Grundlagen der Optimierung

Globalisiertes Newton-Verfahren in der Optimierung

Idee: Kombiniere die globalen Konvergenzeigenschaften des Gradientenverfahrens (Algorithmus 6.1) mit der schnellen lokalen Konvergenz des Newton-Verfahrens (Algorithmus 7.1).

Algorithmus 7.10 (Globalisiertes Newton-Verfahren in der Optimierung)

1: Wähle $x_0 \in \mathbb{R}^n$, $\sigma \in (0, 1/2)$, $\beta \in (0,1)$, $\varrho > 0$, $p > 2$ und setze $k := 0$
2: **while** Abbruchkriterium nicht erfüllt **do**
3: Löse, wenn möglich, das lineare Gleichungssystem $\nabla^2 f(x_k)\, d_k := -\nabla f(x_k)$
4: Ist dieses System nicht oder nicht eindeutig lösbar oder ist die Bedingung

$$\nabla f(x_k)^\top d_k \le -\varrho\, \|d_k\|^p \tag{7.4}$$

verletzt, so setze $d_k := -\nabla f(x_k)$
5: Bestimme eine Schrittweite t_k mit der Armijo-Regel mit der Startschrittweite $s = 1$, d.h., $t_k := \max\{\beta^l : l = 0, 1, 2, \ldots\}$, sodass gilt:

$$f(x_k + t d_k) \le f(x_k) + \sigma\, t\, \nabla f(x_k)^\top d_k$$

6: Setze $x_{k+1} := x_k + t_k d_k$ und $k := k + 1$
7: **end while**

Bemerkung 7.11 (a) Bei unbrauchbarer Newton-Richtung weichen wir also auf einen Gradientenschritt aus. Entweder die Bedingung (7.4) oder aber die Wahl $d_k = -\nabla f(x_k)$ sichert $\nabla f(x_k)^\top d_k < 0$. Die Armijo-Regel liefert also immer eine Schrittweite (Satz 5.1), und der Algorithmus ist wohldefiniert.

(b) Die Vorgabe $\sigma < 1/2$ ist wichtig, damit für hinreichend große $k \in \mathbb{N}$ tatsächlich volle Newton-Schritte ($t_k = 1$) gegangen werden können.

Satz 7.12 **(Globaler Konvergenzsatz für globalisiertes Newton-Verfahren)**
Es sei $\{x_k\}$ eine durch Algorithmus 7.10 erzeugte Folge.

(a) Jeder Häufungspunkt x^* von $\{x_k\}$ ist ein stationärer Punkt von f, erfüllt also $\nabla f(x^*) = 0$.

(b) Ist x^* ein isolierter Häufungspunkt von $\{x_k\}$, dann konvergiert bereits die gesamte Folge $x_k \to x^*$.

Beweis: siehe [1, Satz 9.5 und Satz 9.7] □

Satz 7.13 **(Lokaler Konvergenzsatz für globalisiertes Newton-Verfahren)**
Es seien $\{x_k\}$, $\{d_k\}$ durch den Algorithmus 7.10 erzeugte Folgen. Ist x^* ein Häufungspunkt von $\{x_k\}$ und ist $\nabla^2 f(x^*)$ s.p.d., so gilt:

(a) Die gesamte Folge $\{x_k\}$ konvergiert gegen das strikte lokale Minimum x^*.

(b) Für alle hinreichend großen $k \in \mathbb{N}$ ist die Suchrichtung d_k immer die Newton-Richtung.

(c) Für alle hinreichend großen $k \in \mathbb{N}$ wird die volle Schrittweite $t_k = 1$ akzeptiert.

(d) $\{x_k\}$ konvergiert q-superlinear gegen x^*.

(e) Ist $\nabla^2 f$ Lipschitz-stetig in einer Umgebung von x^*, so konvergiert $\{x_k\}$ q-quadratisch gegen x^*.

Beweis: siehe [1, Satz 9.10] □

Bemerkung 7.14 (a) Ist die Berechnung der Hessematrizen $\nabla^2 f(x_k)$ und/oder das Lösen der linearen Gleichungssysteme $\nabla^2 f(x_k)\, d_k = -\nabla f(x_k)$ zu aufwendig, so weicht man auf sogenannte Quasi-Newton-Verfahren (z.B. BFGS) aus, siehe Vorlesung *Nichtlineare Optimierung* und [1, Abschnitt 11].

(b) In der Praxis kommt in Algorithmus 7.10 auch die nicht-monotone Armijo-Regel zum Einsatz, bei der hinreichender Abstieg nur im Vergleich zum Maximum der letzten Funktionswerte gefordert wird, siehe [1, Ende Abschnitt 9.3, S. 96].

Literatur

[1] C. Geiger and C. Kanzow. *Numerische Verfahren zur Lösung unrestringierter Optimierungsaufgaben.* Springer, New York, 1999.