Kapitel 5

Nichtlineare Gleichungssysteme

5.1 Einführung

Wir betrachten in diesem Kapitel Verfahren zur Lösung von nichtlinearen Gleichungssystemen.

Nichtlineares Gleichungssystem: Gesucht ist eine Lösung $x \in D$ von

$$F(x) = 0$$

mit einer gegebenen Abbildung

$$F = \begin{pmatrix} F_1 \\ \vdots \\ F_n \end{pmatrix} : D \subset \mathbb{R}^n \to \mathbb{R}^n,$$

$D \subset \mathbb{R}^n$ nichtleer und abgeschlossen.

Viele praxisrelevante Probleme, insbesondere im Hochtechnologiebereich, sind nichtlinear und erfordern die Lösung nichtlinearer Gleichungssysteme. So führt zum Beispiel die Schaltkreissimulation und die Diskretisierung nichtlinearer partieller Differentialgleichungen (Wetter- und Klimamodelle, strukturmechanische Berechnungen, Umformprozesse aus der Produktionstechnik...) auf große nichtlineare Gleichungssysteme.

Im Gegensatz zu linearen Gleichungssystemen, bei denen nur genau eine Lösung, keine Lösung oder ein ganzer affiner Unterraum als Lösung auftreten kann, sind bei nichtlinearen Gleichungen auch mehrere oder unendlich viele isolierte Lösungen möglich.

Beispiel 5.1.1

1. $n = 1$, $D = \mathbb{R}$, $F(x) = x^2 - a$, $a > 0$.
   Es gibt zwei reelle Lösungen $x = \pm\sqrt{a}$.

2. $n = 1$, $D = \mathbb{R}$, $F(x) = x^2 + a$, $a > 0$.

   Es existiert keine reelle Lösung.

3. $n = 1$, $D = \mathbb{R}$, $F(x) = x \sin(x)$.

   Es gibt unendlich viele Lösungen $x = k\pi$, $k \in \mathbb{Z}$.

4. Schnittpunkte des Einheitskreises mit der Geraden $G : x_2 = ax_1 + b$, $a, b \in \mathbb{R}$: $n = 2$, $D = \mathbb{R}^2$, $F(x) = \binom{x_1^2+x_2^2-1}{x_2-ax_1-b}$.

   Je nach Wahl von a, b gibt es zwei, eine oder keine relle Lösung.

Sehr oft ist die Funktion F stetig differenzierbar, d.h. die partiellen Ableitungen $\frac{\partial F_i}{\partial x_j}$, $1 \leq i, j \leq n$ existieren und sind stetig. In diesem Fall gilt (Taylorentwicklung erster Ordnung)

$$F(x + s) = F(x) + F'(x)s + R(x; s)$$

mit der Jacobi-Matrix

$$F'(x) = \begin{pmatrix} \frac{\partial F_1}{\partial x_1}(x) & \cdots & \frac{\partial F_1}{\partial x_n}(x) \\ \vdots & & \vdots \\ \frac{\partial F_n}{\partial x_1}(x) & \cdots & \frac{\partial F_n}{\partial x_n}(x) \end{pmatrix}.$$

und einem Restglied $R(x; s)$, wobei

$$\lim_{s \to 0} \frac{\|R(x; s)\|}{\|s\|} = 0, \quad \text{kurz:} \quad R(x; s) = o(\|s\|).$$

Dies ist wesentlich für die Entwicklung schneller Lösungsverfahren.

5.2 Das Newton-Verfahren

Das Newton-Verfahren ist eines der wichtigsten Verfahren zur Lösung nichtlinearer Gleichungssysteme, da es nahe der Lösung sehr schnell konvergiert. Der Einfachheit halber nehmen wir im folgenden den Fall $D = \mathbb{R}^n$ an.

Wir betrachten das Newton-Verfahren zur Lösung eines nichtlinearen Gleichungssystems

$$F(x) = 0 \tag{5.1}$$

mit $F : \mathbb{R}^n \to \mathbb{R}^n$ stetig differenzierbar.

5.2.1 Herleitung des Verfahrens

Anschauliche Herleitung im eindimensionalen Fall

Sei zunächst $n = 1$. Dann ist $F(x)$ eine reelle Funktion. Sei $x^{(k)}$ eine Näherung einer Lösung $\bar{x}$ von (5.1). Die Idee des Newton-Verfahrens besteht darin, F in $x^{(k)}$ durch die Tangente an $(x, F(x))$ im Punkt $x^{(k)}$ zu approximieren und als nächste Iterierte $x^{(k+1)}$ die Nullstelle der Tangente zu wählen.

Die Tangentengleichung lautet

$$y = F(x^{(k)}) + F'(x^{(k)})(x - x^{(k)})$$

und $x^{(k+1)}$ ist die Lösung von

$$F(x^{(k)}) + F'(x^{(k)})(x - x^{(k)}) = 0.$$

Im Falle $F'(x^{(k)}) \neq 0$ ergibt sich

$$x^{(k+1)} = x^{(k)} - F'(x^{(k)})^{-1}F(x^{(k)}).$$

Es gilt also

$$x^{(k+1)} = x^{(k)} + s^{(k)},$$

wobei $s^{(k)}$ die Lösung der Gleichung ist

$$F'(x^{(k)})s^{(k)} = -F(x^{(k)}).$$

Beispiel: Für $F(x) = x^2 - a$, $a > 0$ ergibt sich

$$x^{(k+1)} = x^{(k)} - \frac{1}{2x^{(k)}}((x^{(k)})^2 - a) = \frac{1}{2}\left(x^{(k)} + \frac{a}{x^{(k)}}\right).$$

Der allgemeine Fall

Zur allgemeinen Motivation des Newton-Verfahrens für (5.1) sei $x^{(k)} \in \mathbb{R}^n$ ein gegebener Punkt. Dann ist $\bar{x}$ eine Lösung von (5.1) genau dann, wenn $\bar{x} = x^{(k)} + s$ gilt mit einer Lösung s von

$$F(x^{(k)} + s) = 0. \tag{5.2}$$

Die Idee des Newton-Verfahrens besteht darin, $F(x^{(k)} + s)$ durch die Taylorentwicklung erster Ordnung zu ersetzen: Es gilt

$$F(x^{(k)} + s) = F(x^{(k)}) + F'(x^{(k)})s + o(\|s\|)$$

mit der Jacobi-Matrix $F'(x^{(k)})$ von F in $x^{(k)}$ und das Restglied wird für kurze s klein.

Bei der k-ten Iteration des Newton-Verfahrens ersetzt man daher (5.2) durch die linearisierte Gleichung

$$F(x^{(k)}) + F'(x^{(k)})s = 0.$$

Dies ergibt

Algorithmus 4 Lokales Newton-Verfahren für Gleichungssysteme

Wähle einen Startpunkt $x^{(0)} \in \mathbb{R}^n$.

Für $k = 0, 1, \dots$:

1. *Falls* $F(x^{(k)}) = 0$*: STOP mit Ergebnis* $x^{(k)}$.
2. *Berechne den Newton-Schritt* $s^{(k)} \in \mathbb{R}^n$ *durch Lösen der* Newton-Gleichung
$$F'(x^{(k)})s^{(k)} = -F(x^{(k)}).$$
3. *Setze* $x^{(k+1)} = x^{(k)} + s^{(k)}$.

5.2.2 Superlineare und quadratische lokale Konvergenz des Newton-Verfahrens

Wir werden sehen, dass unter geeigneten Voraussetzungen die schnelle lokale Konvergenz des Newton-Verfahrens gezeigt werden kann.

Wir verwenden im folgenden der Einfachheit halber immer die euklidische Norm $\|\cdot\|_2$ mit induzierter Matrix-Norm $\|\cdot\|_2$, obwohl wir genausogut jede andere Norm verwenden könnten.

Der folgende Satz zeigt die superlineare bzw. quadratische lokale Konvergenz des Newton-Verfahrens.

Satz 5.2.1 (Schnelle lokale Konvergenz des Newton-Verfahrens)
Sei $F : \mathbb{R}^n \to \mathbb{R}^n$ *stetig differenzierbar und sei* $\bar{x} \in \mathbb{R}^n$ *ein Punkt mit* $F(\bar{x}) = 0$ *und* $F'(\bar{x})$ *nichtsingulär. Dann gibt es* $\delta > 0$*, so dass gilt:*

i) $\bar{x}$ *ist die einzige Nullstelle von* F *auf* $B_\delta(\bar{x})$

ii) Für alle $x^{(0)} \in B_\delta(\bar{x})$ *mit der* δ*-Kugel*
$$B_\delta(\bar{x}) := \{x \in \mathbb{R}^n \ : \ \|x - \bar{x}\|_2 < \delta\}$$
terminiert Algorithmus 4 entweder mit $x^{(k)} = \bar{x}$ *oder erzeugt eine Folge* $(x^{(k)}) \subset B_\delta(\bar{x})$*, die superlinear gegen* $\bar{x}$ *konvergiert, d.h.*
$$\lim_{k\to\infty} x_k = \bar{x}, \quad \textit{wobei } \|x_{k+1} - \bar{x}\|_2 \le \nu_k \|x_k - \bar{x}\|_2$$
mit einer Nullfolge $\nu_k \searrow 0$.

iii) Ist F' Lipschitz-stetig auf $B_\delta(\bar{x})$ mit Konstante L, gilt also

$$\|F'(x) - F'(y)\|_2 \leq L\|x - y\|_2 \quad \forall\, x, y \in B_\delta(\bar{x}),$$

dann konvergiert $(x^{(k)})$ sogar quadratisch gegen $\bar{x}$, d.h.

$$\lim_{k\to\infty} x_k = \bar{x}, \quad \textit{wobei} \;\; \|x_{k+1} - \bar{x}\|_2 \leq C\|x_k - \bar{x}\|_2^2,$$

wobei für $\delta > 0$ klein genug $C = \|F'(\bar{x})^{-1}\|_2 L$ gewählt werden kann.

Hinweis: *F' ist automatisch Lipschitz-stetig auf $B_\delta(\bar{x})$, falls F zweimal stetig differenzierbar auf der abgeschlossenen Kugel $\overline{B_\delta(\bar{x})}$ ist.*

Leider konvergiert das Newton-Verfahren aus Algorithmus 4 in der Regel nur für Startpunkte, die nahe genug an einer Lösung $\bar{x}$ liegen.

Beispiel 5.2.1 Betrachte $F(x) = \frac{x}{\sqrt{1+x^2}}$. F hat die eindeutige Nullstelle $\bar{x}$ und ist stetig differenzierbar mit $F'(x) > 0$. Trotzdem konvergiert das Newton-Verfahren für jeden Startpunkt mit $|x^{(0)}| > 1$ nicht. Siehe Übung.

Um Konvergenz für beliebige Startpunkte erzielen zu können, muss man das Newton-Verfahren geeignet globalisieren.

5.2.3 Globalisierung des Newton-Verfahrens

In diesem Abschnitt beschreiben wir eine Modifikation des Newton-Verfahrens, die für eine große Klasse von Funktionen F globale Konvergenz, d.h. Konvergenz von einem beliebigen Startpunkt aus, sicherstellt.

Den Ausgangspunkt bildet die Beobachtung, dass jede Lösung $\bar{x}$ von (5.1) ein globales Minimum des Minimierungsproblems

$$\min_{x\in\mathbb{R}^n} \|F(x)\|_2^2$$

ist.

Wir wenden nun folgende Strategie an:

- Wir verwenden den Newton-Schritt $s^{(k)}$ mit einer Schrittweite $\sigma_k \in]0, 1]$, wählen also als Ansatz für die neue Iterierte

$$x^{(k+1)} = x^{(k)} + \sigma_k s^{(k)}.$$

- Wir bestimmen die Schrittweite σ_k so, dass gilt

(5.3) $$\|F(x^{(k+1)})\|_2 < \|F(x^{(k)})\|_2,$$

und die Abnahme "ausreichend groß" ist.

Durch Taylorentwicklung der Funktion

$$\phi(\sigma) := \|F(x^{(k)} + \sigma s^{(k)})\|_2^2$$

in $\sigma = 0$ erhält man

$$\phi(\sigma) = \phi(0) + \phi'(0)\sigma + o(\sigma) = \|F(x^{(k)})\|_2^2 + 2\sigma F(x^{(k)})^T F'(x^{(k)}) s^k + o(\sigma)$$

und Einsetzen der Newton-Gleichung $F'(x^{(k)})s^k = -F(x^{(k)})$ liefert

$$\|F(x^{(k)} + \sigma s^{(k)})\|_2^2 = \|F(x^{(k)})\|_2^2 - 2\sigma\|F(x^{(k)})\|_2^2 + o(\sigma).$$

Ist $\delta \in]0,1[$ fest, dann gilt im Fall $F(x^{(k)}) \neq 0$ also für σ klein genug

$$\|F(x^{(k)} + \sigma s^{(k)})\|_2^2 \leq \|F(x^{(k)})\|_2^2 - 2\delta\sigma\|F(x^{(k)})\|_2^2.$$

Dies zeigt, dass die folgende Schrittweitenwahl nach Armijo Sinn macht:

Schrittweitenwahl nach Armijo:

Sei $\delta \in]0,1/2[$ (gute Wahl z.B. $\delta = 10^{-3}$) fest gegeben. Wähle das größte $\sigma_k \in \{1, \frac{1}{2}, \frac{1}{4}, \ldots\}$ mit

(5.4) $$\|F(x^{(k)} + \sigma_k s^{(k)})\|_2^2 \leq \|F(x^{(k)})\|_2^2 - 2\delta\sigma_k\|F(x^{(k)})\|_2^2.$$

Wir erhalten insgesamt folgendes Verfahren:

Algorithmus 5 Globalisiertes Newton-Verfahren für Gleichungssysteme

Wähle einen Startpunkt $x^{(0)} \in \mathbb{R}^n$.

Für $k = 0, 1, \ldots$:

1. *Falls $F(x^{(k)}) = 0$: STOP mit Ergebnis $x^{(k)}$.*
2. *Berechne den Newton-Schritt $s^{(k)} \in \mathbb{R}^n$ durch Lösen der* Newton-Gleichung
$$F'(x^{(k)})s^{(k)} = -F(x^{(k)}).$$
3. *Bestimme σ_k nach der Armijo-Regel* (5.4).

4. *Setze* $x^{(k+1)} = x^{(k)} + \sigma_k s^{(k)}$.

Es gilt folgender Konvergenzsatz.

Satz 5.2.2 *Sei* $F : \mathbb{R}^n \to \mathbb{R}^n$ *stetig differenzierbar und* $x^{(0)} \in \mathbb{R}^n$ *beliebig. Ist* $F'(x)$ *invertierbar für alle* x *in der Niveaumenge*

$$N_f(x^{(0)}) := \left\{ y \,:\, f(y) \leq f(x^{(0)}) \right\}, \quad f(x) = \|F(x)\|_2^2$$

und ist $N_f(x^{(0)})$ *kompakt (also beschränkt und abgeschlossen), dann terminiert Algorithmus 5 mit Startpunkt* $x^{(0)}$ *entweder endlich oder erzeugt eine Folge* $(x^{(k)}) \subset N_f(x^{(0)})$*, für die gilt:*

i) $(x^{(k)})$ *konvergiert gegen eine Lösung* $\bar{x}$ *von* (5.1).

ii) Es gibt $l \geq 0$ *mit* $\sigma_k = 1$ *für alle* $k \geq l$*. Das Verfahren geht also in das lokale Newton-Verfahren über und konvergiert superlinear bzw. quadratisch gegen* $\bar{x}$*.*