

Grundlagen der Optimierung

Ein globalisiertes Newton-Verfahren in der Optimierung

Idee: Kombiniere die globalen Konvergenzeigenschaften des Gradientenverfahrens ([Algorithmus 4.4](#)) mit der schnellen lokalen Konvergenz des Newton-Verfahrens ([Algorithmus 5.1](#)).

Algorithmus 5.11 (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 $\nabla^2 f(x_k) d_k := -\nabla f(x_k)$ nach der Newton-Richtung d_k
- 4: Ist dieses System nicht oder nicht eindeutig lösbar oder ist die Bedingung

$$\nabla f(x_k)^\top d_k \leq -\varrho \|d_k\|^p \quad (5.4)$$

verletzt, so setze $d_k := -\nabla f(x_k)$

- 5: Bestimme eine Schrittweite t_k mit der Armijo-Backtracking-Strategie zur Startschrittweite $s = 1$, sodass [\(4.3\)](#) erfüllt ist, also:

$$f(x_k + t d_k) \leq f(x_k) + \sigma t_k \nabla f(x_k)^\top d_k$$

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

Bemerkung 5.12 (zum globalisierten Newton-Verfahren)

- (a) Bei unbrauchbarer Newton-Richtung weichen wir also auf einen Gradientenschritt aus. Entweder die (erfüllte) Bedingung [\(5.4\)](#) oder aber die Wahl $d_k = -\nabla f(x_k)$ sichert $\nabla f(x_k)^\top d_k < 0$. Die Armijo-Backtracking-Strategie liefert also immer eine Schrittweite ([Satz 4.3](#)), und der Algorithmus ist wohldefiniert.
- (b) Anstelle der Gradientenrichtung $d_k = -\nabla f(x_k)$ kann auch hier wieder eine vorkonditionierte Gradientenrichtung $d_k = -\nabla_M f(x_k)$ mit fester s. p. d. Matrix M verwendet werden. Schritt 4 lautet dann:
 - 4: Ist dieses System nicht oder nicht eindeutig lösbar oder ist die Bedingung

$$\nabla f(x_k)^\top d_k \leq -\varrho \|d_k\|_M^p \quad (5.4')$$

verletzt, so setze $d_k := -\nabla_M f(x_k)$

- (c) Als Abbruchbedingungen kommen wiederum diejenigen aus [Bemerkung 4.7](#) zum Einsatz.
- (d) Die Vorgabe $\sigma < 1/2$ ist wesentlich, damit für hinreichend große $k \in \mathbb{N}$ tatsächlich volle Newton-Schritte ($t_k = 1$) gegangen werden können.
- (e) Im praktischen Einsatz kommt in [Algorithmus 5.11](#) auch die nicht-monotone Armijo-Regel zum Einsatz, bei der hinreichender Abstieg nur im Vergleich zum Maximum der letzten Funktionswerte gefordert wird, siehe [Geiger, Kanzow, 1999](#), Ende Abschnitt 9.3, S. 96. \diamond

Wir geben Konvergenzaussagen für [Algorithmus 5.11](#) ohne Abbruchbedingung an, sodass eine unendliche Folge $\{x_k\}$ entsteht.

Satz 5.13 (Globaler Konvergenzsatz für das globalisierte Newton-Verfahren)

Es sei $\{x_k\}$ eine durch [Algorithmus 5.11](#) erzeugte Folge.

- Jeder Häufungspunkt x^* von $\{x_k\}$ ist ein stationärer Punkt von f , erfüllt also $\nabla f(x^*) = 0$.
- Ist x^* ein isolierter Häufungspunkt von $\{x_k\}$, dann konvergiert bereits die gesamte Folge $x_k \rightarrow x^*$.

Beweis: siehe [Geiger, Kanzow, 1999](#), Satz 9.5 und Satz 9.7 □

Satz 5.14 (Lokaler Konvergenzsatz für das globalisierte Newton-Verfahren)

Es seien $\{x_k\}$, $\{d_k\}$ durch den [Algorithmus 5.11](#) erzeugte Folgen. Ist x^* ein Häufungspunkt von $\{x_k\}$ und ist $\nabla^2 f(x^*)$ s. p. d., so gilt:

- Die gesamte Folge $\{x_k\}$ konvergiert gegen das strikte lokale Minimum x^* .
- Für alle hinreichend großen $k \in \mathbb{N}$ ist die Suchrichtung d_k immer die Newton-Richtung, und es wird die volle Schrittweite $t_k = 1$ akzeptiert.
- $\{x_k\}$ konvergiert q-superlinear gegen x^* .
- Ist $\nabla^2 f$ Lipschitz-stetig in einer Umgebung von x^* , so konvergiert $\{x_k\}$ q-quadratisch gegen x^* .

Beweis: siehe [Geiger, Kanzow, 1999](#), Satz 9.10 □

Alle hier besprochenen Basis-Algorithmen zur Lösung freier Optimierungsaufgaben sind **Line-Search-Verfahren**, die in jeder Iteration

- eine Suchrichtung d_k
- und anschließend eine geeignete Schrittweite t_k

bestimmen. Als Alternative sind auch **Trust-Region-Verfahren** etabliert, die beide Schritte gemeinsam durchführen, siehe Vorlesung *Nichtlineare Optimierung* und [Geiger, Kanzow, 1999](#), Abschnitt 14.

Allen besprochenen Verfahren ist gemeinsam, dass sie die Suchrichtung d_k durch Minimierung eines lokalen quadratischen Ersatzmodells

$$q_k(d) := f(x_k) + \nabla f(x_k)^\top d + \frac{1}{2} d^\top B_k d$$

gewinnen, d. h. (bei s. p. d. Matrix B_k) aus dem LGS

$$B_k d_k = -\nabla f(x_k).$$

Folgende Tabelle fasst typische Eigenschaften dieser Verfahren zusammen:

Gradientenverfahren	$B_k = I$	q-linear, einfach
vork. Gradientenverfahren	$B_k = M$	q-linear, einfach
Quasi-Newton-Verf.	je nach Verfahren	q-superlinear, oft guter Kompromiss
Newton-Verfahren	$B_k = \nabla^2 f(x_k)$	q-superlinear oder besser, aufwendig

Mehr dazu in der Vorlesung *Nichtlineare Optimierung*.

Literatur

Geiger, C., Kanzow, C. (1999). *Numerische Verfahren zur Lösung unrestringierter Optimierungsaufgaben*. New York: Springer. DOI: [10.1007/978-3-642-58582-1](https://doi.org/10.1007/978-3-642-58582-1).