.0054\ldots$. Als Näherung x_2 resultiert $x_2 = 2.0946$.

[1] Die Bezeichnung leitet sich aus dem Namen **Al–Khwarazmi** ab, einem der bedeutensten Mathematikers des anfangenden Mittelalters.

[2] Isaac Newton, 1643–1727; „Methodus fluxionum et serierum infinitarum“; siehe [13]

Raphson[3] beschrieb diesen Rechenprozess formal und illustrierte den Formalismus an der allgemeinen Gleichung 3. Grades, die abstrakte Form des Verfahrens mit Benutzung von Ableitungen stammt von Thomas Simpson (1710-1761). Simpson und Fourier (1768-1830) brachten es in eine Form, die in etwa unserer heutigen Formulierung entspricht.

Sei $f : \mathbb{R} \longrightarrow \mathbb{R}$. Eine Nullstelle wird nach folgendem Vorgehen gesucht:

(1) Man rät eine Näherung x_0. O.E. $f(x_0) \neq 0$.
(2) Man berechnet/zeichnet die Tangente T_0 an den Graphen von f im Punkt $(x_0, f(x_0))$.
(3) Man berechnet/konstruiert die Nullstelle x_1 der Tangente T_0.
(4) Man setzt $x_0 := x_1$ und wiederholt den Vorgang, beginnend bei (1).

Klar, um die Tangente bestimmen zu können, müssen wir voraussetzen, dass diese existiert, was die Differenzierbarkeit von f voraussetzt. Dann lautet die Tangentengleichung

$$T_0 \,:\, y = f(x_0) + f'(x_0)(x - x_0) \tag{3.1}$$

und die Berechnung der Nullstelle von T_0 führt zur Formel

$$x_1 = x_0 - f'(x_0)^{-1} f(x_0)\,. \tag{3.2}$$

Hier tritt das Problem auf, dass $f'(x_0) \neq 0$ gelten muss, d. h. dass f in $(x_0, f(x_0))$ keine waagrechte Tangente besitzt. Von der Anschauung her, keine überraschende Forderung, von der Analyse des Verfahrens her eine Forderung, die sukzessive oder a-priori sichergestellt werden muss.

Schreiben wir das Verfahren nun kompakt auf:

$$x_{n+1} := x_n - f'(x_n)^{-1} f(x_n)\,,\; n \in \mathbb{N}_0\,. \tag{3.3}$$

Dabei ist die **Startnäherung** x_0 geeignet zu wählen. Wir nennen dieses Vorgehen nun **Newton–Verfahren**; siehe Abbildung 3.1.

Das Newton–Verfahren ist ein so genanntes lokal konvergentes Verfahren. Konvergenz der in der Newton–Iteration erzeugten Folge zu einer Nullstelle ist also nur garantiert, wenn der Startwert schon „ausreichend nahe“ an der Nullstelle liegt. Ist der Startwert nicht gut genug, so haben wir zu rechnen mit:

- Die Folge divergiert, der Abstand zur Nullstelle wächst über alle Grenzen.
- Die Folge divergiert, bleibt aber beschränkt. Sie kann z.B. periodisch werden, d. h. endlich viele Punkte wechseln sich in immer derselben Reihenfolge ab. Man sagt auch, dass die Folge oszilliert (Bei $f(x) := x^3 - 2x + 2$ ist dies machbar).
- Die Folge konvergiert, falls die Funktion mehrere Nullstellen hat, gegen eine andere als die gewünschte Nullstelle konvergieren; in der Abbildung 3.1 kann man dies erahnen.

Ist der Startwert x_0 so gewählt, dass das Newton–Verfahren konvergiert, so ist die Konvergenz allerdings quadratisch, also mit der Konvergenzordnung 2 (falls die Ableitung an der Nullstelle nicht verschwindet).

[3]Joseph Raphson, 1648–1715; Arbeit „Analysis Aequationum universalis“

Bemerkung 3.1 *Wie ordnet sich das Newtonsche Vorgehen hier nun ein? Ausgehend von der Startnäherung $x_0 = 2$ wird ein Newtonschritt auf die Nullstellengleichung $p(x+2) = 0$ mit $x = 0$ als Startnäherung angewendet:*

$$x_1 := 0 - \frac{p(2)}{p'(2)} = \frac{1}{10}\,.$$

Nun betrachtet man die Nullstellengleichung $p(x + 2.1) = 0$ mit $x = 0$ als Startnäherung und wendet wieder einen Newtonschritt mit Ausgangsnäherung $x = 0$ an:

$$x_2 := 0 - \frac{p(2.1)}{p'(2.1)} = \frac{0.061}{11.23}\,.$$

Und so weiter! □

Viele nichtlineare Gleichungen haben mehrere Lösungen, so hat ein Polynom n-ten Grades bis zu n Nullstellen. Will man alle Nullstellen in einem bestimmten Bereich $D \subset \mathbb{R}$ ermitteln, so muss zu jeder Nullstelle ein passender Startwert in D gefunden werden, für den das Newton–Verfahren konvergiert. Ein beliebtes Vorgehen dazu besteht in Einschachtelungsverfahren: zwischen zwei Punkten z_1, z_2, so dass $f(z_1), f(z_2)$ unterschiedliche Vorzeichen besitzen, liegt immer eine Nullstelle von f, da wir ja Differenzierbarkeit von f (und damit Stetigkeit) voraussetzen.

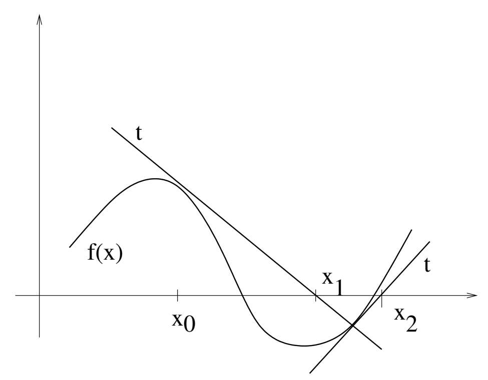

Abbildung 3.1: Newtonverfahren

Beispiel 3.2 *Ein Spezialfall des Newtonschen Näherungsverfahrens ist das Babylonische Wurzelziehen, auch bekannt als Heronverfahren nach Heron von Alexandria: Wendet man das Verfahren zur Nullstellenbestimmung auf die Funktion $f(x) := x^2 - a\,(a > 0)$ an, so erhält man wegen der Ableitungsfunktion $f'(x) = 2x$ für die Lösung $\sqrt{a}$ das Näherungsverfahren*

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

Dieses Verfahren konvergiert für jedes $a \geq 0$ und für jeden beliebigen Anfangswert $x_0 > 0$. □

Beispiel 3.3 *Die Quadratwurzel einer Zahl $a > 0$ sind die Nullstellen der Funktion $f(x) := 1 - a/x^2$. Diese Funktion hat die Ableitung $f'(x) = 2a/x^3$, die Newton-Iteration erfolgt also nach der Vorschrift*

$$x_{n+1} := x_n - \frac{(x_n)^3}{2a} + \frac{x_n}{2} = \frac{x_n}{2}\left(3 - \frac{(x_n)^2}{a}\right)\,.$$

Der Vorteil dieser Vorschrift gegenüber dem Wurzelziehen nach Heron (siehe Beispiel 3.2) ist, dass es divisionsfrei ist, sobald einmal der Kehrwert von a bestimmt wurde. Als Startwert wurde in der Tabelle $x_0 := (1 + a)/2$ gewählt. Die Iterierten wurden an der ersten ungenauen Stelle abgeschnitten. Es ist zu erkennen, dass nach wenigen Schritten die Anzahl gültiger Stellen schnell *wächst.*

n	x_n *bei* $a=2$	x_n *bei* $a=3$	x_n *bei* $a=5$
0	$1,5$	2	3
1	$1,40$	$1,6$	$1,8$
2	$1,4141$	$1,72$	$2,1$
3	$1,41421355$	$1,73203$	$2,22$
4	$1,41421356237309502$	$1,7320508074$	$2,23601$
5	$1,4142135623730950488016887242096\,97$	$1,73205080756887729351$	$2,236067975$

□

Das Newton-Verfahren gilt als ein sehr effizientes Verfahren (in den Naturwissenschaften und anderswo). Worin ist dies begründet, obwohl das Problem der guten Startnäherung und die Tatsache, dass eine Ableitung ausgerechnet werden muss, schwer wiegen? Es liegt an vier Beobachtungen, die in der Literatur ausreichend diskutiert wurden und immer noch werden:

(1) Das Verfahren hat eine naheliegende Erweiterung auf Aufgaben in mehreren Variablen. Im nächsten Abschnitt werden wir es sogar in unendlichdimensionalen Kontext betrachten.

(2) Das Verfahren konvergiert unter gut zu durchschaubaren Voraussetzung (siehe unten) quadratisch.

(3) Das Verfahren kann modifiziert werden, um die Berechnung der Ableitung in jedem Schritt zu vermeiden. Etwa durch:

$$x_{n+1} := x_n - f'(x_0)^{-1} f(x_n)\,,\ n = 0, \dots\,. \tag{3.4}$$

Allerdings ist dann die Konvergenzgeschwindigkeit schlechter.

(4) Das Verfahren kann globalisiert werden, d. h. man kann Vorkehrungen einbauen, die sicherstellen, dass das so abgeänderte Verfahren auch bei „schlechten“ Startwerten konvergiert; das Stichwort ist Schrittweitensteuerung:

$$x_{n+1} := x_n - s_n f'(x_n)^{-1} f(x_n)\,,\ n = 0, \dots\,. \tag{3.5}$$

Satz 3.4 *Sei $f : [a,b] \longrightarrow \mathbb{R}$ zweimal stetig differenzierbar und es gelte*

$$|f'(x)| \geq m\,,\ |f''(x)| \leq M \text{ für alle } x \in [a,b] \tag{3.6}$$

mit $m > 0, M > 0$. Dann gilt:

(a) f hat in $[a,b]$ höchstens eine Nullstelle.

(b) Ist z eine Nullstelle in (a,b), dann ist die Iteration (3.3) definiert für alle $x^0 \in U_r(z) := (z-r, z+r)$ wobei $r := \min(2mM^{-1}, b-z, z-a)$ ist.
Weiterhin gilt mit $q := M(2m)^{-1}|x^0 - z| < 1$ für alle $n \in \mathbb{N}$:

1. $|z - x_n| \leq \frac{M}{2m}|z - x_{n-1}|^2$ ***(Konvergenzordnung)***
2. $|z - x_n| \leq \frac{2m}{M} q^{2^n}$ ***(a priori Abschätzung)***
3. $|z - x_n| \leq \frac{1}{m}|f(x_n)| \leq \frac{M}{2m}|x_n - x_{n-1}|^2$ ***(a posteriori Abschätzung)***

Beweis:
Seien z^1, z^2 Nullstellen von f in $[a,b]$. Aus

$$0 = |f(z^1) - f(z^2)| = |f'(\eta)||z^1 - z^2|$$

erhalten wir $z^1 = z^2$ und $a)$ ist bewiesen.
Mit der Taylorentwicklung folgt

$$\begin{aligned} 0 &= f(z) = f(x_n) + f'(x_n)(z - x_n) + \frac{1}{2}f''(\eta)(z - x_n)^2 \,, \; \eta \in [a,b] \,, \\ 0 &= f(x_n) + f'(x_n)(x_{n+1} - x_n) \,, \end{aligned}$$

und wir erhalten mit Subtraktion

$$0 = (z - x_{n+1})f'(x_n) + \frac{1}{2}f''(\eta)(z - x_n)^2 \,.$$

Dies zeigt

$$|z - x_{n+1}| \le \frac{M}{2m}|z - x_n|^2.$$

Sei $x_0 \in (z - r, z + r)$. Dann folgt

$$|z - x_1| \le \frac{M}{2m}|z - x_0|^2 \le \frac{M}{2m}(\frac{2m}{M})^2 q^2 \,.$$

Mittels vollständiger Induktion erhalten wir die a priori Abschätzung 2..
Es gilt

$$|f(x_{n+1})| = |f(z) - f(x_{n+1})| = |f'(\eta)||z - x_{n+1}| \ge m|z - x_{n+1}|$$

und

$$f(x_{n+1}) = f(x_n - \frac{f(x_n)}{f'(x_n)}) = \frac{1}{2}f''(\xi)(x_{n+1} - x_n)^2$$

was die a posteriori Abschätzung impliziert. ■

Die 1. Abschätzung von (b) in Satz 3.4 besagt, dass die **Konvergenzordnung** der Folge $(x_n)_{n\in\mathbb{N}}$ (mindestens) zwei, also quadratisch ist. Man kann dies so formulieren, dass bei jedem Iterationsschritt die Anzahl der signifikanten Stellen der Approximation x_n sich verdoppelt.

Beispiel 3.5 *Betrachte die Funktion $f(x) := x^2, x \in \mathbb{R}$. Die Nullstelle $z := 0$ von f ist zweifach. Die Newton-Iteration mit Startwert $x_0 \neq 0$ ergibt*

$$x_{n+1} = \frac{1}{2}x_n \textit{ also } |x_{n+1} - z| = \frac{1}{2}|x_n - z| \,, \; n \in \mathbb{N}_0 \,,$$

und die Konvergenzrate ist nur linear.

Bei einer Nullstelle z mit Vielfachheit p einer Funktion f können wir die Iteration

$$x_{k+1} := x_k - p\frac{f(x_k)}{f'(x_k)}$$

betrachten und man kann beweisen, dass wieder quadratische Konvergenz gegen z gegeben ist. Aber die Iteration ist von wenig praktischem Wert, denn nur selten kennt man die Vielfachheit einer Nullstelle im Vorhinein. □

Bemerkung 3.6 *Newton's Methode kann als eine Fixpunktiteration betrachtet werden. Setze $g(x) := x + h(x)f(x), x \in [a,b]$, mit einer glatten Funktion h . Eine Nullstelle von f ist sicher ein Fixpunkt von g . Wir wählen $h(x) := -1/f'(x)$. Wegen $g'(z) = 0$ für jede einfache Nullstelle z von f ist die Kontraktionskonstante von g in einer Nullstelle z von f Null. Dies hat die quadratische Konvergenz der Fixpunktiteration zur Konsequenz.* □

Bemerkung 3.7 *Newton's Methode kann auch aus folgender Beobachtung abgeleitet werden. Ist x_n eine aktuelle Näherung für die Nullstelle der differenzierbaren Funktion $f : \mathbb{R} \longrightarrow \mathbb{R}$, so gilt*

$$\int_{x_n}^{x} f'(t)dt = f(x) - f(x_n)\,.$$

Nutzt man eine Quadraturformel zur Approximation des Integrals

$$Q_n := (x - x_n)\sum_{i=0}^{m} a_i f'(t_i) \ (a_i \ \textit{Gewichte}, t_i \ \textit{Stützpunkte}),$$

so gewinnt man x_{n+1} für die Verbesserung der Näherung x_n folgendermaßen:

$$x_{n+1} = x_n - \frac{f(x_n)}{\sum\limits_{i=0}^{m} a_i f'(t_i)},$$

wobei wir unterstellen, dass schon $f(x) \approx 0$ ist. Verwendet man die Rechteckregel $Q_n := (x - x_n)f'(x_n)$, so gewinnt man das Newtonverfahren zurück. Verwendet man die Mittelpunktregel $Q_n := (x - x_n)f'(\frac{1}{2}(x + x_n))$, so gewinnt man in naheliegender Weise die Iterationsformel

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n - \dfrac{f(x_n)}{2f'(x_n)})} \tag{3.7}$$

Dieses Verfahren („deformiertes Newton-Verfahren") konvergiert unter angepassten Voraussetzungen kubisch; siehe [20]. □

Bemerkung 3.8 *Das Newtonverfahren kann man auch kontinuierlich formulieren. Wir schreiben es im $\mathbb{R}^n$ auf. Sei $f : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ differenzierbar. Dann ist die kontinuierliche Version beschrieben durch die Anfangswertaufgabe*

$$x' = -Df(x)^{-1}f(x)\,,\ x(0) = x_0\,;\ x^0 \ \textit{Startwert}. \tag{3.8}$$

Die einfache Differenzen-Approximation führt zzur diskreten Iteration

$$x(t + h) = x(t) - hDf(x)^{-1}(x(t)f(x(t))\,,\ x(0) = x_0 \tag{3.9}$$

mit der Schrittweite $h > 0$. Es ist damit eine Art „gedämpftes Newtonverfahren" entstanden, von dem man weiß, dass damit eine globale Konvergenz erreicht werden kann. Dies entspricht der Erkenntnis, dass das kontinuierliche Verfahren (3.8) *global konvergiert; siehe [37].* □

3.2 Das Newtonverfahren für ein quadratisches Polynom

Wir wenden nun die Idee des Newton–Verfahrens auf ein spezielles quadratisches Polynom an. Die Betrachtungen dienen als Vorbereitung auf den Beweis des Satzes von Newton-Kantorovich im nächsten Abschnitt, wo wir wie hier [19] folgen. Der Unterschied zum Vorgehen im letzten Abschnitt ist, dass wir die hier die Existenz einer Nullstelle mit dem Newton-Verfahren mitbeweisen. Klar, bei einem quadratischen Polynom wissen wir über die Existenz von Nullstellen ohne Verfahren Bescheid, es geht hier eine Reihe von Beobachtungen bei der Newton-Iterationsfolge, die wir im nächsten Abschnitt nutzen wollen.

Sei $L, b > 0$. Wir betrachten damit das Polynom

$$g : \mathbb{R} \longrightarrow \mathbb{R}\,,\; t \longmapsto \frac{1}{2}Lt^2 - t + b\,.$$

unter der **Voraussetzung** $0 < 2bL \leq 1$.

Wir fassen die benötigten Aussagen in einer Liste von Beobachtungen zusammen:

(1) Nullstellen von g: $t_* = \dfrac{1-\sqrt{1-2bL}}{L}\,,\; t_{**} = \dfrac{1+\sqrt{1-2bL}}{L}\,.$

(2) $g(t) = \frac{1}{2}L(t-t_*)(t-t_{**})\,,\; g'(t) = \frac{1}{2}L((t-t_*)+(t-t_{**})), t \in \mathbb{R}\,.$

(3) $g(t) > 0\,,\; t \in [0, t_*)\,.$

(4) $g'(t) \leq L(t-t_*) < 0\,,\; t \in [0, t_*)\,.$

(5) g ist monoton fallend in $[0, t_*]$.

(6) Der Newton-Operator $n_g : [0, t_*) \ni t \longmapsto t - g'(t)^{-1}g(t) \in \mathbb{R}$ zu g ist wohldefiniert.

(7) $t_* - n_g(t) = -\frac{1}{2}Lg'(t)^{-1}(t_*-t)^2\,,\; t \in [0, t_*)\,.$

(8) $t_{**} - n_g(t) = -\frac{1}{2}Lg'(t)^{-1}(t_{**}-t)^2\,,\; t \in [0, t_*)\,.$

(9) $t < n_g(t) < t_*$ und $g(n_g(t)) = \frac{1}{2}Lg(t)^2g'(t)^{-2}\,,\; t \in [0, t_*)\,.$

(10) $n_g([0, t_*)) \subset [0, t_*)\,.$

(11) Die Newton-Iteration ist wohldefiniert: $t_0 := 0\,,\; t_{k+1} := n_g(t_k)\,,\; k \in \mathbb{N}_0\,.$

(12) $t_k \in [0, t_*)\,,\; k \in \mathbb{N}_0\,.$

(13) Die Folge $(t_k)_{k\in\mathbb{N}_0}$ ist strikt monoton wachsend.

(14) $t_* - t_{k+1} = -\frac{1}{2}Lg'(t_k)^{-1}(t_*-t_k)^2 \leq \frac{1}{2}(t_*-t_k)\,,\; k \in \mathbb{N}_0\,.$

(15) $\dfrac{t_{k+1}-t_*}{t_{k+1}-t_{**}} = \left(\dfrac{t_k-t_*}{t_k-t_{**}}\right)^2\,,\; k \in \mathbb{N}_0\,.$

(16) Ist $2bL < 1$, dann gilt $\theta := t_*/t_{**} < 1$ und wir haben

$$t_* - t_{k+1} = \frac{1-\theta^2}{1+\theta^2}\frac{L}{2\sqrt{1-2bL}}(t_*-t_k)^2 \leq \frac{L}{2\sqrt{1-2bL}}(t_*-t_k)^2\,,\; k \in \mathbb{N}_0\,.$$

(17) Ist $2bL < 1$, dann gilt

$$t_k = t_* - \frac{\theta^{2^k}}{1-\theta^{2^k}}\frac{2\sqrt{1-2bL}}{L}\,,\; k \in \mathbb{N}_0\,.$$

Wir geben Hinweise zum Beweis von Aussagen, die nicht ganz offensichtlich sind.
Zu (2) Satz von Vietá.
Zu (6) Klar.
Zu (7)

$$\begin{aligned} t_* - n_g(t) &= g'(t)^{-1}(g'(t)(t_*-t)+g(t)) \\ &= -g'(t)^{-1}\int_t^{t_*}(g'(s)-g'(t))ds \\ &= -g'(t)^{-1}\int_t^{t_*}L(s-t)ds \;=\; -\frac{1}{2}Lg'(t)^{-1}(t_*-t) \end{aligned}$$

Zu (8) Analog zu (7).
Zu (9) Folgt aus (7) und einer einfachen Rechnung (Taylorentwicklung).
Zu (10) Folgt aus (9).
Zu (11) Klar. Zu (12), (13) Beweist man induktiv.
Zu (14) Die erste Gleichheit ist nach (9) klar, die Ungleichung folgt aus $g'(t_k) \le L(t_* - t_k) < 0$, da $t_k \in [0, t_*)$.
Zu (15) Beweist man induktiv.
Zu (16), (17) Beweist man induktiv.

3.3 Der Satz von Newton-Kantorovich

Erneut gehen wir der Fragestellung „Nullstellen“ nach, und zwar im unendlichdimensionalen Kontext. Dazu betrachten wir die Lösung einer nichtlinearen Gleichung:

$$G(x) = \theta\,. \tag{3.10}$$

Dabei sei $G : U \longrightarrow Y$ mit X, Y Banachräume, $U \subset X$ offen. Es geht hier darum, Existenz und Eindeutigkeit einer Lösung nachzuweisen und sie konstruktiv zu berechnen. Dazu wollen wir das (iterative) **Newton-Verfahren** verwenden. Es unterstellt die (lokale) Fréchet-Differenzierbarkeit der Abbildung G. Dann besteht die Iteration des Newton-Verfahrens in

$$DG(x^k)\Delta x^k = -G(x^k)\,,\; x^{k+1} := x^k + \Delta x^k\,,\; k = 0, 1, 2, \ldots \tag{3.11}$$

Dabei ist der Startwert x^0 geeignet zu wählen. Um das Newton-Verfahren anwenden zu können, muss notwendigerweise $G(x^k)^{-1}$ für jedes x^k existieren und beschränkt sein. Diese Iteration leitet sich aus der Taylorentwicklung ab:

$$G(x^k + \Delta) = G(x^k) + DG(x^k)\Delta + o(\Delta)\,. \tag{3.12}$$

Hier ist das berühmte Theorem von Kantorovich ([23]). Der ursprüngliche Beweis ist ziemlich verwickelt, da die Absicherung, dass die Newtonfolge aus (3.11) eine vorgegebene Kugel nicht verlässt, mehrere ineinandergeschachtelte Argumente verlangt, mehr Klarheit wurde durch die so genannte Majorantenmethode erreicht; siehe [24]. Es sind eine Reihe von erfolgreichen Versuchen unternommen worden, den Beweis zu vereinfachen, und die quantitiven Abschätzungen für die Konvergenzgeschwindigkeit zu verbessern; siehe [14, 31, 39]. Wir folgen ganz eng der Beweisausarbeitung in [19].

Satz 3.9 (Kantorovich, 1939, 1948) *Seien X, Y Banachräume, sei $U \subset X$ offen, und sei G stetig Fréchet-differenzierbar in U. Sei $x^0 \in U$ und seien damit die folgenden quantitativen Annahmen erfüllt:*

(1) $DG(x^0)$ *ist stetig invertierbar*

(2) $\|DG(x^0)^{-1}G(x^0)\| \le b$

(3) $\|DG(x^0)^{-1}(DG(u) - DG(v))\| \le L\|u - v\|\,,\; u, v \in U$

(4) $2bL \le 1$

Setze

$$t_* := (1 - \sqrt{1-2bL})/L\,,\ t_{**} := (1 + \sqrt{1-2bL})/L\,,\ \theta := t_*/t_{**} < 1\,.$$

Ist dann

$$\overline{B}_{t_*}(x^0) \subset U$$

erfüllt, dann gilt:

(a) *Die Folge* $(x^k)_{k\in\mathbb{N}}$ *der Iteration* $x^{k+1} := x^k - DG(x^k)^{-1}G(x^k), k \in \mathbb{N}_0$, *existiert.*

(b) $x^k \in B_{t_*}(x^0), k \in \mathbb{N}\,.$

(c) $(x^k)_{k\in\mathbb{N}}$ *konvergiert gegen ein* $x^* \in \overline{B}_{t_*}(x^0)$ *mit* $G(x^*) = \theta\,.$

(d) x^* *ist die einzige Nullstelle von* G *in* $\overline{B}_{t_*}(x^0)\,.$

(e)
$$\|x^{k+1} - x^*\| \le \frac{1}{2}\|x^k - x^*\|\,,\ k \in \mathbb{N}_0\,. \tag{3.13}$$

(f) *Falls in der Voraussetzung (3) die strikte Ungleichung gilt, also* $2bL < 1$, *dann gilt*

$$\|x^{k+1} - x^*\| \le \frac{1-\theta^{2^k}}{1+\theta^{2^k}}\frac{L}{2\sqrt{1-2bL}}\|x^k - x^*\|^2 \le \frac{L}{2\sqrt{1-2bL}}\|x^k - x^*\|^2\,,\ k \in \mathbb{N}_0\,, \tag{3.14}$$

und x^* *ist die eindeutige Nullstelle von* G *in* $\overline{B}_\rho(x^0)$ *für alle* ρ *in* $[t_*, t_{**})$ *mit* $\overline{B}_\rho(x^0) \subset U\,.$

Beweis:
Die quantitativen Voraussetzungen und die Newton-Iteration aus (a) sind invariant unter einer linearen Transformation, die ein topologischer Isomorphismus zwischen den Räumen X, Y ist. Daher ist es ausreichend, den Satz unter der vereinfachenden Annahme

$$X = Y\,,\ DG(x^0) = I \tag{3.15}$$

zu beweisen, denn die Transformation $x \longmapsto DG(x^0)^{-1}G(x)$ führt auf diese vereinfachte Situation. Sei nun (3.15) zutreffend. Die Argumentation nutzt nun die skalare Funktion g aus dem Abschnitt 3.2 und die dort bereitgestellten Ergebnisse; die Bezeichnungen übernehmen wir.

Resultat 1
Ist $u \in \overline{B}_r(x^0)$ und $v \in B_r(x^0) \subset U$, dann gilt:

$$\|G(u) - (G(v) + DG(v)(u-v))\| \le \frac{1}{2}L\|u-v\|^2\,. \tag{3.16}$$

Beweis dazu:
Der Satz von Hahn-Banach liefert $\lambda \in X^*$ mit

$$\|\lambda\|_* = 1\,,\ \langle\lambda, G(u) - (G(v) + DG(v)(u-v))\rangle = \|G(u) - (G(v) + DG(v)(u-v))\|$$

Setze mit $a \in [0,1]$ $x_a := v + a(u-v)$ und $h(a) := \langle\lambda, G(x_a) - (G(v) + DG(v)(x_a - v))\rangle$. Eine einfache Rechnung ergibt $h'(a) = \langle\lambda, (DG(x_a) - DG(v))(u-v)\rangle$ und h ist stetig differenzierbar. Mit der Voraussetzung (3) folgt $h'(a) \le La\|u-v\|^2$ und Integration ergibt das Resultat.

Resultat 2
Ist $t \in [0, t_*)$ und $x \in \overline{B}_t(x^0)$, dann ist $DG(x)$ stetig invertierbar und es gilt

$$\|DG(x)^{-1}\| \le |g'(t)|^{-1} \le 1 + Lt_*\,. \tag{3.17}$$

Beweis dazu:
Wir wissen $t_* \le 1/L$ und damit $t \in [0, 1/L)$. Damit folgt (unter den vereinfachenden Annahmen aus Voraussetzung (2)

$$\|DG(x) - I\| \le L\|x - x^0\| \le Lt < 1\,.$$

Mit der Neumannschen Reihe (siehe [5], Satz 7.31) folgt

$$\|DG(x)^{-1}\| \le (1 - Lt)^{-1} = |g'(t)|^{-1}\,.$$

Offenbar gilt $|g'(t)|^{-1} \le 1 + Lt_*$ für $t \in [0, t_*)$.

Das *Resultat 2* rechtfertigt die Definition des so genannten *Newton-Operators*:

$$N_G : B_{t_*}(x^0) \ni x \longmapsto x - DG(x)^{-1}G(x) \in X\,.$$

Solange $x \in B_{t_*}(x^0)$ gilt, ist der Newton-Operator definiert, aber es ist zumindest nicht offensichtlich, dass das Bild $N_G(x)$ in $B_{t_*}(x^0)$ liegt. Dies gilt es nun abzusichern. Setze

$$K(t) := \{x \in \overline{B}_t(x) | \, \|G(x)\| \le g(t)\}, t \in [0, t_*)\,,\; K := \cup_{t\in[0,t_*)} K(t)\,.$$

Resultat 3
Für $t \in [0, t_*)$ und $x \in K(t)$ gilt:

(α) $\|DG(x)^{-1}G(x)\| \le -g'(t)^{-1}g(t)$

(β) $\|N_G(x) - x^0\| \le n_g(t)$

(γ) $\|G(N_G(x))\| \le g(n_g(t))$

(δ) $N_G(K(t)) \subset K(n_g(t))$

(η) $N_G(K) \subset K$

Beweis dazu:
Sei $x \in K(t), t \in [0, t_*)$. Wir wissen $\|DG(x)^{-1}\| \le |g(t)|^{-1}$ (*Resultat 2*). Daher folgt

$$\|DG(x)^{-1}G(x)\| \le \|DG(x)^{-1}\|\|G(x)\| \le |g'(t)|^{-1}g(t)$$

und (α) ist gezeigt. (β) folgt aus

$$\|N_G(x) - x^0\| \le \|x - x^0\| + \|DG(x)^{-1}D(x)\| \le t - g(t)g'(t)^{-1} = n_g(t)\,.$$

Daraus folgt nun $N_G(x) \in B_{t_*}(x^0)$. Dies ergibt mit *Resultat 1*

$$\|G(N_D(x)) - (G(x) + DG(x)(N_G(x) - x))\| \le \frac{1}{2}L\|DG(x)^{-1}G(x)\|^2\,.$$

Da $G(x) + DG(x)(N_G(x) - x) = \theta$ gilt, folgt mit der Identität $g(n_g(t)) = \frac{1}{2}Lg(t)^2g'(t)^{-2}$ (siehe Beobachtung (8) im vorhergehenden Abschnitt) die Aussage (γ) . Da $t < n_g(t) < t_*$ gilt (siehe Beobachtung (8) im vorhergehenden Abschnitt), erhalten wir mit (β), (γ) $N_G(x) \in K(n_g(t))$. Damit ist auch (δ) gezeigt. Sei $x \in K$, also $x \in K(t)$ für ein $t \in [0, t_*)$. Dann ist $N_G(x) \in K(n_g(t)) \subset K$. Also gilt (η) .
Resultat 4
Die Folge $(x^k)_{k\in\mathbb{N}}$ der Newton-Iteration ist wohldefiniert und es gilt:

(α) $x^k \in K(t_k) \subset B_{t_*}(x^0), k \in \mathbb{N}$.

(β) $(x^k)_{k\in\mathbb{N}}$ konvergiert gegen ein $x^* \in \overline{B}_{t_*}(x^0)$.

(γ) $G(x^*) = \theta$.

(δ) $\|x^* - x^k\| \le t_* - t_k\,,\; k \in \mathbb{N}$.

Beweis dazu:
Die Definiertheit der Folge ist schon klar, ebenso $x^k \in K, k \in \mathbb{N}$. Da $K \subset B_{t_*}(x^0)$ gilt, ist $x^k \in B_{t_*}(x^0), k \in \mathbb{N}$. Wegen $t_0 = 0$ ist $x^0 \in K(t_0)$ nach Voraussetzung (2). Induktiv folgt $x^k \in K(t_k), k \in \mathbb{N}$. Damit ist ($\alpha$) bewiesen. Nun erhalten wir mit *Resultat 3* (α)

$$\|x^{k+1} - x^k\| \leq t_{k+1} - t_k \, , \, k \in \mathbb{N} \, .$$

Da die Folge $(t_k)_{k\in\mathbb{N}}$ konvergiert (siehe Beobachtung (12), (13) im vorhergehenden Abschnitt), konvergiert auch $\sum_{k\in\mathbb{N}_0}(t_{k+1} - t_k)$ und wir folgern daraus, dass $(x^k)_{k\in\mathbb{N}}$ eine Cauchyfolge ist. Also konvergiert $(x^k)_{k\in\mathbb{N}}$ gegen ein $x^* \in \overline{B}_{t_*}(x^0)$. Ferner

$$\|x^* - x^k\| \leq \sum_{j=k}^{\infty}(t_{j+1} - t_j) = t_* - t_k \, , \, k \in \mathbb{N} \, .$$

Wegen $G(x^k) = DG(x^{k-1})(x^{k+1} - x^k)$ und $\|DG(x)\| \leq 1 + Lt_*, x \in B_{t_*}(x^0)$, erhalten wir $\lim_k G(x^k) = \theta$. Mit der Stetigkeit von G folgt $G(x^*) = \theta$. Damit sind alle Punkte des Resultats gezeigt.

Nun haben wir noch die Eindeutigkeits- und Konvergenzaussagen zu beweisen.

Resultat 5
Seien $u, v \in X, t, s \geq 0, r > 0$. Ist

$$\|u - x^0\| \leq t < t_*, \|v - x^0\| \leq r, G(v) = \theta, g(s) \leq 0, \overline{B}_r(x^0) \subset U,$$

dann gilt

$$s > t \text{ und } \|v - N_G(u)\| \leq (s - n_g(t))\frac{\|v - u\|}{(s - t)^2} \tag{3.18}$$

Beweis dazu:
Wir haben $v - N_G(u) = DG(u)^{-1}(G(u) + DG(u)(v - u)$ und da $G(v) = \theta$ ist, erhalten wir

$$\|DG(u)^{-1}(G(u) + DG(u)(v - u)\| \leq \frac{1}{2}L|g'(t)|^{-1}\|v - u\|^2 \, .$$

Daraus folgt

$$\|v - N_G(u)\| \leq \frac{L}{2|g'(t)|}(s - t)^2\frac{\|v - u\|}{(s - t)^2} \, .$$

Da $g'(t) < 0$ und $g(s) \leq 0$ ist, folgt

$$\begin{aligned} s - n_g(t) &= -g'(t)^{-1}(-g(t) - g'(t)(s - t)) \\ &\geq |g'(t)|^{-1}((g(s) - g(t) - g'(t)(s - t)) \;=\; \frac{L}{2|g'(t)|}(s - t)^2 \, . \end{aligned}$$

Nun folgen die behaupteten Aussagen unmittelbar.

Resultat 6
Ist $v \in \overline{B}_{t_*}(x^0)$ mit $G(v) = \theta$, dann gilt:

$$\|v - x^{k+1}\| \leq \frac{t_* - t_{k+1}}{(t_* - t_k)^2}\|v - x^k\|^2 \, , \, \|v - x^k\| \leq t_* - t_k \, , \, k \in \mathbb{N}_0 \, . \tag{3.19}$$

Insbesondere ist x^* die eindeutig bestimmte Nullstelle von G in $\overline{B}_{t_*}(x^0)$.

Beweis dazu:
Sei $k \in \mathbb{N}_0$. Aus *Resultat 4* folgt $\|x^k - x^0\| \leq t_k$ und mit *Resultat 5* mit $u = x^k, t = t_k, s = t_*$ erhalten wir

$$\|v - N_G(x^k)\| \leq (t_* - n_g(t_k))\frac{\|v - x^k\|^2}{(t_* - t_k)^2}$$

Damit ist die erste Ungleichung schon klar. Wir beweisen die zweite Ungleichung induktiv. Sie gilt für $k = 0$, da $v \in \overline{B}_{t_*}(x^0)$ und $t_0 = 0$. Mit der Induktionsvoraussetzung $\|v - x^k\| \le t_* - t_k$ und der ersten Ungleichung folgt $\|v - x^{k+1}\| \le t_* - t_{k+1}$.
Wir wissen schon $x^* \in \overline{B}_{t_*}(x^0)$ und $G(x^*) = \theta$. Aus der Konvergenz der Folgen $(x^k)_{k\in\mathbb{N}}$ und $(t_k)_{k\in\mathbb{N}}$ gegen x^* bzw. t_*, folgt mit der zweiten Ungleichung $v = x^*$.
Resultat 7
Wir haben

$$\|x^* - x^{k+1}\| \le \frac{t_* - t_{k+1}}{(t_* - t_k)^2}\|x^* - x^k\|^2 \, , \, k \in \mathbb{N}_0 \, , \tag{3.20}$$

$$\|x^* - x^{k+1}\| \le \frac{1}{2}\|x^* - x^k\| \, , \, k \in \mathbb{N}_0 \, . \tag{3.21}$$

Zusatz: Ist $2bL < 1$, dann gilt

$$\|x^* - x^{k+1}\| \le \frac{1 - \theta^{2^k}}{1 + \theta^{2^k}} \frac{L}{2\sqrt{1 - 2bL}}\|x^* - x^k\|^2 \le \frac{L}{2\sqrt{1 - 2bL}}\|x^* - x^k\|^2 \, , \, k \in \mathbb{N}_0 \, . \tag{3.22}$$

Beweis dazu:
Die erste Ungleichung folgt mit $v := x^*$ aus *Resultat 6*. Da $(t_* - t_{k+1})(t_* - t_k)^{-1} \le \frac{1}{2}$, und $\|x^* - x^k\|(t_* - t_k)^{-1} \le 1$, folgt die zweite Ungleichung. Die dritte Aussage folgt aus den vorhergehenden Resultaten und Aussagen zum Hilfspolynom.
Resultat 8
Ist $2bL < 1, t_* \le \rho < t_{**}$, und $\overline{B}_\rho(x^0) \subset U$, dann ist x^* die einzige Nullstelle von G in $\overline{B}_\rho(x^0)$.
Beweis dazu:
Sei $v^* \in \overline{B}_\rho(x^0)$ und $G(v^*) = \theta$.
Mit *Resultat 1* folgt mit $v = x^0, u = v^*$ offenbar $\|G(x^0) + v^* - x^0\| \le \frac{1}{2}L\|v^* - x^0\|^2$ (beachte $DG(x^0) = I$). Mit der Dreiecksungleichung und der Voraussetzung folgt

$$\|G(x^0) + v^* - x^0\| \ge \|v^* - x^0\| - \|G(x^0)\| \ge \|v^* - x^0\| - b$$

und daher

$$\frac{1}{2}L\|v^* - x^0\|^2 \ge \|v^* - x^0\| - b, \text{ d. h. } g(\|v^* - x^0\|) \ge 0 \, .$$

Dies ergibt $\|v^* - x^0\| \le t_*$. Damit folgt mit *Resultat 6*, dass $G(v^*) = \theta$ gilt und daher $v^* = x^*$ ist.

Nun sind alle Aussagen des Satzes von Kantorovich bewiesen. ■

Die Newtoniteration ist invariant gegenüber einer affinen Transformation der Abbildung G. *Glücklicherweise* sind die Voraussetzungen des Satzes auch invariant gegenüber einer affinen Transformation der Abbildung G. Daher heißt diese Fassung des Satzes auch der affin-invariante Satz von Kantorovich; siehe [14]. Wir fügen noch eine nicht affin-invariante Fassung des Satzes an, die zudem auch nicht die Existenz einer Lösung mitbeinhaltet.

Satz 3.10 (Newton-Verfahren) *Seien X, Y Banachräume, sei $U \subset X$ offen und sei $G : U \ni x \longmapsto G(x) \in Y$ Fréchet-differenzierbar. Sei $z \in U$ eine Lösung der Gleichung $G(x) = \theta$ and sei $DF(z)$ stetig invertierbar. Es gelte mit Konstanten $r, \beta, L > 0$:*

$$B_r(z) \subset U \, , \, \|DF(z)^{-1}\| \le \beta \, , \, \|DF(x) - DF(y)\| \le L|x - y| \text{ for all } x, y \in U \, .$$

Dann ist für alle $x^0 \in B_\delta(z)$ mit $\delta := \min\{r, \frac{1}{2\beta L}\}$ die Iteration $x^{k+1} := x^k - DG(x^k)^{-1}DG(x^k)$ definiert und liefert eine Folge $(x^n)_{n\in\mathbb{N}}$ mit

$$|x^{n+1} - z| \le \beta L|x^n - z|^2 \le \frac{1}{2}|x^n - z| \, , \, n = 0, 1, \ldots \, . \tag{3.23}$$

Beweis:
Wir wollen zeigen:

$$\text{Falls } x \in B_\delta(z) \text{ dann ist } DF(x) \text{ stetig invertierbar und } \|DF(x)^{-1}\| \le 2\beta\,.$$

Sei $x \in B_\delta(z)$. Dann setzen wir

$$\eta := \|DF(z)^{-1}(DF(x) - DF(z))\| \le \|DF(z)^{-1}\|\|DF(x) - DF(z)\| \le \beta L|x - z| \le \beta L\delta \le \frac{1}{2}\,.$$

Nun ist $DF(x)$ invertierbar und

$$DF(x)^{-1}\| \le (1-\eta)^{-1}\|DF(z)^{-1}\| \le 2\beta\,.$$

Wir zeigen nun induktiv

$$x^n \in B_\delta(z)\,,\, n = 0, 1, \ldots,\,.$$

Die Induktionsvoraussetzung ist auf Grund der Voraussetzungen des Satzes schon klar. Sei $x^n \in B_\delta(z)$. Wir haben

$$x^{n+1} = x^n - DF(x^n)^{-1}F(x^n) = x^n - DF(x^n)^{-1}(F(x^n) - F(z))$$

und daher

$$x^{n+1} - z = DF(x^n)^{-1}(F(z) - F(x^n) - DF(x^n)(z - x^n))\,.$$

Dies hat

$$|x^{n+1} - z| \le 2\beta\frac{L}{2}|x^n - z|^2 \le \beta L\delta|x^n - z| \le \frac{1}{2}|x^n - z|$$

zur Folge und die Induktion ist abgeschlossen. ■

Bemerkung 3.11 *Beachte, dass die Abschätzung* (3.23) *quadratische Konvergenz der Folge* $(x^n)_{n\in\mathbb{N}}$ *zur Lösung* z *zur Konsequenz hat.* □

Bemerkung 3.12 *Das Newton-Verfahren kann genutzt werden, einen Kandidaten für ein Minimum einer glatten Zielfunktion* $f : U \longrightarrow \mathbb{R}$ *zu berechnen, indem man eine Lösung der notwendigen Bedingung* $F(x) := \nabla f(x) = \theta$ *berechnet. Als Herausforderung kommt hier hinzu, dass man sicherstellen möchte, dass die Zielfunktion entlang der Newton-Iteration monoton nicht wachsend ist. Dazu gibt es eine Reihe von Vorschlägen; siehe etwa [38].* □

Bemerkung 3.13 *Das deformierte Newton-Verfahren (siehe Bemerkung 3.7) wird übertragen in den unendlichdimensionalen Kontext in [25]. Es wird beschrieben durch folgende Iteration:*

$$y^n := x^n - DG(x^n)^{-1}G(x^n)\,,\, x^{n+1} := x^n - DG(\frac{1}{2}(x^n + y^n))^{-1}G(x^n)\,,\, n \in \mathbb{N}_0\,. \tag{3.24}$$

□

3.4 Ein Satz zum Newtonverfahren unter Nutzung einer äußeren Inversen

Wir wollen erneut die Gleichung

$$G(x) = \theta\,. \tag{3.25}$$

mittels eines Newtonverfahrens lösen. Dabei sei $G : U \longrightarrow Y$ mit X, Y Banachräume, $U \subset X$ offen. Wir wollen hier auf die Voraussetzung verzichten, dass DG entlang einer Iteration injektiv ist. Daher wollen wir statt $DG(\cdot)^{-1}$ mit einer stetigen Abbildung $\Gamma : U \longrightarrow \mathcal{B}(Y,X)$ arbeiten. Die Iteration, die wir dann hinschreiben können, ist folgende:

$$x^{k+1} = x^k - \Gamma(x^k)G(x^k)\,,\ k \in \mathbb{N}_0\,. \tag{3.26}$$

Dabei ist $x^0 \in U$ ein Startwert. Wenn wir sicherstellen können, dass die Folge $(x^k)_{k\in\mathbb{N}_0}$ wohldefiniert ist, d. h. $x^k \in U$ gilt, dann gilt:

$$\begin{array}{rcll} \lim_k \Gamma(x^k)G(x^k) & = & \theta & \text{falls } x^* = \lim_k x^k \text{ existiert} \\ \Gamma(x^*)G(x^*) & = & \theta & \text{falls } x^* = \lim_k x^k \text{ existiert} \\ G(x^*) & = & \theta & \text{falls } x^* = \lim_k x^k \text{ existiert und } \Gamma(x^*) \text{ injektiv ist} \end{array}$$

Ist nun aber $\Gamma(x^*)$ nicht injektiv, können wir nur auf $G(x^*) \in \ker(\Gamma(x^*))$ schließen. Da wir $\Gamma(x^*)$ a priori nicht kennen, liegt es nahe, die Abbildung $\Gamma : U \longrightarrow \mathcal{B}(Y,X)$ so zu wählen, dass sich der Kern von Γ entlang der Iteration nicht ändert, was sichergestellt wird durch

$$\ker(\Gamma(x)) = \ker(\Gamma(x^0))\,,\ x \in U\,. \tag{3.27}$$

Unter dieser Voraussetzung können wir dann hoffen, die Gleichung (3.25) in folgendem Sinne zu lösen:

$$G(x) \in \ker(\Gamma(x^0))\,. \tag{3.28}$$

Dies führt uns nun zur eigentlichen Aufgabenstellung:
Gegeben $\Gamma_* \in \mathcal{B}(Y,X)$. Löse

$$G(x) \in \ker(\Gamma_*)\,. \tag{3.29}$$

Wenn es möglich ist, sicherzustellen, dass $\ker(\Gamma(x^0)) = \ker(\Gamma_*)$, dann ist das Verfahren (3.26) eine Iteration zur Berechnung einer Lösung von (3.29).

Definition 3.14 *Seien X,Y Banachräume und seien $T : X \longrightarrow Y$, $S : Y \longrightarrow x$ linear und stetig. S heißt* **äußere Inverse** *von T genau dann, wenn $STS = S$. Wir schreiben dann $S = T^{\#}$.* □

Eigenschaften der äußeren Inversen sind im Anhang 3.8 aufgeführt. Beachte, dass äußere Inverse (im Sinne unserer Definition) nicht eindeutig bestimmt sind; der Nulloperator ist ja immer schon eine äußere Inverse. Es kommt also immer auf den Kontext an, welche äußere Inverse Verwendung finden soll.

Wir starten mit einer Approximation $J : U \longrightarrow \mathcal{B}(X,Y)$ von DG. Sei $x^0 \in U$ und sei $\Gamma(x^0)$ eine äußere Inverse von $J(x^0)$ mit $\ker(\Gamma_{x^0})$. Damit können wir eine Abbildung $\Gamma : U \longrightarrow \mathcal{B}(Y,X)$ konstruieren gemäß

$$\Gamma(x) := (I - \Gamma(x^0)(J(x^0) - J(x)))^{-1}\Gamma(x^0)\,,\ x \in U\,. \tag{3.30}$$

Nun ist jedes $\Gamma(x)$ ist eine äußere Inverse von $J(x)$ mit $\ker(\Gamma(x)) = \ker(\Gamma(x^0))$, falls gilt:

$$\|\Gamma(x^0)(J(x^0) - J(x))\| < 1\,;$$

siehe Lemma 3.28.

Satz 3.15 (Chen-Nashed, 1993) *Seien X,Y Banachräume, sei $U \subset X$ offen, und sei G stetig Fréchet-differenzierbar in U. Sei $x^0 \in U$ und seien mit den Konstanten $b,L,M,m,N,n \geq 0$ die folgenden quantitativen Annahmen erfüllt:*

(1) $\|\Gamma(x^0)G(x^0)\| \leq b$

(2) $\|\Gamma(x^0)(DG(u) - DG(v))\| \le L\|u - v\|\,,\ u, v \in U$

(3) $\|\Gamma(x^0)(DG(x) - J(x))\| \le M\|x - x^0\| + m\,,\ x \in U$

(4) $\|\Gamma(x^0)(J(x) - J(x^0))\| \le N\|x - x^0\| + n\,,\ x \in U$

Setze

$$p := m + n\,,\ K := \max(L, M + N)\,,\ \kappa := Kb\,.$$

Ist

$$K > 0\,,\ p < 1\,,\ \kappa \le \frac{1}{2}(1-p)^2\,,\ \overline{B}_{t_*}(x^0) \subset U\,,$$

$$t_* := (1 - p - \sqrt{(1-p)^2 - 2\kappa})/K\,,\ t_{**} := (1 - p + \sqrt{(1-p)^2 - 2\kappa})/K\,,$$

dann gilt:

(a) *Die Folge* $(x^k)_{k\in\mathbb{N}}$ *der Iteration* $x^{k+1} := x^k - \Gamma(x^k)G(x^k), k \in \mathbb{N}_0$, *existiert.*

(b) $x^k \in B_{t_*}(x^0), k \in \mathbb{N}\,.$

(c) $(x^k)_{k\in\mathbb{N}}$ *konvergiert gegen ein* $x^* \in \overline{B}_{t_*}(x^0)\,.$

(d) $G(x^*) \in ker(\Gamma(x^0))\,.$

(e) x^* *ist lokal eindeutig bestimmte Lösung von* $G(x) \in ker(\Gamma(x^0))$ *in folgendem Sinne:*

$$V \cap (\Gamma(x^0)G)^{-1} = \{x^*\} \text{ mit } V := (\overline{B}_{t_*}(x^0) \cup B_{t_{**}}(x^0) \cap U)) \cap (x^0 + \Gamma(x^0))$$

(f) *Ist* $J = DG$ *und* $\kappa = \frac{1}{2}$, *dann haben wir die Abschätzung*

$$\|x^{k+1} - x^*\| \le \frac{1}{2}\|x^k - x^*\|\,,\ k \in \mathbb{N}_0\,. \tag{3.31}$$

(g) *Ist* $J = DG$ *und* $\kappa < \frac{1}{2}$, *dann kann die Abschätzung 3.32 verschärft werden zu*

$$\|x^{k+1} - x^*\| \le \frac{1}{2\sqrt{1-2\kappa}}\|x^k - x^*\|^2\,,\ k \in \mathbb{N}_0\,, \tag{3.32}$$

Zur Vorbereitung des Beweises zwei Lemmata zu Hilfsfunktionen, die dann zur Majorisierung beim Beweis Verwendung finden.

Lemma 3.16 *Definiere mit den Konstanten* $K, p, b, \kappa, M, m, N, n, t_*, t_{**}$ *aus Satz 3.15 die Funktionen* $f, g : \mathbb{R} \longrightarrow \mathbb{R}$ *durch*

$$f(t) := \frac{1}{2}Kt^2 + (p-1)t + b\,,\ g(t) := Nt + (n-1)\,,\ t \in \mathbb{R}\,.$$

Ferner sei $t_+ := (1-n)/N$ *. Dann gilt:*

(1) $0 < t_* \le t_{**}\,;\ t_* \le t_+\,,\ t_* = t_{**} \iff \kappa = \frac{1}{2}(1-p)^2\,.$

(2) $f^{-1}(-\infty, 0) = (t_*, t_{**})\,,\ f^{-1}(0) = \{t_*, t_{**}\}\,,\ f^{-1}(0, \infty) = \mathbb{R}\backslash[t_*, t_{**}]\,.$

(3) $g^{-1}(-\infty, 0) = (\infty, t_+)\,,\ g^{-1}(0) = \{t_+\}\,,\ g^{-1}(0, \infty) = (t_+, \infty)\,.$

(4) $f(s) = f(t) + f'(t)(s-t) + \frac{1}{2}K(s-t)^2\,,\ s, t \in \mathbb{R}\,.$

(5) $0 \le Mt + m \le f'(t) - g(t)\,,\ t \in \mathbb{R}\,.$

Beweis:
Diese Aussagen sind einfach nachzurechnen. ■

Lemma 3.17 *Definiere mit den Konstanten $K, p, b, \kappa, M, m, N, n, t_*, t_{**}$ aus Satz 3.15 die Funktionen $f, g : \mathbb{R} \longrightarrow \mathbb{R}$ durch*

$$f(t) := \frac{1}{2}Kt^2 + (p-1)t + b\,,\; g(t) := Nt + (n-1)\,,\; t \in \mathbb{R}\,.$$

Betrachte damit die Funktion

$$\psi : [0, t_*) \ni t \longmapsto t - \frac{f(t)}{g(t)} \in \mathbb{R}\,.$$

Dann gilt:

(1) $0 < \psi(t) < t_*\,,\; t \in [0, t_*)\,.$

(2) $\psi([0, t_*)) \subset [0, t_*)\,.$

(3) Die Folge $(t_k)_{k\in\mathbb{N}}$, definiert durch

$$t_0 := 0\,,\; t_{k+1} := \psi(t_k)\,,\; k \in \mathbb{N}_0\,,$$

ist wohldefiniert, strikt monoton wachsend, konvergent, und es gilt $\lim_k t_k = t_*\,.$

Beweis:
Zu (1) Sei $t \in [0, t_*)$. Dann ist $f(t) > 0, g(t) < 0$, und daher $\psi(t) > 0$. Ferner

$$0 = f(t_*) = f(t) + f'(t)(t_* - t) + \frac{1}{2}K(t_* - t)^2 \geq f(t) + g(t)(t_* - t) + \frac{1}{2}(t_* - t)^2\,.$$

Daraus lesen wir dann ab:

$$t_* - \psi(t) \geq -\frac{K}{2g(t)}(t_* - t)^2 > 0$$

Zu (2) Konsequenz von (1).
Zu (3) Wegen (2) ist die Folge wohldefiniert und wir wissen $t_k \in [0, t_*)$ für alle $k \in \mathbb{N}_0$. Aus (1) folgt die Monotonieaussage. Also ist $(t_k)_{k\in\mathbb{N}}$ konvergent; sei $l := \lim_k t_k$. Aus

$$|f(l)| = \limsup_k |f(t_k)| \leq \limsup_j |g(t_j)| \limsup_k \frac{|f(t_k)|}{|g(t_k)|} \leq (1-n)\limsup_k |t_{k+1} - t_k| = 0$$

und $f^{-1}(0) \cap [0, t_*] = \{t_*\}$ schließen wir $l = t_*$. ■

Kommen wir nun zum **Beweis von Satz 3.15**. Wir teilen ihn in Teilresultate auf. Zunächst eine Definition:

$$W(t) := \{x \in X | \|x - x^0\| \leq t, \|\Gamma(x^0)G(x)\| \leq f(t)\}, t \in [0, t_*)\,,\; W := \cup_{t\in[0,t_*)} W(t)\,.$$

Resultat 1
Ist $t \in [0, t_*), x \in W(t)$, dann gilt $\|\Gamma(x)G(x)\| \leq -g(t)^{-1}f(t)\,.$
Beweis dazu:

$$\begin{aligned}\|\Gamma(x)G(x)\| &= \|\Gamma(x)J(x^0)\Gamma(x^0)G(x)\| (\text{ siehe Lemma 3.28}) \leq \|\Gamma(x)J(x^0)\|\|\Gamma(x^0)G(x)\| \\ &\leq \frac{f(t)}{1 - \|\Gamma(x^0)(J(x^0) - J(x))\|} \leq \frac{f(t)}{1 - N\|x - x^0\| - n} \leq -\frac{f(t)}{g(t)}\,.\end{aligned}$$

Resultat 2
Ist $t \in [0, t_*), x \in W(t)$, dann gilt $\|x^0 - \Psi(x)\| \le \psi(t)$.
Beweis dazu:

$$\|x^0 - \Psi(x)\| \le \|x^0 - x\| + \|\Gamma(x)G(x)\| \le t - \frac{f(t)}{g(t)} = \psi(t)\,.$$

Resultat 3
Ist $t \in [0, t_*), x \in W(t)$, dann gilt $\|\Gamma(x^0)G(\Psi(x))\| \le \psi(f(t))$.
Beweis dazu: Wir schreiben $\Gamma(x^0)G(\Psi(x)) = u + v + w$ mit

$$\begin{aligned}
u &:= \Gamma(x^0)G(\Psi(x)) - \Gamma(x^0)G(x) - \Gamma(x^0)DG(x)(\Psi(x) - x)\\
v &:= \Gamma(x^0)G(x) + \Gamma(x^0)J(x)(\Psi(x) - x)\\
w &:= \Gamma(x^0)DG(x)(\Psi(x) - x) - \Gamma(x^0)J(x)(\Psi(x) - x)
\end{aligned}$$

Dann folgt

$$\begin{aligned}
\|u\| &\le \frac{1}{2}L\|\Psi(x) - x\|^2 \le \frac{1}{2}K\|\Gamma(x)G(x)\|^2 \le \frac{1}{2}K\frac{f(t)^2}{g(t)^2}\,,\\
\|v\| &= \Gamma(x^0)(G(x) + J(x)(\psi(x) - x)) = \Gamma(x^0)(I - J(x)\Gamma(x))G(x) = \theta\,,\\
\|w\| &\le \|\Gamma(x^0)(J(x) - DG(x))\|\|\Psi(x) - x\| \le (M\|x - x^0\| + m)\|\Gamma(x)G(x)\|\\
&\le -(Mt + m)\frac{f(t)}{g(t)} \le (g(t) - f'(t))\frac{f(t)}{g(t)}\,.
\end{aligned}$$

Dies ergibt nun

$$\begin{aligned}
\|\Gamma(x^0)G(\Psi(x))\| &\le \frac{1}{2}K\frac{f(t)^2}{g(t)^2} + (g(t) - f'(t))\frac{f(t)}{g(t)}\\
&= f(t) + f'(t)(\psi(t) - t) + \frac{1}{2}K(\psi(t) - t)^2 = f(\psi(t))\,.
\end{aligned}$$

Resultat 4
Wir haben $\Psi(W(t)) \subset W(\psi(t)) \subset W, t \in [0, t_*)$, und $\Psi(W) \subset W$.
Beweis dazu: Folgt aus *Resultat (2),(3)* und einer einfachen Rechnung.
Resultat 5
Ist $u, v \in V$ mit $\Gamma(x^0)G(v) = \theta$, dann gilt

$$\|\Psi(u) - v\| \le \|\Gamma(u)J(x^0)\|(\frac{1}{2}L\|u - v\|^2 + (M\|u - v\| + m)\|u - v\|)\,.$$

Beweis dazu: Wir haben (siehe 3.28,3.26)

$$\mathrm{ran}(\Gamma(x^0)) = \mathrm{ran}(\Gamma(u)) = \mathrm{ran}(\Gamma(u)J(u)) = \ker(I - \Gamma(u)J(u))$$

und daher

$$v - u = (v - x^0) - (u - x^0) \in \ker(I - \Gamma(u)J(u))\,.$$

Wir setzen

$$\begin{aligned}
y &:= \Gamma(u)J(x^0)(\Gamma(x^0)(G(v)) - \Gamma(x^0)(G(u)) - \Gamma(x^0)DG(u)(v - u)\,,\\
z &:= \Gamma(u)J(x^0)(\Gamma(x^0)DG(u)(v - u) - \Gamma(x^0)J(u)(v - u))\,.
\end{aligned}$$

Dann folgt

$$\begin{aligned}
y + z &= -\Gamma(u)J(x^0)\Gamma(x^0)(G(u) + J(u)(v - u))\\
&= -\Gamma(u)(G(u)) - \Gamma(u)J(u)(v - u)\\
&= u - \Gamma(u)(G(u)) - v = \psi(u) - v\,.
\end{aligned}$$

Aus der Voraussetzung (2) folgt

$$\|y\| \leq \frac{1}{2} L \|\Gamma(u) J(x^0)\| \|u - v\|^2 ,$$

$$\|z\| \leq \|\Gamma(u) J(x^0)\| (M\|u - v\| + m) \|u - v\| .$$

Daraus leitet man die Aussage ab.
Resultat 5
$x^k \in W(t_k)$, $k \in \mathbb{N}_0$, und es gilt die Aussage (a) des Satzes.
Indukion: $k = 0$ Wir haben $t_0 = 0$ und daher $\|\Gamma(x^0) G(x^0)\| \leq b = f(t_0)$.
$k \to k + 1$ Wegen

$$\|x^{k+1} - x^0\| = \|\Psi(x^k) - x^0\| \leq \psi(t_k) = t_{k+1}$$

und

$$\|\Gamma(x^0) G(x^{k+1})\| = \|\Gamma(x^0) G(\Psi(x^k))\| \leq f(\psi(t_k)) = f(t_{k+1})$$

ist der Induktionsschluss abgeschlossen.
Aussage (a) ergibt sich daraus.
Resultat 6
Aussage (b) gilt.
Beweis dazu: Aus folgt

$$\|x^{k+1} - x^k\| = \|\Gamma(x^k) G(x^k)\| \leq -\frac{f(t_k)}{g(t_k)} = t_{k+1} - t_k \, , \; k \in \mathbb{N}_0 \, .$$

Daraus folgt, da $(t_k)_{k\in\mathbb{N}_0}$ konvergent ist gegen t_*, ist auch $(x^k)_{k\in\mathbb{N}_0}$ konvergent und $x^* := \lim_k x^k \in \overline{B}_{t_*}(x^0)$ existiert.
Resultat 7
Aussage (c) gilt.
Beweis dazu: Wir haben

$$\sup_{k\in\mathbb{N}_0} \|\Gamma(x^0)(J(x^k) - J(x^0))\| \leq N \sup_{k\in\mathbb{N}_0} \|x^k - x^0\| + n \leq N t_* + n$$

und erhalten daher

$$\begin{aligned}
\|\Gamma(x^0) G(x^*)\| &= \limsup_k \|\Gamma(x^0) G(x^k)\| \\
&= \limsup_k \|(I - \Gamma(x^0)(J(x^0) - J(x^k)) \Gamma(x^k) G(x^k)\| \\
&\leq (1 + \sup_{j\in\mathbb{N}_0} \|\Gamma(x^0)(J(x^j) - J(x^0)\|) \limsup_k \|\Gamma(x^k) G(x^k)\| \\
&\leq (1 + N t_* + n) \limsup_k \|x^{k+1} - x^k\| \; = \; 0
\end{aligned}$$

Resultat 8
Aussage (d) gilt.
Beweis dazu: Sei $v \in V$ mit $\Gamma(x^0) G(v) = \theta$. Setze $t := \|v - x^0\|$. Dann folgt

$$\|x^1 - v\| = \|\Psi(x^0) - v\| \leq \|\Gamma(x^0) J(x^0)\| (\frac{1}{2} L t^2 + m t) \leq \frac{1}{2} K t^2 + p t \, .$$

Wir haben auch $\|x^1 - x^0\| \leq a$ und daher

$$t \leq \|x^1 - v\| + \|x^1 - x^0\| \leq \frac{1}{2} K t^2 + b t + a \, .$$

Daraus folgt $f(t) \geq 0$ und es muss daher $t \leq t_*$ gelten. Dies zeigt $v \in B_{t_*}(x^0)$.
Wir zeigen nun $x^k \in V$ für alle $k \in \mathbb{N}_0$. Wir wissen schon $x^k \in W(t_k) \subset \overline{B}_{t_*}(x^0), k \in \mathbb{N}_0$, und $x^{j+1} - x^j = -\Gamma(x^j)G(x^j) \in \operatorname{ran}(\Gamma(x^j)) = \operatorname{ran}(\Gamma(x^0))$, $j \in \mathbb{N}_0$. Daher

$$x^k - x^0 = \sum_{j=0}^{k-1} x^{j+1} - x^j = -\sum_{j=0}^{k-1} \Gamma(x^j)G(x^j) \in \operatorname{ran}(\Gamma(x^0)) .$$

Nun beweisen wir induktiv $\|x^k - v\| \leq t_* - t_k$, $k \in \mathbb{N}_0$, was dann unmittelbar $v = \lim_k x^k = x^*$ zur Konsequenz hat.
$k = 0$ Dies ist schon klar, $t_0 =$ ist.
$k \to k+1$

$$\begin{aligned} \|\Gamma(x^k)J(x^k)\| &\leq (1 - \|\Gamma(x^0)(J(x^k) - J(x^0))\|)^{-1} \\ &\leq (1 - N\|x^k - x^0\| - n)^{-1} \leq -g(t_k)^{-1} \end{aligned}$$

Mit der Induktionsvoraussetzung $x^k \in V$ folgt

$$\begin{aligned} \|x^{k+1} - v\| &\leq \|\Gamma(x^k)J(x^0)\|(\frac{1}{2}L\|x^k - v\|^2 + (M\|x^k - x^0\| + m)\|x^k - v\|) \\ &\leq -g(t_k)^{-1}(\frac{1}{2}K(t_* - t_k)^2 + (Mt_k + m)(t_* - t_k)) \\ &\leq -g(t_k)^{-1}(\frac{1}{2}K(t_* - t_k)^2 + (f'(t_k) - g(t_k))(t_* - t_k)) \\ &\leq -g(t_k)^{-1}(f(t_*) - f(t_k)) + t_* - t_k \\ &= t_* + \frac{f(t_k)}{g(t_k)} - t_k = t_* - t_{k+1} \end{aligned}$$

Resultat 9
Aussage (e) gilt.
Beweis dazu: Im Spezialfall $J = DG$ können die Konstanten vereinfacht werden zu

$$M = m = n = p = 0 , L = K = N .$$

Wir haben nun

$$\|x^{k+1} - x^*\| \leq \|\Gamma(x^k)DG(x^0)\|(\frac{1}{2}L\|x^k - x^*\|^2) \leq -Lg(t_k)^{-1}\|x^k - x^*\|^2 = \frac{L}{2(1 - Lt_k)}\|x^k - x^*\|^2 .$$

Ist $\kappa = \frac{1}{2}$, dann folgt mit $v := x^*$ und $Lt_* \leq 1$

$$\frac{L}{2(1 - Lt_k)}\|x^k - x^*\|^2 \leq \frac{1}{2}\frac{L(t_* - t_k)}{1 - Lt_k}\|x^k - x^*\| \leq \frac{1}{2}\|x^k - x^*\| .$$

Ist $\kappa < \frac{1}{2}$, dann haben wir $0 \leq Lt_* < 1$. Also

$$\frac{L}{2(1 - Lt_k)}\|x^k - x^*\|^2 \leq \frac{L}{2(1 - Lt_*)}\|x^k - x^*\|^2 = \frac{L}{2\sqrt{1 - 2\kappa}}\|x^k - x^*\|^2 .$$

■

Bemerkung 3.18 *Die Voraussetzungen des Satzes 3.15 können als affin invariant angesehen werden, da man die Transformation, der man G unterwirft, auch J unterwerfen kann.* □

Eine Anwendung des Satzes 3.15 findet man in [2]. Dort wird die Existenz einer Lösung der Euler-Lagrange-Gleichung eines nichtlienaren Pendels nachgewiesen.

3.5 Der Satz von Kantorovich unter einer Surjektivitätsbedingung

Betrachte erneut die Gleichung

$$G(x) = \theta\,. \tag{3.33}$$

Dabei sei $G : U \longrightarrow Y$ mit X, Y Banachräume, $U \subset X$ offen. Wir betrachten nun folgende Iterationsmethode zur Lösung dieser Gleichung:

$$x^{k+1} = x^k - \Phi_k(x^k)\,,\ k \in \mathbb{N}_0\,. \tag{3.34}$$

Dabei sei $x^0 \in U$ ein Startwert und $(\Phi_k)_{k\in\mathbb{N}_0}$ eine Familie von Abbildungen, definiert auf U mit Werten in X .

Satz 3.19 (Polyak, 1964) *Seien X, Y Banachräume, sei $U \subset X$ offen, und sei G stetig Fréchet-differenzierbar in U . Sei $x^0 \in U$ und seien damit die folgenden quantitativen Annahmen erfüllt:*

(1) $\|G(x^0)\| \le a$

(2) $\|DG(u) - DG(v)\| \le L\|u - v\|\,,\ u, v \in U$

(3) $\|\Phi_k(x)\| \le b\|G(x)\|\,,\ x \in U\,, k \in \mathbb{N}_0\,.$

(4) $\|G(x) - DG(x)\Phi_k(x)\| \le c\|G(x)\|\,,\ x \in U, k \in \mathbb{N}_0\,.$

(5) $\kappa := \frac{1}{2}Lab^2 + c < 1\,.$ *3.15*

Setze $\tau_* := \dfrac{ab}{1-\kappa}$ *. Ist dann*

$$\overline{B}_{\tau_*}(x^0) \subset U$$

erfüllt, dann gilt:

(a) *Die Folge $(x^k)_{k\in\mathbb{N}}$ der Iteration $x^{k+1} := x^k - \Phi_k(x^k)G(x^k), k \in \mathbb{N}_0,$ existiert.*

(b) $x^k \in B_{\tau_*}(x^0), k \in \mathbb{N}\,.$

(c) *$(x^k)_{k\in\mathbb{N}}$ konvergiert gegen ein $x^* \in \overline{B}_{\tau_*}(x^0)$ mit $G(x^*) = \theta$.*

(d) *Wir haben die a priori-Abschätzung*

$$\|x^{k+1} - x^*\| \le \tau_*\kappa^k\,,\ k \in \mathbb{N}_0\,. \tag{3.35}$$

(e) *Falls in der Voraussetzung (4) $c = 0$ gilt, dann gilt $x^k \in B_{s_*}(x^0), k \in \mathbb{N}_0,$ mit $s_* := ab\sum_{j=0}^{\infty}\kappa^{2^j-1}$ und die a priori-Abschätzung 3.35 kann verschärft werden zu*

$$\|x^{k+1} - x^*\| \le ab\frac{\kappa^{2^k-1}}{1+\kappa^{2^k}} \le \tau_*\kappa^{2^k-1}\,,\ k \in \mathbb{N}_0\,. \tag{3.36}$$

Beweis:
Zu (a) Sei $x^k \in U$ mit $x^{k+1} \in U$. Dann gilt

$$\begin{aligned}
\|G(x^{k+1})\| &\le \|G(x^{k+1}) - (G(x^k) + DG(x^k)(x^{k+1} - x^k))\| + \|G(x^k) - DG(x^k)\Phi_k(x^k)\| \\
&\le \frac{1}{2}L\|x^{k+1} - x^k\|^2 + c\|G(x^k)\| \\
&\le (\frac{1}{2}Lb^2\|G(x^k)\| + c)\|G(x^k)\|
\end{aligned}$$

Definiere die Folge $(\tau_k)_{k\in\mathbb{N}_0} \subset [0,\infty)$ durch

$$\tau_k := ab\sum_{j=0}^{k-1}\kappa^j \,,\; k\in\mathbb{N}_0\,.$$

Offenbar ist diese Folge konvergent; $\tau_* := \lim_k \tau_k$. Wir zeigen induktiv

$$x^k \in \overline{B}_{\tau_k}(x^0) \subset \overline{B}_{\tau_*}(x^0)\,,\; \|G(x^k)\| \le a\kappa^k \le a\,,\; k\in\mathbb{N}_0\,. \tag{3.37}$$

$k=0$ Klar, da $\tau_0 = 0$ ist; siehe Vorausetzung (1).
$k \to k+1$

$$\begin{aligned}\|x^{k+1}-x^0\| &\le \|x^{k+1}-x^k\| + \|x^k - x^0\| \\ &\le \|\Phi_k(x^k)\| + \tau_k \;\le\; ab\kappa^k + \tau_k \;\le\; \tau_{k+1}\,.\end{aligned}$$

Dies zeigt $x^{k+1} \in \overline{B}_{\tau_{k+1}}(x^0)$. Ferner

$$\|G(x^{k+1})\| \;\le\; (\frac{1}{2}Lb^2\|G(x^k)\| + c)\|G(x^k)\| \;\le\; (\frac{1}{2}Lab^2 + c)a\kappa^k \;=\; a\kappa^{k+1}$$

Damit ist nun klar, dass die Folge $(x^k)_{k\in\mathbb{N}_0}$ wohldefiniert ist.
Zu (b) Wir haben nun

$$\|x^{k+1}-x^k\| \le b\|G(x^k)\| \le ab\kappa^k = \tau_{k+1} - \tau_k\,,\; k\in\mathbb{N}_0\,.$$

Da $(\tau_k)_{k\in\mathbb{N}_0}$ konvergent ist, ist auch $(x^k)_{k\in\mathbb{N}_0}$ konvergent. Also existiert $x^* := \lim_k x^k \in \overline{B}_{\tau_*}(x^0)$.
Zu (c) Nun folgt

$$\|G(x^*)\| = \limsup_k \|G(x^k)\| \le a \limsup_k \kappa^k = 0\,.$$

Zu (d) Wir erhalten

$$\|x^* - x^k\| \le \tau_* - \tau_k = ab\sum_{j=0}^{\infty}\kappa^j = ab\frac{\kappa^k}{1-\kappa} = \tau_*\kappa^k\,,\; k\in\mathbb{N}_0\,.$$

Zu (e) Definiere die Folge $(t_k)_{k\in\mathbb{N}_0}$ durch

$$t_k := ab\sum_{j=0}^{k-1}\kappa^{2^j-1}\,,\; k\in\mathbb{N}_0\,.$$

Klar, $t_* := \lim_k t_k$ existiert. Dann können wir zeigen

$$\|x^k - x^0\| \le t_k - t_* \le t_k \le t_*\,,\; k\in\mathbb{N}_0\,,$$

und

$$\|x^* - x^k\| \le t_* - t_k = ab\sum_{j=0}^{\infty}\kappa^{2^j-1}\; k\in\mathbb{N}_0\,.$$

Ferner gilt

$$\sum_{j=0}^{\infty}\kappa^{2^j-1} = \kappa^{-1}\sum_{j=0}^{\infty}(\kappa^{2^k})^{2^j} \le \kappa^{-1}\sum_{j=1}^{\infty}(\kappa^{2^k})^j = \frac{\kappa^{2^k}-1}{1-\kappa^{2^k}} \le \frac{\kappa^{2^k}-1}{1-\kappa}\,.$$

■

Nun können wir ein Resultat zum Newtonverfahren aufschreiben, das nur eine Surjektivitätsbedingung nutzt; siehe Anhang 3.8. Wir schreiben es als ein **relaxiertes** Verfahren auf. **Relaxation** meint, dass der Newtonschritt nicht ohne „Schrittweitensteuerung“ ausgeführt wird.

Satz 3.20 (Newton-Mysovskikh-Theorem, Polyak, 1964) *Seien X, Y Banachräume, sei $U \subset X$ offen, und sei G stetig Fréchet-differenzierbar in U. Sei $x^0 \in U$ und seien mit den Konstanten a, L, b damit die folgenden quantitativen Annahmen erfüllt:*

(1) $\|G(x^0)\| \leq a$

(2) $\|DG(u) - DG(v)\| \leq L\|u - v\|$, $u, v \in U$

(3) $sur(DG(x)) > 1/b$, $x \in U$.

Seien $\omega_-, \omega_+ \in (0, 2)$ und setze $c := \max(|\omega_- - 1|, |\omega_+ - 1|)$. Setze $\kappa := \frac{1}{2}\omega_+^2 Lab^2$. Ist dann $\kappa < 1$ und ist $(\omega_k)_{k\in\mathbb{N}}$ eine Folge mit $\omega_k \in [\omega_-, \omega_+], k \in \mathbb{N}_0$, und gilt $\overline{B}_{\tau_*}(x^0) \subset U$ mit $\tau_* := \omega_+ \frac{ab}{1-\kappa}$, dann gilt:

(a) *Es gibt eine stetige Abbildung $\Gamma : U \longrightarrow \mathcal{B}(Y, X)$ mit $\|\Gamma(x)\| \leq b, x \in U$.*

(b) *Die Folge $(x^k)_{k\in\mathbb{N}}$, definiert durch die Iteration*

$$x^{k+1} = x^k - \omega_k \Gamma(x^k) G(x^k)\,,\ k \in \mathbb{N}_0, \tag{3.38}$$

ist wohldefiniert.

(c) $x^k \in B_{\tau_*}(x^0), k \in \mathbb{N}_0$.

(d) *$(x^k)_{k\in\mathbb{N}}$ konvergiert gegen ein $x^* \in \overline{B}_{\tau_*}(x^0)$ mit $G(x^*) = \theta$.*

(e) *Wir haben die a priori-Abschätzung*

$$\|x^{k+1} - x^*\| \leq \tau_* \kappa^k\,,\ k \in \mathbb{N}_0\,. \tag{3.39}$$

(f) *Falls $\omega_k = 1, k \in \mathbb{N}_0$, gilt, dann gilt $x^k \in B_{s_*}(x^0), k \in \mathbb{N}_0$ mit $s_* := ab\sum_{j=0}^{\infty} \kappa^{2^j - 1}$ und die a priori-Abschätzung 3.39 kann verschärft werden zu*

$$\|x^{k+1} - x^*\| \leq ab\frac{\kappa^{2^k - 1}}{1 + \kappa^{2^k}} \leq \tau_* \kappa^{2^k - 1}\,,\ k \in \mathbb{N}_0\,. \tag{3.40}$$

Beweis:
Zu (a) Dies folgt aus 3.33. Also ist der Newtonschritt ausführbar, solange x^k in U ist.
Zu (b),(c),(d) Wir haben

$$\Phi_k(x) = \omega_k \Gamma(x) G(x)\,,\ x \in U, k \in \mathbb{N}_0\,.$$

Nun sieht man einfach ein, dass alle Voraussetzungen des Satzes 3.19 erfüllt sind. Wir haben für $x \in U$

$$\|G(x) - DG(\omega_k \Gamma(x) G(x))\| = |1 - \omega_k| \|G(x)\| \leq c\|G(x)\|$$

und

$$\|\omega_k \Gamma(x) G(x)\| \leq \omega_k \|\Gamma(x)\| \|G(x)\| \leq \omega_+ b\|G(x)\|\,.$$

Im Vergleich zu Satz 3.19 haben wir b durch $\omega_+ b$ zu ersetzen.
Zu (e) Wir beobachten, dass $c = 0$ genau dann gilt, wenn $\omega_k = 1, k \in \mathbb{N}_0$, ist. ■

Bemerkung 3.21 *In der Originalarbeit [33] ist die Voraussetzung $sur(DG(x)) > 1/b$ ersetzt durch $inj(DG(x)^*) > 1/b$; siehe dazu Lemma 3.34.*

Die Voraussetzungen in Satz 3.20 sind nicht affin invariant gestellt. Sie können auch nicht die Eindeutigkeit garantieren. □

3.6 Anwendung: Ein Konvexitätsprinzip

Als Anwendung von Satz 3.20 beweisen wir das so genannte **Konvexitätsprinzip**. Es handelt von der Idee, dass (auch) nichtlineare Abbildungen lokal die gleichen Eigenschaften haben wie die Linearisierung[4].

Lemma 3.22 *Sei Z ein Banachraum und sei $K \subset Z$ mit der Eigenschaft*

$$\frac{1}{2}(u+v) \in int(K) \text{ für alle } u,v \in K, u \neq v\,. \tag{3.41}$$

Dann ist K strikt konvex.[5]

Beweis:
Es genügt zu zeigen, dass K konvex ist. Seien $u,v \in K, u \neq v$, und sei damit $y_t := (1-t)u+tv, t \in [0,1]$. Sei $Q := \{t \in [0,1] | y_t \in K\}$. Setze

$$W_k := \{j2^{-k} | j \in \mathbb{Z}, 1 \leq j \leq 2^k - 1\}\,,\ W := \cup_{k=0}^{\infty} W_k\,.$$

Induktiv folgt mit (3.41) $W_k \subset Q, k \in \mathbb{N}_0$, und daher gilt $W \subset Q$. Setze

$$a := \sup\{s \in [0,\frac{1}{2}] | [\frac{1}{2}-s, \frac{1}{2}+s] \subset Q\}\,.$$

Wegen (3.41) haben wir $a > 0$. Ferner

$$(\frac{1}{2}-a, \frac{1}{2}+a) = \cup_{s \in (0,a)} [\frac{1}{2}-s, \frac{1}{2}+s] \subset Q\,.$$

Wenn wir $a = \frac{1}{2}$ zeigen können, ist K konvex.
Annahme: $a < \frac{1}{2}$.
Da W dicht in $[0,1]$ ist, können wir $d \in (a, \min(\frac{1}{2}, 3a)) \cap W$ wählen. Diese Wahl ergibt

$$0 < 1-a-d < 1-2a < 1+a-d < 1\,,\ 1 < 1-a+d < 1+2a < 1+a+d < 2\,.$$

Mit (3.41) erhalten wir

$$(\frac{1}{2}(1-a-d), \frac{1}{2}(1+a-d)) \subset Q\,,\ (\frac{1}{2}(1-a+d), \frac{1}{2}(1+a+d)) \subset Q\,.$$

Dann haben wir

$$\begin{aligned}(\frac{1}{2}-\frac{1}{2}(a+d), \frac{1}{2}+\frac{1}{2}(a+d)) &= \\ = (\frac{1}{2}(1-a-d), \frac{1}{2}(1+a-d)) &\cup (\frac{1}{2}(1-2a), \frac{1}{2}(1+2a)) \cup (\frac{1}{2}(1-a+d), \frac{1}{2}(1+a+d)) \subset Q\,.\end{aligned}$$

Insbesondere gilt $[\frac{1}{2}-s, \frac{1}{2}+s] \subset Q$ für alle $s \in (a, \frac{1}{2}(a+d))$, was ein Widerspruch zur Definition von a ist. ∎

Satz 3.23 *Seien X, Y Banachräume, sei $U \subset X$ offen, und sei G stetig Fréchet-differenzierbar in U. Sei $\rho > 0$ und $B_\rho(x^0) \subset U$. Seien damit mit den Konstanten K, L, b die folgenden quantitativen Annahmen erfüllt:*

[4] Beispiele sind: G invertierbar, $DG(x)$ invertierbar (unter Voraussetzungen!), G kompakt, $DG(x)$ kompakt, G offen, $DG(x)$ offen,...

[5] K heisst **strikt konvex**, wenn für alle $u,v \in K, u \neq v$, und $t \in (0,1)$ $tu + (1-t)v \in \text{int}(K)$ gilt.

(1) $\|G(x^0)\| \le \frac{\rho}{b}$

(2) $\|DG(u) - DG(v)\| \le L\|u - v\|$ *,* $u, v \in B_\rho(x^0)$

(3) *sur*$(DG(x)) > 1/b$ *,* $x \in B_\rho(x^0)$ *.*

Dann hat die Gleichung $G(x) = \theta$ *eine Lösung in* $\overline{B}_\rho(x^0)$ *.*

Beweis:
Wir wollen Satz 3.20 anwenden für $U := B_\rho(x^0)$ und $\omega_k := \omega \in (0,1), k \in \mathbb{N}_0$. Die Wahl von ω ist noch geeignet zu treffen. Wir haben dazu sicherzustellen, dass

$$1 > \kappa(\omega) := 1 - \omega + \frac{1}{2}\omega^2 Lab^2 \, , \, \rho > \tau_*(\omega) := \omega \frac{ab}{1-\kappa} = \frac{2ab}{2 - \omega Lab^2}$$

gilt. Dazu setze

$$\overline{\omega} := \min(1, \frac{2}{Lab^2}(1 - ab\rho^{-1})) > 0$$

und wähle $\omega \in (0, \overline{\omega})$. Dann sind alle Voraussetzungen von Satz 3.20 erfüllt und wir haben eine Lösung $x^*(\omega)$ der Gleichung $G(x) = \theta$ in $\overline{B}_{\tau_*(\omega)} \subset \overline{B}_\rho(x^0)$. ■

Satz 3.24 (Konvexitätsprinzip, Polyak, 2003) *Seien* X, Y *Hilberträume, sei* $U \subset X$ *offen, und sei* G *stetig Fréchet-differenzierbar in* U *. Sei* $z \in U$ *und* $R > 0$ *mit* $\overline{B}_R(z) \subset U$ *. Seien damit mit den Konstanten* L, b *die folgenden quantitativen Annahmen erfüllt:*

(1) $\|DG(u) - DG(v)\| \le L\|u - v\|$ *,* $u, v \in \overline{B}_R(z)$

(2) *sur*$(DG(z)) > 1/c$ *.*

Sei nun $r \in (0, \min(\frac{1}{2cL}, R)$ *. Dann gilt:*

(a) $G(\overline{B}_R(z))$ *ist strikt konvex.*

(b) $G(B_R(z))$ *ist offen.*

Beweis:
Setze $1/b := 1/c - Lr > 0$. Dann gilt $\text{sur}(DG(x)) > 1/b, x \in \overline{B}_r(z)$, denn mit Lemma

$$\text{sur}(DG(x)) \ge \text{sur}(DG(z)) - \|DG(x) - DG(z)\| > 1/c - L\|x - z\| \ge 1/c - Lr \, , \, x \in \overline{B}_r(z) \, .$$

Zu (a) Seien $x_0, x_1 \in \overline{B}_r(z)$. Setze

$$y_0 := G(x_0), y_1 := G(x_1) \, , \, x_{1/2} := \frac{1}{2}(x_0 + x_1), y_{1/2} := \frac{1}{2}(y_0 + y_1) \, .$$

Setze $\rho := \dfrac{\|x_0 - x_1\|}{8r}$. Dann haben wir $\overline{B}_\rho(x_{1/2}) \subset B_r(z)$[6] und

$$y_i = G(x_{1/2}) + DG(x_{1/2})(x_i - x_{1/2}) + R_i \, , \, \|R_i\| \le \frac{1}{2}L\|x_i - x_{1/2}\|^2 = \frac{1}{8}L\|x_0 - x_1\|^2 \, , \, i = 0, 1 \, .$$

Wir setzen $R_{1/2} := \frac{1}{2}(R_0 + R_1)$ und erhalten

$$\|R_{1/2}\| \le \frac{1}{2}(\|R_0\| + \|R_1\|) \le \frac{1}{8}\|x_0 - x_1\| \, .$$

[6] Hier wird benutzt, dass ein Hilbertraum gleichmäßig konvex ist. Dies folgt aus der Parallelogrammidentität sofort.

Ferner gilt mit $Lr = 2Lr - Lr < 1/c - Lr = 1/b$

$$\|G(x_{1/2}) - y_{1/2}\| \leq \frac{1}{8}\|x_0 - x_1\|^2 = \rho Lr < \rho b^{-1} .$$

Wir betrachten nun $\hat{G} : \overline{B}_\rho(x_{1/2}) \ni x \longmapsto G(x) - y_{1/2} \in Y$. Aus Satz 3.23 folgt die Existenz von $x^* \in \overline{B}_\rho(x_{1/2})$ mit $G(x^*) = y_{1/2}$ und daher $y_{1/2} \in G(\overline{B}_R(z))$. Nun ist $G(x^*) = y_{1/2}$ mit $x^* \in B_r(z)$. Also folgt $y_{1/2} \in \mathrm{int}(G(\overline{B}_r(z)))$. Damit folgt die Konvexität von $G(\overline{B}_r(z))$ aus Lemma 3.22.
Sei $\overline{x} \in B_r(z)$ und setze $\overline{y} := G(\overline{x})$. Dann gibt es $\overline{r} > 0$, so dass $\overline{B}_{\overline{r}}(\overline{z}) \subset B_r(z)$. Sei $y \in Y$ mit $\|G(\overline{x}) - y\| < \overline{r}b^{-1}$. Dann können wir Satz 3.23 auf $\hat{G} : \overline{B}_{\overline{r}}(\overline{x}) \ni x \longmapsto G(x) - y \in Y$ anwenden und erhalten $x \in \overline{B}_{\overline{r}}(\overline{x})$ mit $G(x) = y$. Dies zeigt

$$\overline{B}_{\overline{x}b^{-1}}(\overline{y}) \subset G(\overline{B}_{\overline{r}}(\overline{x})) \subset G(B_r(z))$$

und daher ist $\overline{y}$ ein innerer Punkt von $G(B_r(z))$. ■

3.7 Anhang: Äußere Inverse

Wir fassen hier Ergebnisse zur Existenz und zu Eigenschaften einer äußeren Inversen zusammen. Nicht zu allen Ergebnissen werden wir die Beweise anführen.

Definition 3.25 *Seien X, Y Banachräume und seien $T \in \mathcal{B}(X, Y)$, $S \in \mathcal{B}(Y, X)$. Wir nennen S eine* **äußere Inverse** *von T falls $STS = S$. Wir schreiben dann $T^\# := S$.* □

Äußere Inverse sind nicht eindeutig bestimmt. Beispielsweise ist der Nulloperator stets eine äußere Inverse. Ist ein Operator $T \in \mathcal{B}(X, Y)$ stetig invertierbar, dann ist diese Inverse auch eine äußere Inverse.

Lemma 3.26 *Seien X, Y Banachräume, sei $T \in \mathcal{B}(X, Y)$, und sei $S \in \mathcal{B}(Y, X)$ eine äußere Inverse. Dann gilt:*

(1) $(ST)(ST) = (ST)$, $(TS)(TS) = (TS)$.

(2) $\mathit{ran}(ST) = \mathit{ker}(I - ST)$, $\mathit{ran}(TS) = \mathit{ker}(I - TS)$.

(3) $X = \mathit{ker}(ST) \oplus \mathit{ran}(ST)$, $Y = \mathit{ker}(TS) \oplus \mathit{ran}(TS)$.

(4) $\mathit{ker}(TS) = \mathit{ker}(S)$, $\mathit{ran}(ST) = \mathit{ran}(S)$.

Beweis:
Zu (1),(2),(3) Sei $P \in \{TS, ST\}$. Klar, P ist linear stetig und offenbar gilt $PP = P$ (P ist ein Projektor). Damit folgen (2),(3) in wohlbekannter Weise.
Zu (4) Aus $STS = S$ folgt

$$S^{-1}(\theta) \subset S^{-1}(T^{-1}(\theta)) = (TS)^{-1}(\theta) \subset (TS)^{-1}(S^{-1}(\theta)) = (STS)^{-1}(\theta) = S^{-1}(\theta)$$

und

$$S(Y) = (STS)(Y) = (ST)(S(Y)) \subset (ST)(X) = S(T(X)) \subset S(Y)$$

■

Zur Erinnerung: Der Rang eines linearen Operators T ist die Dimension seines Bildes, also $\mathrm{rank}(T) = \dim \mathrm{ran}(T)$.

Lemma 3.27 *Seien X, Y Banachräume und sei $T \in \mathcal{B}(X,Y)$. Dann gilt:*

(1) Für alle $k \in \mathbb{Z}, 0 \le k \le rank(T)$, existiert eine äußere Inverse S von T mit $k = rank(S)$.

(2) Es gibt eine äußere Inverse S von T mit $rank(S) = \infty$ genau dann, wenn ein abgeschlossener linearer Teilraum X_0 von X existiert, so dass $X_0 \cap ker(T) = \{\theta\}$ gilt und der lineare Teilraum $Y_0 := T(X_0)$ von Y ein Komplement in Y besitzt.

Beweis:
Siehe [36], Theorem 2.3. ■

Lemma 3.28 *Seien X, Y Banachräume, sei $T \in \mathcal{B}(X,Y)$ und sei $T^{\#}$ eine äußere Inverse von T. Ferner sei $S \in \mathcal{B}(Y,X)$, und es gelte: $\|T^{\#}(T-S)\| < 1$. Dann ist $I - T^{\#}(T-S)$ invertierbar und durch*

$$S^{\#} := (I - T^{\#}(T-S))^{-1}T^{\#}$$

wird eine äußere Inverse von S definiert mit

$$ker(S^{\#}) = ker(T^{\#}),\ ran(S^{\#}) = ran(T^{\#})$$

und es gilt:

$$\|S^{\#}T\| \le \frac{1}{1 - \|T^{\#}(T-S)\|}$$

Beweis:
Sei $R := I - T^{\#}(T-S)$. Da $\|T^{\#}(T-S)\| < 1$ gilt, hat R eine beschränkte Inverse (Neumannsche Reihe!). Also ist $S^{\#}$ wohldefiniert. Da $T^{\#}$ eine äußere Inverse von T ist, folgt $T^{\#}TR = T^{\#}S$ und daher $T^{\#}SR^{-1} = T^{\#}T$. Nun erhalten wir

$$S^{\#}SS^{\#} = R^{-1}T^{\#}SR^{-1}T^{\#} = R^{-1}T^{\#}TT^{\#} = R^{-1}T^{\#} = S^{\#}.$$

Also ist $S^{\#}$ eine äußere Inverse von S.
$\ker(S^{\#}) = \ker(T^{\#})$ folgt unmittelbar aus $S^{\#} = R^{-1}T^{\#}$. Die Gleichheit $\operatorname{ran}(S^{\#}) = \operatorname{ran}(T^{\#})$ ist offenbar äquivalent zu

$$\ker(I - S^{\#}S) = \ker(I - T^{\#}T),$$

eine Aussage, die wir nun beweisen wollen.
Sei $(I - T^{\#}T)(x) = \theta$. Dann ergibt sich

$$\begin{aligned} RS^{\#}S(x) &= T^{\#}S(x) = T^{\#}ST^{\#}T(x) = (T^{\#}TT^{\#} - T^{\#}(T-S)T^{\#})(T(x)) \\ &= (I - T^{\#}(T-S))(T^{\#}T(x))) = R(x) \end{aligned}$$

Da R invertierbar ist, gilt $(I - S^{\#}S)(x) = \theta$.
Sei $(I - S^{\#}S)(x) = \theta$. Setze $Q := I - S^{\#}(S-T)$. Man erhält in analoger Weise $R^{-1}T^{\#}T(x) = Q(x)$. Daraus ergibt sich

$$RQ = R(I - R^{-1}T^{\#}(S-T)) = R - T^{\#}(S-T) = I.$$

Also ist $Q^{-1}R$.
Wir haben

$$\begin{aligned} \|S^{\#}T\| &= \sup\{\|R^{-1}T^{\#}Tx\| | x \in X, \|x\| = 1\} \\ &= \sup\{\|R^{-1}T^{\#}Tx\| | x \in \ker(I - T^{\#}T), \|x\| = 1\} \\ &= \sup\{\|R^{-1}x\| | x \in \ker(I - T^{\#}T), \|x\| = 1\} \\ &= \sup\{\|R^{-1}x\| | x \in X, \|x\| = 1\} \\ &= \|R^{-1}\| = \frac{1}{1 - \|T^{\#}(T-S)\|} \end{aligned}$$

■

Lemma 3.29 *Seien X, Y Banachräume, seien $T, S \in \mathcal{B}(X,Y)$ und seien $T^\#, S^\#$ äußere Inverse von T bzw. S. Dann sind äquivalent:*

(1) $S^\#(I - TT^\#) = 0$.

(2) $ker(T^\#) \subset ker(S^\#)$.

Beweis:
Wir haben

$$\begin{aligned} S^\#(I-TT^\#) = 0 \quad &\Longleftrightarrow \quad S^\#(I-TT^\#)_{|\ker(TT^\#)} = 0 \iff S^\#_{|\ker(TT^\#)} = 0 \\ &\Longleftrightarrow \quad S^\#_{|\ker(T^\#)} = 0 \iff \ker(T^\#) \subset \ker(S^\#) \end{aligned}$$

■

Bemerkung 3.30 *Es gibt viele Versuche, einen Ersatz für die Inverse eines nichtinvertierbearen linearen Operators zu definieren. Hier ist eine Art Klassifikation.*

Seien X, Y Hilberträume – Hilberträume setzen wir voraus, um die Adjungierte eines Operators auf den Ausgangsräumen zur Verfügung zu haben – und seien $T \in \mathcal{B}(X,Y), S \in \mathcal{B}(Y,X)$. Betrachte folgende Eigenschaften:

$$\begin{array}{rcll} TST &=& T & \quad (1) \\ STS &=& S & \quad (2) \\ (TS)^* &=& TS & \quad (3) \\ (ST)^* &=& ST & \quad (4) \end{array}$$

Sei M die Menge der Gleichungen, die gültig sind für das Paar (T, S). Wir nennen S eine **M-Inverse** *für T, falls $M \neq \emptyset$ ist. Beispielsweise:*

(1) Eine $\{1\}$-Inverse wird **innere Inverse** *genannt.*

(2) Eine $\{2\}$-Inverse wird **äußere Inverse** *genannt; siehe oben.*

(3) Eine $\{1,2,3,4\}$-Inverse wird **Moore-Penrose-Inverse** *genannt.*

□

3.8 Anhang: Surjektivitäts- und Injektivitätsmodul

Wir haben die Newtoniteration

$$DG(x^k)v^k = -G(x^k)\,,\ x^{k+1} := x^k + v^k\,,\ k \in \mathbb{N}_0\,, \tag{3.42}$$

für die Gleichung

$$G(x) = \theta \tag{3.43}$$

betrachtet. Wenn die Gleichung $DG(x^k)v^k = -G(x^k)$ lösbar ist, ist die Iteration definiert; die Surjektivität von $DG(x^k)$ reicht dafür aus. Ist diese Gleichun3.15g nicht eindeutig lösbar, stellt sich aber die Frage der Auswahl einer Lösung. Als vernünftige Auswahl könnte

$$v^k := \arg\min\{\|v\| | v \in X, DG(x^k)v^k = -G(x^k)\}$$

angesehen werden. Hier stellt sich dann im unendlichdimensionalen Kontext wieder eine Existenzproblem. Für damit zusammenhängende Fragen dienen die folgenden Ergebnisse.

Definition 3.31 *Seien* X, Y *Banachräume und sei* $T \in \mathcal{B}(X,Y)$. *Dann heißen*

$$inj(T) := \inf\{\|Tx\| \| \|x\| = 1\}\,,\ sur(T) := \sup\{s \geq 0 | \overline{B}_s \subset T(\overline{B}_1)\}$$

der **Injektivitätsmodul** *bzw. der* **Surjektivitätsmodul** *von* T. □

Lemma 3.32 *Seien* X, Y *Banachräume und sei* $T \in \mathcal{B}(X,Y)$. *Dann sind äquivalent:*

(a) T *ist injektiv und hat abgeschlossenes Bild.*

(b) $inj(T) > 0$.

Beweis:
Sei $Z := \mathrm{ran}(T)$. Z ist ein normierter linearer Teilraum von Y. Zu $(a) \implies (b)$ Z ist ein Banachraum und $\hat{T} : X \ni x \longmapsto Tx \in Z$ kann als bijektiver linearer stetiger Operator betrachtet werden. Nach dem Satz über die stetige Inverse (siehe [41]) hat $\hat{T}$ eine stetige lineare Inverse und es gibt $c > 0$ mit $\|\hat{T}^{-1}z| \leq c\|z\|$ für alle $z \in Z$. Also haben wir $\mathrm{inj}(Z) \geq c^{-1} > 0$.
Zu $(b) \implies (a)$ Die Injektivität ist klar. Sei $(z_n)_{n\in\mathbb{N}}$ eine Folge in Z mit $\lim_n z_n = z$. Da T injektiv ist, gibt es eine Folge $(x_n)_{n\in\mathbb{N}}$ mit $Tx_n = z_n, n \in \mathbb{N}$. Wegen

$$\|z_m - z_n\| = \|Tx_m - Tx_n\| \geq \mathrm{inj}\|x_m - x_n\|\,,\ m, n \in \mathbb{N}$$

ist $(x_n)_{n\in\mathbb{N}}$ eine Cauchyfolge in X. Sei $x := \lim_n x_n$. Dann folgt mit der Stetigkeit von $Tx = \lim_T x_n = \lim_n z_n = z$. Also ist $\mathrm{ran}(T)$ abgeschlossen. ■

Hier ist eine etwas erweiterte Fassung des Satzes über die Offenheit linearer Abbildungen.

Satz 3.33 *Seien* X, Y *Banachräume und sei* $T \in \mathcal{B}(X,Y)$. *Dann sind äquivalent:*

(a) T *ist surjektiv.*

(b) T *ist eine offenen Abbildung.*

(c) $sur(T) > 0$.

(d) *Es gibt* $Q : Y \longrightarrow X$ *(Rechtsinverse, nicht notwendigerweise linear) mit* $TQ = I$.

Zusatz zu (d): Ist $b \in (0, \infty)$ *mit* $b^{-1} < sur(T)$, *dann ist* $\|Qy\| \leq b\|y\|$ *für alle* $y \in Y$.

Beweis:
Zu $(a) \implies (b)$ Dies ist Inhalt des klassischen Satzes über die Offenheit von bijektiven stetigen linearen Operatoren; siehe etwa [6, 41].
Zu $(b) \implies (a), (b) \implies (c), (d) \implies (a)$ Trivial.
Zu $(c) \implies (d)$ Sei $b^{-1} \in (0, \mathrm{sur}(T))$. Dann ist offenbar $\overline{B}_{b^{-1}} \backslash B_{b^{-1}} \subset T(\overline{B}_1)$. Definiere $Q : Y \longrightarrow X$ in folgender Weise:

$Q(y) := \theta$, falls $y = \theta$, $Q(y) := b\|y\|\hat{x}$ wobei $T\hat{x} = \hat{y}$ mit $\hat{y} = b^{-1}\|y\|^{-1}y \in \overline{B}_{b^{-1}} \backslash B_{b^{-1}}$, für $y \neq \theta$.

Dann gilt $T(Q(y)) = y$ für alle $y \in Y$. Q ist also eine Rechtsinverse von T. Ferner

$$\|Q(y)\| = b\|y\|\,,\ y \in Y\,.$$

■

Lemma 3.34 *Seien* X, Y *Banachräume und sei* $T \in \mathcal{B}(X,Y)$. *Dann gilt:*

$$inj(T) = sur(T^*)\,,\ sur(T) = inj(T^*)$$

Beweis:
Wir beweisen nur die zweite Gleichheit; für einen Beweis der ersten Identität siehe [32].
Es gilt

$$\begin{aligned} \operatorname{sur}(T) > 0 \quad &\iff \quad \operatorname{ran}(T) = Y, \operatorname{ran}(T) \text{ abgeschlossen} \\ &\iff \quad \ker(T^*) = \{\theta\}, \operatorname{ran}(T^*) \text{ abgeschlossen} \quad \iff \quad \operatorname{inj}(T^*) > 0\,. \end{aligned}$$

Dies zeigt insbesondere $\operatorname{sur}(T) = 0$ genau dann, wenn $\operatorname{inj}(T^*) = 0$ ist. Es bleibt die zweite Ungleichung zu zeigen unter der Bedingung $\operatorname{sur}(T) > 0, \operatorname{inj}(T^*) > 0$.
Zu $\operatorname{sur}(T) \le \operatorname{inj}(T^*)$. Sei $s \in (0, \operatorname{sur}(T))$.
Sei $\lambda \in X^*$ mit $\|\lambda\| = 1$ und sei $\vartheta \in (0,1)$. Dann können wir $y \in \overline{B}_1$ wählen mit $|\langle \lambda, y\rangle| \ge \vartheta$.
Wegen $\overline{B}_s \subset T(\overline{B}_1)$ finden wir $x \in \overline{B}_{s^{-1}}$ mit $Tx = y$. Da $sx \in \overline{B}_1$ ist, folgt

$$\|T^*\lambda\| \ge |\langle T^*(\lambda), sx\rangle| = s|\langle \lambda, Tx\rangle| = s|\langle \lambda, y\rangle| \ge s\vartheta\,.$$

Daraus folgt $\operatorname{sur}(T) \le \operatorname{inj}(T^*)$ durch Bildung der Suprema bzgl. ϑ, λ und s.
Zu $\operatorname{sur}(T) \ge \operatorname{inj}(T^*)$. Sei $s \in (0, \operatorname{inj}(T^*))$.
Annahme: Es gilt nicht $\overline{B}_s \subset \overline{T(\overline{B}_1)}$.
Dann gibt es $z \in \overline{B}_s$ mit $z \notin C := \overline{T(\overline{B}_1)}$. Mit dem Satz von Hahn-Banach für konvexe Mengen (siehe [6, 41]) folgt die Existenz von $\lambda \in Y^*$ mit

$$\langle \lambda, y\rangle \le 1 \le \langle \lambda, z\rangle \text{ für alle } y \in C\,.$$

Also

$$\|T^*\lambda\| = \sup_{x\in X, \|\lambda\|\le 1} |\langle T^*\lambda, x\rangle| = \sup_{x \in X, \|x\| \le 1} |\langle \lambda, Tx\rangle| \le \sup_{y\in C} |\langle \lambda, y\rangle| \le 1 < \lambda, z\rangle \le s\|\lambda\|\,,$$

also $\|T^*(\lambda\|\lambda\|^{-1})\| \le s < \operatorname{inj}(T^*)$, was der Definition von $\operatorname{inj}(T^*)$ widerspricht.
Also gilt nun $\overline{B}_s \subset \overline{T(\overline{B}_1)}$. Daraus schließt man

$$\overline{B}_{\vartheta s} \subset T(\overline{B}_1)\,,\ \vartheta \in (0,1)\,.$$

Dies zeigt $\operatorname{sur}(T) \ge \vartheta s$. Daraus folgt $\operatorname{sur}(T) \ge \operatorname{inj}(T^*)$. ■

Lemma 3.35 *Seien X, Y Banachräume und seien $T, S \in \mathcal{B}(X,Y)$. Dann gilt:*

$$|\mathit{inj}(T) - \mathit{inj}(S)| \le \|T - S\|\,,\ |\mathit{sur}(T) - \mathit{sur}(S)| \le \|T-S\|$$

Beweis:
Zur ersten Abschätzung: O. E. können wir annehmen: $\operatorname{inj}(T) \ge \operatorname{inj}(S)$. Dann haben wir

$$\begin{aligned} \operatorname{inj}(S) \ &= \inf_{x\in X, \|x\|=1} \|Sx\| \ \ge \inf_{x \in X, \|x\|=1} (\|Tx\| - \|Tx - Sx\|) \\ &\ge \inf_{x\in X, \|x\|=1} \|Tx\| - \sup_{x\in X, \|x\|=1} \|(T-S)x\| \ = \ \operatorname{inj}(T) - \|T - S\| \end{aligned}$$

also

$$|\operatorname{inj}(T) - \operatorname{inj}(S)| = \operatorname{inj}(T) - \operatorname{inj}(S) \le \|T - S\|\,.$$

Zur zweiten Abschätzung: Wenn wir die Identitäten aus Lemma 3.34 und die eben bewiesene erste Abschätzung nutzen, folgt

$$|\operatorname{sur}(T) - \operatorname{sur}(S)| = |\operatorname{inj}(T^*) - \operatorname{inj}(S^*)| \le \|T^* - S^*\| = \|T - S\|$$

■

3.9 Übungen

1.) Betrachte das Polynom $f(x) := x^3 - 3x^2 + 3x - 1$ und dazu die Nullstellengleichung $f(x) = 0$.

   (a) Berechne zwei Newtonnäherungen x^1, x^2, ausgehend vom Startwert $x^0 := 1.1$.

   (b) Zeige, dass das Newtoverfahren für alle Startwerte $x^0 \in \mathbb{R}$ konvergiert.

   (c) Welche Nullstellen hat das Polynom f.

2.) Betrachte in $\mathbb{R}$ die Gleichung

$$5 - x = \frac{x}{\sqrt{x^2 - 1}}.$$

   Löse diese Gleichung mit einer Genauigkeit von 10^{-2}.

3.) Betrachte die Funktion $f(x) := \sin(x) - 0.5x - 0.1$ und dazu die Nullstellengleichung $f(x) = 0$.

   (a) Zeige: f hat eine Nullstelle x^* in $(0, 0.3)$.

   (b) Zeige: $|x^0 - x^*| < 0.2$ für $x^0 := 0.1$.

   (c) Zeige für die Newtoniterierten $(x^n)_{n\in\mathbb{N}}$ (Startwert $x^0 := 0.1$)

$$|x^{n+1} - x^*| \leq \frac{2}{3}|x^n - x^*|^2, n \in \mathbb{N}_0\,,\ |x^n - x^*| < 0.2, n \in \mathbb{N}\,,$$

   und $\lim_n x^n = x^*$.

4.) Sei $f : \mathbb{R} \longrightarrow \mathbb{R}$ stetig differenzierbar. Betrachte mit einem Startwert x^0 dazu die Iteration

$$x^{k+1} := x^k - \frac{f(x^k)^2}{f(x^k) - f(x^k - f(x^k))}\,,\ k \in \mathbb{N}_0\,.$$

   Zeige: $\lim_k x^k = x^*$, wenn f nur eine einfache Nullstelle besitzt.

5.) Sei $f : \mathbb{R} \ni x \longmapsto \cosh(x) - 2 \in \mathbb{R}$. Für welche Starwerte x^0 konvergiert die Newtoniteration?

6.) Betrachte das Gleichungssystem

$$x^2 + y^2 - 9 = 0\,,\ x + y - 1 = 0\,.$$

   (a) Zeige: Es gibt genau zwei Lösungen.

   (b) Schreibe das Newton–Verfahren zur Lösung des Systems auf.

   (c) Gib zu jeder Lösung einen Startwert an, so dass das Newton-Verfahren gegen diese Lösung konvergiert.

7.) Betrachte

$$F(x, y, z) := (x\cos(y) - z, x^2 + z, e^{x+z}\sin(\frac{y}{2} + z))\,,\ (x, y, z) \in \mathbb{R}^3\,.$$

   (a) Gib mindestens zwei Lösungen für die Gleichung

$$F(x, y.z) = \theta$$

   an.

   (b) Formuliere das Newton–Verfahren mit Startwert $(\frac{1}{4}, \frac{\pi}{2}, -\frac{1}{2})$ zur Berechnung einer Nullstelle und berechne damit eine Nullstellennäherung $(\overline{x}, \overline{y}, \overline{z})$ mit $|F(\overline{x}, \overline{y}, \overline{z})| < 10^{-2}$.

8.) Für eine invertierbare Matrix $A \in \mathbb{R}^{n,n}$ ist die Matrix $X := A^{-1}$ offensichtlich eine Lösung der (nichtlinearen) Gleichung

$$X^{-1} - A = \Theta\,. \tag{3.44}$$

Zeige:

(a) Die Iteration

$$X^{n+1} := X^n + X^n(E - AX^n)\,,\; n = 0.1, \dots, \tag{3.45}$$

stellt das Newton–Verfahren zur Lösung von (3.44) dar.

(b) Für jede Startmatrix X^0 mit $\|E - AX^0\| \le q < 1$ ($\|\cdot\|$ ist eine Matrixnorm) konvergiert die in (3.45) erklärte Folge $(X^n)_{n\in\mathbb{N}}$ gegen die Matrix A^{-1} und es gelten die Abschätzungen

$$\|X^n - A^{-1}\| \le (1-q)^{-1}\|X^0\|\|E - AX^n\| \le (1-q)^{-1}\|X^0\|q^{2^n}\,,\; n = 0, 1, \dots\,.$$

9.) Sei X ein normierter Raum und sei $F : X \longrightarrow X$ Fréchet-differenzierbar. Betrachte die Newton-Iteration:

$$x^{n+1} := x^n - DF(x^n)^{-1}F(x^n), n \in \mathbb{N}_0; x^0 \text{ Startwert}\,.$$

Die Iterationsfolge $(x^n)_{n\in\mathbb{N}}$ sei wohldefiniert und konvergent mit Grenzwert x^*. Zeige: Ist DF stetig in x^*, so gilt $F(x^*) = \theta$.

10.) Sei X ein normierter Raum und sei $F : X \longrightarrow X$ zweimal Fréchet-differenzierbar. Betrachte die Newton-Iteration:

$$x^{n+1} := x^n - DF(x^n)^{-1}F(x^n), n \in \mathbb{N}_0; x^0 \text{ Startwert}\,.$$

Die Iterationsfolge $(x^n)_{n\in\mathbb{N}}$ sei wohldefiniert und konvergent mit Grenzwert x^*. Zeige: Ist D^2F beschränkt in einer Umgebung von x^*, so gilt $F(x^*) = \theta$.

11.) Betrachte die Funktion

$$F : (0, \infty) \ni x \longmapsto \frac{x^{1+1/n}}{1 + 1/n} + c_1x + c_0 \in \mathbb{R}$$

mit $c_1, c_0 > 0$. Sei $x^0 \in (0, \infty)$. Zeige:

(a) F ist nicht Lipschitzstetig.

(b) Es gibt $L \ge 0$ mit

$$|DF(x^0)^{-1}(DF(x) - DF(x^0))| \le L|x - x^0|,\,,\; x \in (0, \infty)\,.$$

12.) Betrachte im Banachraum $X := C[0, 1]$ (versehen mit der Maximumsnorm) die Integralgleichung $F(x) = \theta$ mit

$$F(x)(t) := \frac{1}{4}x(t)\int_0^1 \frac{t}{t+s}x(s)ds + 1 - x(t), t \in [0, 1]\,.$$

Berechne die Konstanten in den Voraussetzungen des Satzes von Kantorovich in x^0 mit $x^0(t) = 1, t \in [0, 1]$.

13.) Betrachte die Funktion $F : \mathbb{R} \longrightarrow \mathbb{R}$ mit

$$F(t) := \frac{1}{6}t^3 + \frac{1}{6}t^2 - \frac{5}{6}t + \frac{1}{3}, t \in \mathbb{R}\,.$$

Zeige: F hat eine Nullstelle in $\mathbb{R}$ und zwei Nullstellen in $\mathbb{C}$.

14.) Betrachte die Abbildung $G : \mathbb{R}^2 \ni (x_1, x_2) \longmapsto (x_1x_2 - x_1, x_1x_2 + x_2) \in \mathbb{R}^2$. Zeige: Es gibt $r^* > 0$ mit

$$G(\overline{B}_r) \text{ ist konvex für alle } r \in [0, r^*] .$$

15.) Sei X ein Banachraum und sei $K \subset X$ abgeschlossen. Zeige: K ist konvex genau dann, wenn $\frac{1}{2}(x_1 + x_2) \in K$ gilt für alle $x_1, x_2 \in K$.

16.) Sei X ein gleichmäßig konvexer Banachraum mit Konvexitätsmodul

$$\delta_X(\varepsilon) := \inf\{1 - \frac{1}{2}\|x_1 + x_2\| \,|x_1, x_2 \in \overline{B}_1, \|x_1 - x_2\| \geq \varepsilon\} \geq c\varepsilon^2 , \, \varepsilon \in (0, 2]\, (c > 0) .$$

Zeige: Für alle $x_0, x_1, x_2 \in X, r > 0$, mit $x_1, x_2 \in \overline{B}_r(x_0)$ gilt mit $s := \frac{c\|x_1 - x_2\|^2}{r}$ $\frac{1}{2}(x_1 + x_2) \in \overline{B}_r(x_0)$.

17.) Betrachte das Polynom $f : \mathbb{C} \ni z \longmapsto z^3 - 1 \in \mathbb{C}$.

(a) Berechne die Nullstellen von f.

(b) Die Übertragung des Newtonverfahrens zur Berechnung der Nullstellen ist:

$$z_{k+1} = z_k - f'(z_k)f(z_k), k = 0, 1, \ldots ; z_0 \text{ Startwert}$$

Betrachte $S_{0,k} := \{z_0 | z_k = 0\}, k = 0, 1, \ldots,$ und bestimme ein $z_0 \in S_{0,1}$.

(c) Bestimme $\#S_{0,k}, k = 0, 1, \ldots$.

18.) Betrachte die Abbildung $f : \mathbb{R} \ni x \longmapsto x(1 - |x|) \in \mathbb{R}$.

(a) Berechne die Nullstelle(n) von f.

(b) Ist f differenzierbar?

(c) Ist der Satz von Kantorovich anwendbar für $x^0 \in [-\frac{1}{2}, \frac{1}{2}]$?

19.) Sei $a \in \mathbb{R}, b \in (0, 1)$. Zeige, dass die Gleichung

$$x = a + b\sin(x)$$

genau eine Lösung besitzt.

3.10 Bibliographische und historische Anmerkungen

Die Literatur zum Newtonverfahren und zum Satz von Kantorovich ist dank der Qualität des Verfahrens bzw. des Satzes umfangreich. Grundzüge der historischen Entwicklung findet man etwa in [13, 42]. Hier sind einige Anmerkungen zu unterschiedlichen Entwicklungen.

Beweistechniken zum Newtonverfahren und zum Satz von Kantorovich und Anwendungen findet man etwa in [3, 11, 12, 19, 31, 33, 34, 35]. Verbindungen zum Satz der inversen Funktion werden beschrieben in [7, 10, 17, 22]. Hier ordnet sich auch das Verfahren von Moser ein; siehe [21, 27] ein.

Das deformierte Newtonverfahren im Reellen (siehe [20]) wird auf Banachräume übertragen in [25]. Inexakte Varianten werden beschrieben in [16, 18]. Verbindungen zwischen den Nullstellensätzen von Kantorovich, Miranda und Borsuk werden diskutiert in [1, 26]. Die Verallgemeinerung des Satzes von Kantorovich in den Kontext von Mannigfaltigkeiten ist zu finden in [8]. Kontiniuierliche Analoga werden u. a. beschrieben in [9, 30].

Zu den Abschnitten 3.5, 3.6, 3.8, 3.7 siehe [4, 33, 28, 29, 32, 36, 40]. Wir folgen im wesentlichen [15],

Literaturverzeichnis

[1] G. Ahlefeld, A. Frommer, G. Heindl, and J. Mayer. On the existence theorems of Kantorovich, Miranda and Borsuk. *Electronic transactions on numerical analysis*, 17:102–111, 2004.

[2] M. Anitescu, D.L. Coroian, M.Z. Nashed, and F.A. Potra. Outer inverses and multi-body system simulation. *Numer. Funct. Anal. Optim.*, 17:661–678, 1996.

[3] I.K. Argyros. On a theorem of I.V. Kantorovich concerning Newton's method. *J. Comp. Applied Math.*, 155:223–230, 2003.

[4] I. Banakh, T. Banakh, A. Plichko, and A. Prykarpatsky. On the local convexity of nonlinear mappings between Banach spaces. *Centr. Eur. J. Math.*, 10:2264–2271, 2012.

[5] J. Baumeister. Inverse problems in finance. In *Recent Developments in Computational Finance: Foundations, Algorithms and Applications*, volume 14 of *Interdisciplinary Mathematical Sciences*. World Scientific, 2012.

[6] J. Baumeister. Funktionalanalysis. Goethe–Universität Frankfurt/Main, 2013. Skriptum einer Vorlesung.

[7] M. Berti, P. Bolle, and M. Processi. An abstract Nash-Moser theorem with parameters and applications to PDEs. *Annales IHP (C)*, 27:377–399, 2010.

[8] T. Bittencourt and O.P. Ferreira. Robust Kantorovich's theorem on Newton's method under majorant conditions in Riemannian manifolds. *arXiv:1505.0557301 [math.OC]*, 21. May 2015.

[9] L. Bittner. Einige kontinuierliche Analogien von Iterationsverfahren. In *International Series of Numerical Mathematics*, pages 114–135, Basel, 1967. Birkhäuser.

[10] A. Castro and J. Neuberger. An inverse function theorem via continuous Newton's method. *Nonlinear Analysis*, 47:3223–3229, 2001.

[11] X. Chen and M.Z. Nashed. Convergence of Newton-like methods for singular operator equations using outer inverses. *Numer. Math.*, 66:235–257, 1993.

[12] P.G. Ciarlet and C. Mardare. On the Newton-Kantorovich theorem. *Analysis and Applications*, 10:249–269, 2012.

[13] P. Deuflhard. A short history of Newton's method. *Documenta Mathematica*, Extra volume ISMP:25–30, 2012.

[14] P. Deuflhard and G. Heindl. Affine invariant convergence theorems for Newton's method and extensions to related methods. *SIAM J. Numer. Anal.*, 16:1–10, 1979.

[15] K. Dietrich. Variants of the Newton-Kantorovich theorem. Technische Universität München, 2013. Bachelorarbeit.

[16] S.C. Eisenstat and H.F. Walker. Choosing the forcing terms in an inexact newton method. *SIAM J. Sci. Computing*, 17:16–32, 1994.

[17] I. Ekeland. An inverse function theorem in Fréchet spaces. *Annales IHP (C)*, 28:91–105, 2011.

[18] O.P. Ferreira and B.F. Svaiter. A robust Kantorovich's theorem on inexact Newton method with relative residual error tolerance, 2011. arXiv:1110.3430v1 [math.NA] 15 Oct 2011.

[19] O.P. Ferreira and B.F. Svaiter. Kantorovich's theorem and Newton's method, 2012. arXiv:1209.5704v1 [math.NA] 25 Sep 2012.

[20] M. Frontini and E. Sormani. Some variants of Newton's method with third-order convergence. *Applied Mathematics and Computation*, 140:419–426, 2003.

[21] O.H. Hald. On a Newton-Moser type method. *Numerische Mathematik*, 23:411–426, 1975.

[22] R.S. Hamilton. The inverse function teorem of Nash-Moser. *Bull. Amer. Math. Soc.*, 7:65–222, 1982.

[23] L.V. Kantorovich. On Newton's method for functional equations. *Akad. Nauk. SSSR*, 59:1237–1240, 1948.

[24] L.V. Kantorovich and G.P. Akilov. *Functional analysis in normed spaces.* Pergamon, New York, 1964.

[25] R. Lin, Y. Zhao, Z. Smarda, Y. Khan, and Q. Wu. Newton-Kantorovich and Smale uniform type convergence theorem for a deformed Newton method in Banach spaces. *Abstract and Applied Analysis*, page 8 pages, 2013.

[26] J. Mayer. A generalized theorem of miranda and the theorem of newton-kantorovich. *Computing*, 23:313–357, 2002.

[27] J. Moser. A rapidly convergent iteration method and non-linear differential equations. *Ann. Scuola Norm. Sup. Pisa*, 20:499–535, 1966.

[28] V. Müller. *Spectral Theory of Linear Operators and Spectral Systems in Banach Algebras.* Birkhäuser, Basel, 2007.

[29] M.Z. Nashed. Inner, outer and generalized inverses in Banach and Hilbert spaces. *Numer. Funct. Anal. Optim.*, 9:261–325, 1987.

[30] J. Neuberger. Continuous newton's method for polynomials. *Math. Intelligencer*, 21:18–23, 1999.

[31] J.M. Ortega. The Newton-Kantorovich Theorem. *The American Mathematical Monthly*, 75:658–660, 1968.

[32] A. Pietsch. *Operator Ideals.* North-Holland, Amsterdam, 1980.

[33] B.T. Polyak. Gradient methods for solving equations and inequalities. *USSR Comp. Math. Phys.*, 4:17–32, 1964.

[34] B.T. Polyak. Convexity principle and its applications. *Bull. Braz. Math. Soc.*, 34:59–75, 2003.

[35] B.T. Polyak. Newton-Kantorovich method and its global convergence. *J. of Math. Sciences*, 133:1513–1523, 2006.

[36] B. Shekhtman. On some problems of m. z. nashed on outer inverses. *Linear Algebra*, 76:149–152, 1986.

[37] S. Smale. A convergent process of price adjustment and global newton methods. *J. Math. Econom.*, 3:107–120, 1976.

[38] P. Spellucci. *Numerische Verfahren der nichtlinearen Optimierung.* Birhäuser, Basel, 1993.

[39] R.A. Tapia. The Kantorovich Theorem for Newton's method. *The American Mathematical Monthly*, 78:389–392, 1971.

[40] A. Uderzo. On the Polyak convexity principle and its applications to variational analysis. *arXiv:1303.7443[math.OC]*, 2013.

[41] D. Werner. *Funktionalanalysis.* Springer, 2002.

[42] T. Ypma. Historical developments of th Newton-Raphson method. *SIAM Review*, 124:1–23, 2000.