Das Newton-Verfahren für Optimierungsaufgaben mit hochdimensionalen Box-Restriktionen

Diplomarbeit

vorgelegt von

Stefan Härtel

aus

Hildesheim

angefertigt im

Institut für Numerische und Angewandte Mathematik

der Georg-August-Universität zu Göttingen

2002

Inhaltsverzeichnis

Kapitel 1

Einleitung

1.1 Problemstellung

Das hier betrachtete *Newton-Verfahren* ist ein spezielles *Trust-Region-Verfahren* für restringierte Optimierungsaufgaben, das auf zweimal stetig differenzierbare Funktionen anwendbar ist. Wir benutzen als Restriktionsmengen die Mengen M und $[l, u]$, die im folgenden definiert werden.

Definition 1.1.1 (Linear und Box-restringierte Menge) Seien $m, n \in \mathbb{N}$ sowie

$$\begin{aligned} \{c_1, \ldots, c_m\} &\subset \mathbb{R}^n, \\ \{l_1, \ldots, l_m\} &\subset \mathbb{R} \cup \{-\infty\} \text{ und} \\ \{u_1, \ldots, u_m\} &\subset \mathbb{R} \cup \{\infty\}. \end{aligned}$$

Die *linear restringierte Menge* M ist durch

$$M := \{x \in \mathbb{R}^n \mid l_i \leq {c_i}^T x \leq u_i \text{ für alle } 1 \leq i \leq m\}$$

und für $m := n$ die *Box-restringierte Menge* $[l, u]$ durch

$$[l, u] := \{x \in \mathbb{R}^n \mid l_i \leq x_i \leq u_i \text{ für alle } 1 \leq i \leq n\}.$$

definiert.

Die Menge M ist ein *Polyeder* und $[l, u]$ kann man auch als *n-dimensionales Intervall* oder als *n-dimensionalen Quader* bezeichnen. Insbesondere ist die Box-restringierte Menge $[l, u]$ eine spezielle Variante von M, die uns hauptsächlich bei der Implementation des Newton-Verfahrens in MATLAB begegnen wird, da sie technisch einfacher zu handhaben ist als die linear restringierte Menge M. Viele Aussagen dieser Arbeit lassen sich auch für allgemeinere Mengen zeigen, das sind die nichtleeren, abgeschlossenen und konvexen Teilmengen des

$\mathbb{R}^n$, die wir durchgängig mit Ω bezeichnen werden. Sei nun $f : \mathbb{R}^n \to \mathbb{R}$ die von uns betrachtete Zielfunktion, die in der Regel mindestens einmal stetig differenzierbar sein sollte. Die zu lösenden Optimierungsaufgaben (P) und (Q) lauten mit Hilfe dieser Definitionen

(P) $$\text{Minimiere } f(x) \quad \text{auf} \quad M$$

sowie

(Q) $$\text{Minimiere } f(x) \quad \text{auf} \quad [l, u].$$

1.2 Motivation

Die Besonderheiten des hier vorgestellten Verfahrens sind geometrischer Natur. Globale und superlineare Konvergenz bei der Lösung der Optimierungsaufgabe (P) wurden mit anderen Verfahren bisher nur unter der Voraussetzung der *strikten Komplementarität* an einem *stationären Punkt* (Definition 2.2.2) $x^* \in M$ erzielt. Das bedeutet, daß in der Darstellung

$$\nabla f(x^*) = \sum_{i=1}^{m} \lambda_i c_i$$

nach dem Satz von Karush-Kuhn-Tucker für die Lagrange-Multiplikatoren $\{\lambda_1, \ldots, \lambda_m\} \subset \mathbb{R}$ nicht nur

$$\lambda_i \begin{cases} \geq 0 & \text{falls } {c_i}^T x^* = l_i, \\ = 0 & \text{falls } l_i < {c_i}^T x^* < u_i, \\ \leq 0 & \text{falls } {c_i}^T x^* = u_i, \end{cases}$$

sondern sogar

$$\lambda_i \begin{cases} > 0 & \text{falls } {c_i}^T x^* = l_i, \\ = 0 & \text{falls } l_i < {c_i}^T x^* < u_i, \\ < 0 & \text{falls } {c_i}^T x^* = u_i \end{cases}$$

gilt. Zusätzlich benötigen diese Verfahren die exakte Lösung von linearen Gleichungssystemen. Wir werden ein Verfahren vorstellen, daß weder auf die strikte Komplementarität an einem Häufungspunkt noch auf das genaue Lösen linearer Gleichungssysteme angewiesen ist. Insbesondere ermöglicht es uns das Lösen hochdimensionaler Optimierungsaufgaben und wir legen nicht einmal speziellen Wert darauf, daß die Vektoren der Ungleichungsrestriktionen linear unabhängig sind. Der überwiegende Teil dieser Arbeit beruht auf den Ergebnissen von Chih-Jen Lin und Jorge J. Moré, *Newton's method for large bound-constrained optimization problems* (Lin and Moré, 1999), wo man die

entsprechenden Literaturangaben dazu findet. In dieser Arbeit werden zwei Verfahren vorgestellt, wobei das zweite eine spezielle Version des ersten ist. Zunächst betrachten wir ein *Trust-Region-Verfahren* (Verfahren 3.1.1) für linear restringierte Optimierungsaufgaben und entwickeln dafür Konvergenzaussagen. Beim *Newton-Verfahren* (Verfahren 5.1.1) werden dann bestimmte Parameter des Trust-Region-Verfahrens fest gewählt, so daß dieses Verfahren aber nur noch auf zweimal stetig differenzierbare Funktionen anwendbar ist. Wir werden insgesamt vier Hauptresultate entwickeln.

(i) Das erste wird durch die Sätze 3.5.4 und 3.5.5 repräsentiert. Es wird das Resultat gezeigt, daß jeder Häufungspunkt der durch das Trust-Region-Verfahren erzeugten Folge $(x_k)_{k\in\mathbb{N}}$ ein stationärer Punkt der Optimierungsaufgabe (P) ist. Dieses Ergebnis beruht im wesentlichen darauf, daß wir die Minderung in der Modellfunktion des Trust-Region-Subproblems mit Hilfe des *Cauchy-Schritts* (Definition 3.3.1) abschätzen können. Der Cauchy-Schritt ist eine „projizierte Version des steilsten Abstiegs“. Das heißt, daß man in Richtung des negativen Gradienten nach einem Punkt sucht, der eine gewisse Minderung in der Modellfunktion garantiert, ohne die Menge M zu verlassen. Das wird dadurch erreicht, daß man den Punkt wieder auf M projiziert.

(ii) Das zweite beinhaltet den geometrischen Aspekt des Trust-Region-Verfahrens und sagt aus, daß letztendlich alle Iterationspunkte des Verfahrens in der durch den negativen Gradienten eines stationären Punktes *exponierten Seitenfläche* (Definition 4.1.1) enthalten sind. Satz 4.1.12 liefert dieses Resultat. Die exponierte Seitenfläche ist eine spezielle Seitenfläche des Polyeders M. Wir benötigen dieses Ergebnis für die Hauptresultate (iii) und (iv), da wir keine strikte Komplementarität voraussetzen.

(iii) Das dritte Resultat gilt speziell für das Newton-Verfahren und beinhaltet die Aussage, daß unter der Annahme der positiven Definitheit der Hesseschen Matrix an einem stationären Punkt $x^* \in M$ auf einem speziellen Untervektorraum des $\mathbb{R}^n$ die gesamte Folge $(x_k)_{k\in\mathbb{N}}$ gegen x^* konvergiert. Satz 5.2.8 liefert die dazugehörigen Details.

(iv) Die vierte Aussage wird durch Satz 5.3.5 ausgedrückt und lautet, daß unter den gleichen Voraussetzungen wie für Hauptresultat (iii) und einer weiteren Bedingung an die Iterationen zur Lösung des Trust-Region-Subproblems die Folge $(x_k)_{k\in\mathbb{N}}$ mindestens linear oder superlinear konvergiert.

Zur Entwicklung dieser Ergebnisse sind aber noch andere wichtige Hilfsmittel notwendig, die auch in den Kapiteln davor behandelt werden.

- In Kapitel 2 werden die theoretischen Grundlagen bereitgestellt, die im weiteren Verlauf benötigt werden. Dabei handelt es sich um *Projektionen* (Definition 2.1.1) und deren fundamentale Eigenschaften sowie um den *projizierten Gradienten* (Definition 2.2.1), der uns die Abbruchbedingung des Trust-Region-Verfahrens liefert. Wir entwickeln die Richtungsableitung der Projektion, die wir später mit dem projizierten Gradienten in Verbindung bringen. Die meisten Sätze dieses Kapitels stammen aus Zarantonello, 1971.

- In Kapitel 3 wird ein Trust-Region-Verfahren für linear restringierte Optimierungsaufgaben vorgestellt und das Konvergenzverhalten untersucht. Wir führen den Cauchy-Schritt ein, der das Trust-Region-Subproblem zwar nicht notwendigerweise löst, aber bei der Konvergenzanalyse unerläßlich ist. Dieses Verfahren dient als Grundlage für das Newton-Verfahren und man findet auch das Hauptresultat (i). Als Quelle dazu dient Burke et al., 1990.

- In Kapitel 4 werden gewisse geometrische Aspekte des Trust-Region-Verfahrens dargestellt. Wir definieren die exponierte Seitenfläche und den *Normalkegel* (Definition 4.1.4), um damit Hauptresultat (ii) zu entwickeln. Wir führen die *Zwischenschritte* (Definition 4.2.1) zur besseren Näherung an die Lösung des Trust-Region-Subproblems ein, um mit dem Newton-Verfahren hochdimensionale Probleme lösen zu können und gleichzeitig lineare oder superlineare Konvergenz zu gewährleisten.

- In Kapitel 5 wird das Newton-Verfahren erläutert und dessen Konvergenzeigenschaften beschrieben. Insbesondere entwickeln wir in diesem Kapitel die Hauptresultate (iii) und (iv). Unter weiteren Annahmen, die eher theoretischer Natur sind, können wir auch quadratische Konvergenz zeigen.

- In Kapitel 6 werden die numerischen Ergebnisse präsentiert, die mit einer Implementation des Newton-Verfahrens in MATLAB erzielt wurden. Wir beschränken uns allerdings auf die Optimierungsaufgabe (Q), betrachten aber verschiedene Funktionen und verändern wichtige Parameter des Verfahrens.

Kapitel 2

Theoretische Grundlagen

In diesem Kapitel werden die theoretischen Grundlagen bereitgestellt, die im weiteren Verlauf benötigt werden. Dabei handelt es sich zum einen um Projektionen und deren fundamentale Eigenschaften. Darauf aufbauend werden wir eine Funktion kennenlernen, die zwar, da sie nicht linear ist, nicht als Ableitung der Projektion betrachtet werden kann, aber eine Näherung der Ordnung $o(\|h\|)$ für Änderungen im Argument der Projektion der Gestalt $h \in \mathbb{R}^n$ liefert. Daraus erhält man die Richtungsableitung der Projektion, die wir später dazu verwenden werden, um die Durchführbarkeit des Trust-Region-Verfahrens zu zeigen. Anschließend führen wir den projizierten Gradienten ein, der uns die Abbruchbedingung des Trust-Region-Verfahrens liefert und den wir später mit der Richtungsableitung der Projektion in Verbindung bringen werden.

2.1 Projektion

Als erstes sollen an dieser Stelle Projektionen auf abgeschlossene und konvexe Mengen eingeführt und deren Eigenschaften beschrieben werden. Sei dazu $\Omega \subseteq \mathbb{R}^n$ eine nichtleere, abgeschlossene und konvexe Menge und $x \in \mathbb{R}^n$ beliebig vorgegeben. Anschaulich gesehen ist die Projektion von x auf Ω der Punkt aus Ω, der den geringsten Abstand zu x besitzt, hier bezüglich der euklidischen Norm. Da das kanonische Skalarprodukt aus dem $\mathbb{R}^n$ insbesondere einen Hilbertraum macht, ist die Projektion wohldefiniert, da zu diesem x genau ein $y_0 \in \Omega$ mit der Eigenschaft

$$\|y_0 - x\| = \inf_{y \in \Omega} \|y - x\|$$

existiert. In der gesamten Arbeit wird aussschließlich die euklidische Norm benutzt, aus diesem Grund lassen wir den Index unten rechts zur Bezeichnung der Norm weg.

Definition 2.1.1 (Projektion) Sei $\Omega \subseteq \mathbb{R}^n$ eine nichtleere, abgeschlossene und konvexe Menge und $x \in \mathbb{R}^n$. Die *Projektion* auf Ω ist die Abbildung $\Pi_\Omega : \mathbb{R}^n \to \Omega$ mit

$$\Pi_\Omega(x) := y_0,$$

wobei $\|y_0 - x\| = \inf_{y \in \Omega} \|y - x\|$ gelte.

Das nächste Lemma liefert eine grundlegende Charakterisierung der Projektion.

Lemma 2.1.2 *Sei $\Omega \subseteq \mathbb{R}^n$ eine nichtleere, abgeschlossene und konvexe Menge. Für $x \in \mathbb{R}^n$ und $y_0 \in \Omega$ gilt*

$$\|y_0 - x\| = \inf_{y \in \Omega} \|y - x\|$$

genau dann, wenn

$$(y_0 - x)^T (y - y_0) \geq 0 \quad \text{für alle } y \in \Omega$$

ist. Durch diese Eigenschaft ist die Projektion auf Ω eindeutig charakterisiert.

Beweis. Siehe den Beweis von Lemma 1.1 in Zarantonello, 1971. ■

Die erste Eigenschaft im nächsten Satz wird oft auch als *Monotonie* der Projektion bezeichnet. Die zweite Eigenschaft sagt aus, daß Projektionen nicht *expandierend* sind. Das bedeutet, daß bei Abbildung von Punkten die Abstände der Bildpunkte durchaus gleich bleiben können, die Abbildung aber sonst kontrahiert. Insbesondere ergibt sich aus der zweiten Aussage die Stetigkeit der Projektion.

Satz 2.1.3 *Sei $\Omega \subseteq \mathbb{R}^n$ eine nichtleere, abgeschlossene und konvexe Menge. Dann gelten für alle $x, y \in \mathbb{R}^n$*

(i) $(\Pi_\Omega(x) - \Pi_\Omega(y))^T (x - y) \geq 0$ *und im Fall $\Pi_\Omega(x) \neq \Pi_\Omega(y)$ gilt sogar die strikte Ungleichung,*

(ii) $\|\Pi_\Omega(x) - \Pi_\Omega(y)\| \leq \|x - y\|$.

Beweis. Siehe den Beweis von Lemma 1.2 in Zarantonello, 1971. ■

Teil (i) des nächsten Satzes stammt aus Toint, 1988, Teil (ii) aus Calamai and Moré, 1987.

Satz 2.1.4 *Sei $\Omega \subseteq \mathbb{R}^n$ eine nichtleere, abgeschlossene und konvexe Menge. Dann gelten für alle $x, d \in \mathbb{R}^n$*

(i) die Funktion $\varphi : \mathbb{R}_+ \cup \{0\} \to \mathbb{R}$, *definiert durch*

$$\varphi(\alpha) := \|\Pi_\Omega(x + \alpha d) - x\|,$$

ist monoton nichtfallend,

(ii) die Funktion $\psi : \mathbb{R}_+ \to \mathbb{R}$, *definiert durch*

$$\psi(\alpha) := \frac{\|\Pi_\Omega(x + \alpha d) - x\|}{\alpha},$$

ist monoton nichtsteigend.

Beweis.

(i) Seien $x, d \in \mathbb{R}^n$ und $\alpha > \beta \geq 0$. Wir werden

$$\|\Pi_\Omega(x + \alpha d) - x\| \geq \|\Pi_\Omega(x + \beta d) - x\|$$

zeigen. Durch einfaches Ausmultiplizieren erhält man

$$\begin{aligned}
&\|\Pi_\Omega(x + \alpha d) - x\|^2 \\
=&\|\Pi_\Omega(x + \beta d) - x\|^2 \\
&+ 2(\Pi_\Omega(x + \beta d) - x)^T(\Pi_\Omega(x + \alpha d) - \Pi_\Omega(x + \beta d)) \\
&+ \|\Pi_\Omega(x + \alpha d) - \Pi_\Omega(x + \beta d)\|^2 \\
\geq&\|\Pi_\Omega(x + \beta d) - x\|^2 \\
&+ 2(\Pi_\Omega(x + \beta d) - (x + \beta d))^T(\Pi_\Omega(x + \alpha d) - \Pi_\Omega(x + \beta d)) \\
&+ 2\beta d^T(\Pi_\Omega(x + \alpha d) - \Pi_\Omega(x + \beta d)).
\end{aligned}$$

Hierbei sieht man, daß die letzten beiden Terme nichtnegativ sind, denn es ist

$$(\Pi_\Omega(x + \beta d) - (x + \beta d))^T(\Pi_\Omega(x + \alpha d) - \Pi_\Omega(x + \beta d)) \geq 0$$

aufgrund von Lemma 2.1.2 und es gilt weiterhin nach Satz 2.1.3, Teil (i)

$$\begin{aligned}
&\beta d^T(\Pi_\Omega(x + \alpha d) - \Pi_\Omega(x + \beta d)) \\
=&\frac{\beta}{\alpha - \beta}((x + \alpha d) - (x + \beta d))^T(\Pi_\Omega(x + \alpha d) - \Pi_\Omega(x + \beta d)) \\
\geq&0.
\end{aligned}$$

Insgesamt ist damit die oben angegebene Ungleichung gezeigt. $\surd$

(ii) Seien nun $x, d \in \mathbb{R}^n$ und $\alpha > \beta > 0$. Wir wollen die Ungleichung

$$\frac{\|\Pi_\Omega(x + \alpha d) - x\|}{\alpha} \leq \frac{\|\Pi_\Omega(x + \beta d) - x\|}{\beta}$$

zeigen. Falls

$$\Pi_\Omega(x + \alpha d) = \Pi_\Omega(x + \beta d)$$

gilt, dann ist sogar die Ungleichung

$$\frac{\|\Pi_\Omega(x + \alpha d) - x\|}{\alpha} < \frac{\|\Pi_\Omega(x + \beta d) - x\|}{\beta}$$

erfüllt und wir sind fertig. Daher betrachten wir jetzt den Fall, daß

$$\Pi_\Omega(x + \alpha d) \neq \Pi_\Omega(x + \beta d)$$

ist. Seien dazu beliebige $u, v \in \mathbb{R}^n$ mit

$$v^T(u - v) > 0$$

vorgegeben. Daraus ergibt sich sofort $\|v\|^2 < v^T u \leq \|v\| \, \|u\|$ und somit $\|v\| < \|u\|$. Weiterhin folgt daraus

$$\begin{aligned}
\|v\| \, u^T(u - v) &= \|v\| \, (u - v)^T(u - v) + \|v\| \, v^T(u - v) \\
&= \|v\| \, \|u - v\|^2 + \|v\| \, v^T(u - v) \\
&\geq \|v\| \, v^T(u - v) \\
&\geq \|u\| \, v^T(u - v)
\end{aligned}$$

und damit die Ungleichung

$$\frac{\|u\|}{\|v\|} \leq \frac{u^T(u - v)}{v^T(u - v)}.$$

Setzt man nun

$$\begin{aligned}
u :=& \Pi_\Omega(x + \alpha d) - x \text{ und} \\
v :=& \Pi_\Omega(x + \beta d) - x,
\end{aligned}$$

so folgt mit Lemma 2.1.2

$$\begin{aligned}
u^T(u - v) =& (\Pi_\Omega(x + \alpha d) - x)^T(\Pi_\Omega(x + \alpha d) - \Pi_\Omega(x + \beta d)) \\
=& (\Pi_\Omega(x + \alpha d) - (x + \alpha d)^T(\Pi_\Omega(x + \alpha d) - \Pi_\Omega(x + \beta d)) \\
&+ \alpha d^T(\Pi_\Omega(x + \alpha d) - \Pi_\Omega(x + \beta d)) \\
\leq& \alpha d^T(\Pi_\Omega(x + \alpha d) - \Pi_\Omega(x + \beta d))
\end{aligned}$$

und

$$\begin{aligned} v^T(u-v) =& (\Pi_\Omega(x+\beta d)-x)^T(\Pi_\Omega(x+\alpha d)-\Pi_\Omega(x+\beta d)) \\ =& (\Pi_\Omega(x+\beta d)-(x+\beta d)^T(\Pi_\Omega(x+\alpha d)-\Pi_\Omega(x+\beta d)) \\ &+ \beta d^T(\Pi_\Omega(x+\alpha d)-\Pi_\Omega(x+\beta d)) \\ \geq& \beta d^T(\Pi_\Omega(x+\alpha d)-\Pi_\Omega(x+\beta d)). \end{aligned}$$

Aus $\Pi_\Omega(x+\alpha d) \neq \Pi_\Omega(x+\beta d)$ ergibt sich jetzt mit Hilfe von Satz 2.1.3, Teil (i)

$$\begin{aligned} & (\alpha-\beta)d^T(\Pi_\Omega(x+\alpha d)-\Pi_\Omega(x+\beta d)) \\ =& ((x+\alpha d)-(x+\beta d))^T(\Pi_\Omega(x+\alpha d)-\Pi_\Omega(x+\beta d)) \\ >& 0 \end{aligned}$$

und zusammen mit $\alpha-\beta>0$ die Ungleichung $v^T(u-v)>0$. Jetzt erhält man das gewünschte Ergebnis

$$\frac{\|\Pi_\Omega(x+\alpha d)-x\|}{\|\Pi_\Omega(x+\beta d)-x\|} = \frac{\|u\|}{\|v\|} \leq \frac{u^T(u-v)}{v^T(u-v)} \leq \frac{\alpha}{\beta}.$$

√

■

Direkt daraus erhält man das folgende Korollar.

Korollar 2.1.5 *Sei $\Omega \subseteq \mathbb{R}^n$ eine nichtleere, abgeschlossene und konvexe Menge. Dann gelten für alle $x, d \in \mathbb{R}^n$ und alle $\alpha, \beta, \gamma > 0$ mit $\alpha \geq \gamma\beta > 0$*

$$\|\Pi_\Omega(x+\alpha d)-x\| \geq \min\{\gamma, 1\}\,\|\Pi_\Omega(x+\beta d)-x\|$$

Beweis. Seien $x, d \in \mathbb{R}^n$ und $\alpha, \beta, \gamma > 0$ mit $\alpha \geq \gamma\beta > 0$ beliebig gewählt. Dann folgt aus Satz 2.1.4

$$\begin{aligned} \|\Pi_\Omega(x+\alpha d)-x\| &\geq \|\Pi_\Omega(x+\min\{\gamma,1\}\beta d)-x\| \\ &= \min\{\gamma,1\}\beta\frac{\|\Pi_\Omega(x+\min\{\gamma,1\}\beta d)-x\|}{\min\{\gamma,1\}\beta} \\ &\geq \min\{\gamma,1\}\beta\frac{\|\Pi_\Omega(x+\beta d)-x\|}{\beta} \\ &= \min\{\gamma,1\}\,\|\Pi_\Omega(x+\beta d)-x\|\,. \end{aligned}$$

■

Obwohl Projektionen im allgemeinen nichtlinear sind, kann man gewissermaßen Vektoren und Skalare aus dem Argument der Projektion herausnehmen, wenn man die gleiche Prozedur mit der Menge vornimmt, auf die projiziert werden soll. Die genaue Formulierung dazu liefert der nächste Satz.

Satz 2.1.6 *Sei $\Omega \subseteq \mathbb{R}^n$ eine nichtleere, abgeschlossene und konvexe Menge. Dann gelten für alle $x \in \mathbb{R}^n$*

(i) $\Pi_{z+\Omega}(z+x) = z + \Pi_\Omega(x)$ *für alle* $z \in \mathbb{R}^n$,

(ii) $\Pi_{t\Omega}(tx) = t\Pi_\Omega(x)$ *für alle* $t > 0$.

Beweis.

(i) Seien $x, z \in \mathbb{R}^n$ und $y_0 := \Pi_{z+\Omega}(z+x) - z$. Dann ist $\Pi_{z+\Omega}(z+x) = z + y_0$ und somit

$$\begin{aligned} \|y_0 - x\| &= \|z + y_0 - (z+x)\| \\ &= \inf_{y\in\Omega} \|z + y - (z+x)\| \\ &= \inf_{y\in\Omega} \|y - x\| . \end{aligned}$$

Daraus ergibt sich $\Pi_\Omega(x) = y_0$ wie gefordert. √

(ii) Seien $x \in \mathbb{R}^n$ und $t > 0$ sowie $y_0 := \frac{1}{t}\Pi_{t\Omega}(tx)$. Dann ist $\Pi_{t\Omega}(tx) = ty_0$ und es gilt

$$\begin{aligned} t\,\|y_0 - x\| &= \|ty_0 - tx\| \\ &= \inf_{y\in\Omega} \|ty - tx\| \\ &= t \inf_{y\in\Omega} \|y - x\| . \end{aligned}$$

Daraus ergibt sich wie eben $\Pi_\Omega(x) = y_0$. √

■

Eine Menge $K \subseteq \mathbb{R}^n$ mit der Eigenschaft, daß $tx \in K$ für alle $t \geq 0$ und $x \in K$ ist, nennen wir *Kegel*. Anschaulich gesprochen ist das eine Menge mit der Eigenschaft, daß jede Halbgerade vom Nullpunkt durch einen beliebigen Punkt dieser Menge ganz in der Menge liegt.

Definition 2.1.7 (Polarkegel) Sei $K \subseteq \mathbb{R}^n$ ein nichtleerer Kegel. Der *Polarkegel* von K ist durch

$$K^\perp := \{x \in \mathbb{R}^n \mid x^T y \leq 0 \text{ für alle } y \in K\}$$

definiert.

Offenbar ist auch für den Polarkegel die Bezeichnung Kegel gerechtfertigt, denn es ist $tx \in K^\perp$ für alle $t \geq 0$ und alle $x \in K^\perp$. Die Aussagen bezüglich Kegeln findet man im zweiten Abschnitt von Zarantonello, 1971.

Lemma 2.1.8 *Sei $K \subseteq \mathbb{R}^n$ ein nichtleerer, abgeschlossener und konvexer Kegel. Dann gilt*

$$(K^\perp)^\perp = K.$$

Beweis.

$\subseteq$: Sei $x \in (K^\perp)^\perp$ beliebig vorgegeben und sei $y := x - \Pi_K(x)$. Wir wollen $x \in K$ zeigen. Da K ein konvexer Kegel ist, gilt $w = z + \Pi_K(x) \in K$ für alle $z \in K$ und daher

$$y^T z = -(\Pi_K(x) - x)^T (w - \Pi_K(x)) \leq 0 \quad \text{für alle } z \in K$$

nach Lemma 2.1.2. Somit folgt $y \in K^\perp$ nach der Definition des Polarkegels. Für das gewählte x ist deshalb $x^T y \leq 0$, wendet man die Definition des Polarkegels auf $K^\perp$ an. Da $0 \in K$ ist, erhalten wir wieder mit Lemma 2.1.2

$$\begin{aligned} 0 \geq & x^T y \\ = & x^T (x - \Pi_K(x)) \\ = & \|x - \Pi_K(x)\|^2 + (\Pi_K(x) - 0)^T (x - \Pi_K(x)) \\ \geq & \|x - \Pi_K(x)\|^2 \\ \geq & 0. \end{aligned}$$

Aus $\|x - \Pi_K(x)\|^2 = 0$ folgt letztendlich $x = \Pi_K(x) \in K$. $\surd$

$\supseteq$: Sei nun $x \in K$ beliebig gewählt. Nach der Definition des Polarkegels haben wir

$$x^T y = y^T x \leq 0 \quad \text{für alle } y \in K^\perp$$

und somit $x \in (K^\perp)^\perp$. $\surd$

■

Lemma 2.1.9 *Sei $K \subseteq \mathbb{R}^n$ ein nichtleerer, abgeschlossener und konvexer Kegel. Dann kann jedes $x \in \mathbb{R}^n$ eindeutig als Summe zweier orthogonaler Vektoren aus K und $K^\perp$ dargestellt werden und es gilt*

$$x = \Pi_K(x) + \Pi_{K^\perp}(x).$$

Beweis. Sei $x \in \mathbb{R}^n$ beliebig vorgegeben. Zuerst zeigen wir die Eindeutigkeit der oben behaupteten Darstellung. Sei dazu

$$x = z + \hat{z}$$

mit $z \in K$ und $\hat{z} \in K^\perp$ sowie $z^T\hat{z} = 0$. Dann gilt

$$\begin{aligned}(z-x)^T(y-z) &= -\hat{z}^T(y-z)\\ &= -\hat{z}^T y\\ &\geq 0 \quad \text{für alle } y \in K\end{aligned}$$

nach der Definition des Polarkegels und aus Lemma 2.1.2 erhalten wir $\Pi_K(x) = z$. Lemma 2.1.8 impliziert $z \in (K^\perp)^\perp$ und daher gilt

$$\begin{aligned}(\hat{z}-x)^T(y-\hat{z}) &= -z^T(y-\hat{z})\\ &= -z^T y\\ &\geq 0 \quad \text{für alle } y \in K^\perp.\end{aligned}$$

Aus Lemma 2.1.2 erhält man wie eben $\Pi_{K^\perp}(x) = \hat{z}$. Jetzt zeigen wir die Existenz der behaupteten Darstellung. Trivialerweise ist

$$x = \Pi_K(x) + (x - \Pi_K(x)).$$

Aufgrund dieser Darstellung muß nur noch $x - \Pi_K(x) \in K^\perp$ sowie die Orthogonalität zwischen $\Pi_K(x)$ und $x - \Pi_K(x)$ gezeigt werden. Da K ein konvexer Kegel ist, ist für $y \in K$ auch $y + \Pi_K(x) \in K$ und aus diesem Grund erhält man aus der Charakterisierung der Projektion in Lemma 2.1.2

$$(\Pi_K(x) - x)^T(y - \Pi_K(x)) \geq 0 \quad \text{für alle } y \in K$$

die Ungleichung

$$(\Pi_K(x) - x)^T y \geq 0 \quad \text{für alle } y \in K.$$

Mit der Definition des Polarkegels erhalten wir nun $x - \Pi_K(x) \in K^\perp$. Setzt man in der ersten der letzten beiden Ungleichungen $y := 0$ und in der zweiten $y := \Pi_K(x)$, so bekommt man

$$\begin{aligned}(x - \Pi_K(x))^T\Pi_K(x) &\geq 0 \text{ und}\\ (x - \Pi_K(x))^T\Pi_K(x) &\leq 0\end{aligned}$$

und damit $(x - \Pi_K(x))^T\Pi_K(x) = 0$, womit auch die Orthogonalität zwischen $\Pi_K(x)$ und $x - \Pi_K(x)$ gezeigt und damit der ganze Beweis beendet ist. ■

Als nächstes führen wir den Begriff des Tangentialkegels ein.

Definition 2.1.10 (Tangentialkegel) Sei $\Omega \subseteq \mathbb{R}^n$ eine nichtleere, abgeschlossene und konvexe Menge und $x \in \Omega$. Der *Tangentialkegel* an Ω in x ist durch

$$T(\Omega; x) := \left\{ p \in \mathbb{R}^n \;\middle|\; \begin{array}{l}\text{Es existieren Folgen } (t_k)_{k\in\mathbb{N}} \subset \mathbb{R}_+ \text{ und } (r_k)_{k\in\mathbb{N}} \subset \mathbb{R}^n \text{ mit}\\ x + t_k p + r_k \in \Omega \text{ für alle } k \in \mathbb{N}, \lim\limits_{k\to\infty} t_k = 0, \lim\limits_{k\to\infty} \dfrac{r_k}{t_k} = 0\end{array} \right\}.$$

definiert.

Der nach dem nächsten Lemma folgende Satz bietet für die Projektion unter Verwendung des Tangentialkegels eine der Ableitung ähnliche Funktion an, die zwar nichtlinear ist, aber für kleine Änderungen $h \in \mathbb{R}^n$ im Argument der Projektion vom Fehler $o(\|h\|)$ ist. In der Literatur wird diese Approximation auch als *konisches Differential* bezeichnet. Die nächsten beiden Aussagen stammen aus dem vierten Abschnitt von Zarantonello, 1971.

Lemma 2.1.11 *Sei $\Omega \subseteq \mathbb{R}^n$ eine nichtleere, abgeschlossene und konvexe Menge und $x \in \Omega$. Dann hat die Funktion $\vartheta : T(\Omega; x) \to \mathbb{R}^n$, definiert durch*

$$\vartheta(h) := \Pi_\Omega(x+h) - x - h,$$

die Eigenschaft

$$\lim_{\substack{h \to 0 \\ h \in T(\Omega;x)}} \frac{\vartheta(h)}{\|h\|} = 0.$$

Beweis. Der Beweis erfolgt in zwei Schritten.

- Als erstes zeigen wir

$$\lim_{0<t\to 0} \frac{\vartheta(th)}{\|th\|} = 0 \quad \text{für alle } h \in T(\Omega; x).$$

Sei dazu $0 \neq h \in T(\Omega; x)$ beliebig vorgegeben. Nach der Definition des Tangentialkegels existieren Folgen $(t_k)_{k\in\mathbb{N}} \subset \mathbb{R}_+$ und $(r_k)_{k\in\mathbb{N}} \subset \mathbb{R}^n$ mit

$$x_k := x + t_k h + r_k \in \Omega \text{ für alle } k \in \mathbb{N}, \quad \lim_{k\to\infty} t_k = 0 \text{ und } \lim_{k\to\infty} \frac{r_k}{t_k} = 0.$$

Sei nun $\varepsilon > 0$ beliebig gewählt. Dann existiert ein $\hat{k} \in \mathbb{N}$, so daß

$$\frac{\|x_k - x - t_k h\|}{t_k} = \frac{\|r_k\|}{t_k} < \varepsilon \|h\| \quad \text{für alle } k \geq \hat{k}$$

gilt. Wähle jetzt ein $0 < t < t_{\hat{k}}$. Dann folgt

$$\begin{aligned}
\frac{\|\Pi_\Omega(x+th) - x - th\|}{t} \leq & \frac{\left\| \frac{t}{t_{\hat{k}}} x_{\hat{k}} + \left(1 - \frac{t}{t_{\hat{k}}}\right) x - x - th \right\|}{t} \\
= & \frac{\left\| \frac{t}{t_{\hat{k}}} x_{\hat{k}} - \frac{t}{t_{\hat{k}}} x - th \right\|}{t} \\
= & \frac{\|x_{\hat{k}} - x - t_{\hat{k}} h\|}{t_{\hat{k}}} \\
< & \varepsilon \|h\|
\end{aligned}$$

und damit

$$\lim_{0<t\to 0} \frac{\vartheta(th)}{\|th\|} = 0.$$

√

- Jetzt zeigen wir, daß

$$\lim_{\substack{h\to 0\\ h\in T(\Omega;x)}} \frac{\vartheta(h)}{\|h\|} = 0 \quad \text{für alle } h \in T(\Omega;x)$$

gilt. Das heißt, daß $\frac{\vartheta(h)}{\|h\|}$ für $h \to 0$ mit $h \in T(\Omega;x)$ gleichmäßig gegen Null konvergiert. Der Beweis dazu wird durch Widerspruch geführt. Angenommen, der Limes ist nicht gleichmäßig. Dann existiert ein $\varepsilon > 0$ und eine Folge $(h_k)_{k\in\mathbb{N}} \subset T(\Omega;x)$ mit $\lim_{k\to\infty} h_k = 0$ und

$$\frac{\|\Pi_\Omega(x+h_k) - x - h_k\|}{\|h_k\|} \geq \varepsilon \quad \text{für alle } k \in \mathbb{N}.$$

Nach einem möglicherweise notwendigen Übergang zu Teilfolgen sei

$$u := \lim_{k\to\infty} \frac{h_k}{\|h_k\|}.$$

Es ist $u \in T(\Omega;x)$, da der Tangentialkegel abgeschlossen ist. Nach Umformungen mit Hilfe von Satz 2.1.6 erhalten wir

$$\begin{aligned}\left\|\Pi_{\|h_k\|^{-1}(\Omega-x)}\left(\frac{h_k}{\|h_k\|}\right) - \frac{h_k}{\|h_k\|}\right\| &\geq \frac{\|\Pi_{\Omega-x}(h_k) - h_k\|}{\|h_k\|}\\ &\geq \varepsilon \quad \text{für alle } k \in \mathbb{N}.\end{aligned}$$

Da die Menge $s(\Omega - x)$ mit wachsendem $s > 0$ immer größer bezüglich der Inklusionsrelation wird, rückt anschaulich gesprochen die Projektion von $\frac{h_k}{\|h_k\|}$ auf $s(\Omega - x)$ mit wachsendem $s > 0$ immer näher an $\frac{h_k}{\|h_k\|}$ selber heran. So ergibt sich jedenfalls

$$\left\|\Pi_{s(\Omega-x)}\left(\frac{h_k}{\|h_k\|}\right) - \frac{h_k}{\|h_k\|}\right\| \geq \varepsilon \quad \text{für alle } s > 0$$

und alle $k \in \mathbb{N}$, für die $\|h_k\|^{-1} > s$ ist. Der Übergang zum Grenzwert für $k \to \infty$ und die erneute Anwendung von Satz 2.1.6 ergeben

$$\left\|\Pi_{\Omega-x}\left(\frac{u}{s}\right)s - \frac{u}{s}s\right\| = \left\|\Pi_{s(\Omega-x)}(u) - u\right\| \geq \varepsilon \quad \text{für alle } s > 0$$

und somit

$$\frac{\left\|\Pi_\Omega(x+\frac{u}{s}) - x - \frac{u}{s}\right\|}{\left\|\frac{u}{s}\right\|} \geq \varepsilon \quad \text{für alle } s > 0.$$

Das aber steht im Widerspruch zu der bereits bewiesenen Aussage, daß

$$\lim_{0<t\to 0} \frac{\vartheta(th)}{\|th\|} = 0 \quad \text{für alle } h \in T(\Omega;x)$$

gilt. $\surd$

■

Satz 2.1.12 *Sei $\Omega \subseteq \mathbb{R}^n$ eine nichtleere, abgeschlossene und konvexe Menge und $x \in \Omega$. Dann hat die Funktion $\vartheta : \mathbb{R}^n \to \mathbb{R}^n$, definiert durch*

$$\vartheta(h) := \Pi_\Omega(x+h) - x - \Pi_{T(\Omega;x)}(h),$$

die Eigenschaft

$$\lim_{h\to 0} \frac{\vartheta(h)}{\|h\|} = 0.$$

Beweis. Sei $h \in \mathbb{R}^n$ beliebig gewählt. Es ist

$$\begin{aligned}
&\|x+h-\Pi_\Omega(x+h)\|^2 \\
=&\left\|x+h-\Pi_{x+T(\Omega;x)}(x+h)\right\|^2 \\
&+2(x+h-\Pi_{x+T(\Omega;x)}(x+h))^T(\Pi_{x+T(\Omega;x)}(x+h)-\Pi_\Omega(x+h)) \\
&+\left\|\Pi_{x+T(\Omega;x)}(x+h)-\Pi_\Omega(x+h)\right\|^2.
\end{aligned}$$

Aus Lemma 2.1.2 folgt

$$(x+h-\Pi_{x+T(\Omega;x)}(x+h))^T(\Pi_{x+T(\Omega;x)}(x+h)-\Pi_\Omega(x+h)) \geq 0,$$

so daß man zusammen mit der ersten Gleichung und der Definition der Projektion

$$\begin{aligned}
\left\|x+h-\Pi_\Omega(\Pi_{x+T(\Omega;x)}(x+h))\right\|^2 \geq& \|x+h-\Pi_\Omega(x+h)\|^2 \\
\geq& \left\|x+h-\Pi_{x+T(\Omega;x)}(x+h)\right\|^2 \\
&+\left\|\Pi_{x+T(\Omega;x)}(x+h)-\Pi_\Omega(x+h)\right\|^2
\end{aligned}$$

erhält. Mit Satz 2.1.6 kann man diese Terme in

$$\begin{aligned}
\left\|x+h-\Pi_\Omega(x+\Pi_{T(\Omega;x)}(h))\right\|^2 \geq& \left\|h-\Pi_{T(\Omega;x)}(h)\right\|^2 \\
&+\left\|x+\Pi_{T(\Omega;x)}(h)-\Pi_\Omega(x+h)\right\|^2
\end{aligned}$$

umformen. Nach Lemma 2.1.9 existiert die Zerlegung

$$h = \Pi_{T(\Omega;x)}(h) + \Pi_{T(\Omega;x)^\perp}(h)$$

und deshalb gilt

$$\begin{aligned}
&\left\|\Pi_\Omega(x+h)-x-\Pi_{T(\Omega;x)}(h)\right\|^2 \\
\leq&\left\|\Pi_{T(\Omega;x)^\perp}(h)+x+\Pi_{T(\Omega;x)}(h)-\Pi_\Omega(x+\Pi_{T(\Omega;x)}(h))\right\|^2-\left\|\Pi_{T(\Omega;x)^\perp}(h)\right\|^2 \\
\leq&(2\Pi_{T(\Omega;x)^\perp}(h)+(x+\Pi_{T(\Omega;x)}(h)-\Pi_\Omega(x+\Pi_{T(\Omega;x)}(h))))^T \\
&(x+\Pi_{T(\Omega;x)}(h)-\Pi_\Omega(x+\Pi_{T(\Omega;x)}(h))).
\end{aligned}$$

Sei nun die Funktion $\xi : T(\Omega; x) \to \mathbb{R}^n$ definiert durch

$$\xi(h) := x + h - \Pi_\Omega(x + h).$$

Jetzt können wir Lemma 2.1.11 anwenden und es ist

$$\lim_{h \to 0} \frac{\xi(h)}{\left\|\Pi_{T(\Omega;x)}(h)\right\|} = 0,$$

da die Projektion stetig und $\lim_{h\to 0} \Pi_{T(\Omega;x)}(h) = 0$ ist. Nach der Ungleichung von Cauchy-Schwarz und der eben definierten Funktion u kann man die letzte Ungleichung zu

$$\begin{aligned} &\left\|\Pi_\Omega(x + h) - x - \Pi_{T(\Omega;x)}(h)\right\|^2 \\ \leq &\left(2 \left\|\Pi_{T(\Omega;x)^\perp}(h)\right\| + \xi(\Pi_{T(\Omega;x)}(h))\right) \xi(\Pi_{T(\Omega;x)}(h)) \end{aligned}$$

umformen. Da $0 \in T(\Omega; x)$ ist, haben wir

$$\left\|\Pi_{T(\Omega;x)}(h) - h\right\| \leq \|0 - h\| = \|h\|$$

und somit die Abschätzung $\left\|\Pi_{T(\Omega;x)}(h)\right\| \leq 2\|h\|$, die selbstverständlich auch für $T(\Omega; x)^\perp$ statt $T(\Omega; x)$ gilt. Man erhält damit

$$\begin{aligned} &\frac{\left(2 \left\|\Pi_{T(\Omega;x)^\perp}(h)\right\| + \xi(\Pi_{T(\Omega;x)}(h))\right) \xi(\Pi_{T(\Omega;x)}(h))}{\|h\|^2} \\ \leq &4 \left(\frac{\left\|\Pi_{T(\Omega;x)^\perp}(h)\right\|}{\|h\|} + \frac{\xi(\Pi_{T(\Omega;x)}(h))}{\left\|\Pi_{T(\Omega;x)}(h)\right\|} \right) \frac{\xi(\Pi_{T(\Omega;x)}(h))}{\left\|\Pi_{T(\Omega;x)}(h)\right\|}. \end{aligned}$$

Zusammengenommen ergibt das die Ungleichungskette

$$\begin{aligned} &\lim_{h \to 0} \frac{\left\|\Pi_\Omega(x + h) - x - \Pi_{T(\Omega;x)}(h)\right\|^2}{\|h\|^2} \\ \leq &\lim_{h \to 0} 4 \left(\frac{\left\|\Pi_{T(\Omega;x)^\perp}(h)\right\|}{\|h\|} + \frac{\xi(\Pi_{T(\Omega;x)}(h))}{\left\|\Pi_{T(\Omega;x)}(h)\right\|} \right) \frac{\xi(\Pi_{T(\Omega;x)}(h))}{\left\|\Pi_{T(\Omega;x)}(h)\right\|} \\ \leq &4 \left(\underbrace{\lim_{h \to 0} \frac{2\|h\|}{\|h\|}}_{=2} + \underbrace{\lim_{h \to 0} \frac{\xi(\Pi_{T(\Omega;x)}(h))}{\left\|\Pi_{T(\Omega;x)}(h)\right\|}}_{=0} \right) \underbrace{\lim_{h \to 0} \frac{\xi(\Pi_{T(\Omega;x)}(h))}{\left\|\Pi_{T(\Omega;x)}(h)\right\|}}_{=0} \\ = &0 \end{aligned}$$

und damit das zu zeigende Ergebnis

$$\lim_{h\to 0} \frac{\left\|\Pi_\Omega(x+h) - x - \Pi_{T(\Omega;x)}(h)\right\|}{\|h\|} = 0.$$

■

Aus dem gerade bewiesenen Satz gewinnt man die Richtungsableitung der Projektion, die wir später zum Beweis der Durchführbarkeit des Trust-Region-Verfahrens benötigen. Die Details dazu liefert das folgende Korollar.

Korollar 2.1.13 *Sei $\Omega \subseteq \mathbb{R}^n$ eine nichtleere, abgeschlossene und konvexe Menge und $x \in \Omega$. Dann gilt*

$$\lim_{0<t\to 0} \frac{\Pi_\Omega(x+th) - x}{t} = \Pi_{T(\Omega;x)}(h) \quad \textit{für alle } h \in \mathbb{R}^n.$$

Beweis. Sei $h \in \mathbb{R}^n$ beliebig gewählt. Da $T(\Omega;x) = tT(\Omega;x)$ für alle $t > 0$ ist, bekommt man mit den Umformungen aus Satz 2.1.6 und Satz 2.1.12

$$\begin{aligned}
0 &= \lim_{t\to 0} \frac{\vartheta(th)}{\|th\|} \\
&= \lim_{t\to 0} \frac{\Pi_\Omega(x+th) - x - \Pi_{T(\Omega;x)}(th)}{\|th\|} \\
&= \lim_{0<t\to 0} \frac{\Pi_\Omega(x+th) - x - \Pi_{tT(\Omega;x)}(th)}{t\,\|h\|} \\
&= \lim_{0<t\to 0} \frac{\Pi_\Omega(x+th) - x - t\Pi_{T(\Omega;x)}(h)}{t\,\|h\|}
\end{aligned}$$

und somit auch

$$\lim_{0<t\to 0} \frac{\Pi_\Omega(x+th) - x}{t} - \Pi_{T(\Omega;x)}(h) = 0.$$

■

2.2 Projizierter Gradient

In diesem Abschnitt wird der projizierte Gradient definiert und seine Eigenschaften beschrieben.

Definition 2.2.1 (Projizierter Gradient) Sei $\Omega \subseteq \mathbb{R}^n$ eine nichtleere, abgeschlossene und konvexe Menge und $f : \mathbb{R}^n \to \mathbb{R}$ eine auf einer offenen

Obermenge von Ω differenzierbare Abbildung. Der *projizierte Gradient* von f an der Stelle $x \in \Omega$ ist durch

$$\nabla_\Omega f(x) := \Pi_{T(\Omega;x)}(-\nabla f(x))$$

definiert.

Aufgrund der Konvexität der Menge Ω ist der Tangentialkegel an Ω in x auch gleichzeitig der Abschluß der zulässigen Richtungen an Ω in x und ebenfalls konvex. Damit ist die Projektion auf $T(\Omega; x)$ und somit $\nabla_\Omega f(x)$ wohldefiniert. Der projizierte Gradient spielt bei dem im nächsten Kapitel beschriebenen Trust-Region-Verfahren eine entscheidende Rolle. Dazu müssen wir allerdings den Begriff des stationären Punktes einführen.

Definition 2.2.2 (Stationärer Punkt) Sei $\Omega \subseteq \mathbb{R}^n$ eine nichtleere, abgeschlossene und konvexe Menge und $f : \mathbb{R}^n \to \mathbb{R}$ eine auf einer offenen Obermenge von Ω differenzierbare Abbildung. Dann ist $x \in \Omega$ ein *stationärer Punkt* der Aufgabe, f auf Ω zu minimieren, wenn

$$\nabla f(x)^T (y - x) \geq 0 \quad \text{für alle } y \in \Omega$$

gilt.

So wird durch den nächsten Satz ein einfaches Abbruchkriterium für das Verfahren bereitgestellt, das auf stationäre Punkte testet. Nach diesem Satz liegt ein stationärer Punkt vor, wenn der projizierte Gradient der Zielfunktion an dem Iterationspunkt der Nullvektor ist. Die nächsten beiden Aussagen findet man in Calamai and Moré, 1987.

Satz 2.2.3 *Sei $\Omega \subseteq \mathbb{R}^n$ eine nichtleere, abgeschlossene und konvexe Menge und $f : \mathbb{R}^n \to \mathbb{R}$ eine auf einer offenen Obermenge von Ω differenzierbare Abbildung. Dann gelten für alle $x \in \Omega$*

(i) $\nabla f(x)^T \nabla_\Omega f(x) = -\|\nabla_\Omega f(x)\|^2$,

(ii) $\min\{\nabla f(x)^T p \mid p \in T(\Omega; x) \text{ und } \|p\| \leq 1\} = -\|\nabla_\Omega f(x)\|$,

(iii) x ist genau dann ein stationärer Punkt der Aufgabe, f auf Ω zu minimieren, wenn $\nabla_\Omega f(x) = 0$ ist.

Beweis.

(i) Sei $x \in \Omega$ beliebig vorgegeben. Da $T(\Omega; x)$ ein Kegel ist, gilt

$$t\nabla_\Omega f(x) \in T(\Omega; x) \quad \text{für alle } t \geq 0.$$

Aus Lemma 2.1.2 erhalten wir

$$(\nabla_\Omega f(x) + \nabla f(x))^T(\lambda\nabla_\Omega f(x) - \nabla_\Omega f(x)) \geq 0 \quad \text{für alle } t \geq 0.$$

Setzt man in diese Ungleichung $\lambda := 0$ und $\lambda := 2$ ein, so ergeben sich

$$\begin{aligned}(\nabla_\Omega f(x) + \nabla f(x))^T\nabla_\Omega f(x) \leq& 0 \text{ und}\\ (\nabla_\Omega f(x) + \nabla f(x))^T\nabla_\Omega f(x) \geq& 0\end{aligned}$$

und damit das gewünschte Ergebnis

$$\|\nabla_\Omega f(x)\|^2 = -\nabla f(x)^T\nabla_\Omega f(x).$$

$\surd$

(ii) Sei $x \in \Omega$ beliebig gewählt. Wir betrachten zwei Fälle.

- Wir betrachten zunächst die Möglichkeit, daß $\|\nabla_\Omega f(x)\| = 0$ ist. Sei dazu ein $p \in T(\Omega; x)$ mit $\|p\| \leq 1$ vorgegeben. Dann erhält man aus der Definition der Projektion

$$\begin{aligned}\|\nabla f(x)\|^2 =& \|\nabla_\Omega f(x) + \nabla f(x)\|^2\\ \leq& \|tp + \nabla f(x)\|^2 \quad \text{für alle } 0 < t < 1,\end{aligned}$$

  woraus man nach Ausmultiplizieren des letzten Termes die Ungleichung

$$\nabla f(x)^T(tp) + \|tp\|^2 \geq 0 \quad \text{für alle } 0 < t < 1$$

  bekommt. Division durch t und Übergang zum Limes für $t \to 0$ implizieren das gewünschte Resultat. $\surd$

- Sei nun $\|\nabla_\Omega f(x)\| > 0$ und es sei $q \in T(\Omega; x)$ mit $\|q\| \leq \|\nabla_\Omega f(x)\|$ beliebig gewählt. Aus der Definition der Projektion und nach Quadrieren und Ausmultiplizieren erhält man

$$\begin{aligned}&\|\nabla_\Omega f(x)\|^2 + 2\nabla f(x)^T\nabla_\Omega f(x) + \|\nabla f(x)\|^2\\ =& \|\nabla_\Omega f(x) + \nabla f(x)\|^2\\ \leq& \|q + \nabla f(x)\|^2\\ \leq& \|\nabla_\Omega f(x)\|^2 + 2\nabla f(x)^T q + \|\nabla f(x)\|^2\end{aligned}$$

  und damit

$$\nabla f(x)^T\nabla_\Omega f(x) \leq \nabla f(x)^T q.$$

  Wähle nun ein beliebiges $p \in T(\Omega; x)$ mit $\|p\| \leq 1$. Dann ist $\|\|\nabla_\Omega f(x)\| p\| \leq \|\nabla_\Omega f(x)\|$. Aus diesem Grund können wir jetzt

$q := \|\nabla_\Omega f(x)\|\, p$ setzen und erhalten zusammen mit Teil (i) dieses Satzes das Resultat

$$\begin{aligned} -\|\nabla_\Omega f(x)\|^2 &= \nabla f(x)^T \nabla_\Omega f(x) \\ &\leq \|\nabla_\Omega f(x)\|\, \nabla f(x)^T p. \end{aligned}$$

$\surd$

(iii) $\Rightarrow$: Sei $x \in \Omega$ ein stationärer Punkt der Aufgabe, f auf Ω zu minimieren, das heißt es sei

$$\nabla f(x)^T (y - x) \geq 0 \quad \text{für alle } y \in \Omega.$$

Sei $p \in \mathbb{R}^n$ eine zulässige Richtung an Ω in x. Dann existiert ein $t > 0$, so daß $x + tp \in \Omega$ ist. Setzt man $y := x + tp$, so erhält man $\nabla f(x)^T p \geq 0$. Da $T(\Omega; x)$ der Abschluß der zulässigen Richtungen an Ω in x ist, folgt daraus

$$\nabla f(x)^T p \geq 0 \quad \text{für alle } p \in T(\Omega; x).$$

Mit Teil (ii) dieses Satzes erhalten wir $\|\nabla_\Omega f(x)\| = 0$. $\surd$

$\Leftarrow$: Sei $x \in \Omega$ und $\nabla_\Omega f(x) = 0$. Dann folgt mit Teil (ii), daß

$$\nabla f(x)^T p \geq 0 \quad \text{für alle } p \in T(\Omega; x)$$

ist. Da $y - x \in T(\Omega; x)$ für alle $y \in \Omega$ gilt, bekommt man

$$\nabla f(x)^T (y - x) \geq 0 \quad \text{für alle } y \in \Omega.$$

Das heißt aber, daß x ein stationärer Punkt der Aufgabe, f auf Ω zu minimieren, ist. $\surd$

■

Weiterhin ist es wichtig, gewisse Stetigkeitseigenschaften der Abbildung $\|\nabla_\Omega f(\cdot)\| : \Omega \to \mathbb{R}$, definiert durch $x \mapsto \|\nabla_\Omega f(x)\|$, zu ermitteln, da bei den Konvergenzsätzen bestimmte Teilfolgen der Folge der Iterationspunkte konvergieren. Da dann häufig auch die Folge der projizierten Gradienten oder eine ähnliche Folge konvergiert, kann man mit dem nächsten Lemma Aussagen darüber machen, ob man in irgendeiner Weise gegen einen stationären Punkt konvergiert. Dafür ist es aber hier das erstemal notwendig, daß die Funktion f nicht nur differenzierbar, sondern sogar stetig differenzierbar auf einer offenen Obermenge von Ω ist. Diese Voraussetzung wird daher auch im ganzen nächsten Kapitel beibehalten, da die wichtigen Konvergenzsätze dieses Resultat benötigen. Vorher allerdings rekapitulieren wir eine einfache Tatsache aus

der Analysis.

Eine Funktion $h : \mathbb{R}^n \to \mathbb{R}$ heißt *nach unten halbstetig in* $x \in \Omega$, wenn es zu jedem $\varepsilon > 0$ ein $\delta > 0$ gibt, so daß für alle $y \in \Omega$ aus $\|y - x\| < \delta$ folgt, daß $h(y) > h(x) - \varepsilon$ ist. Dementsprechend heißt die Funktion h *nach unten halbstetig auf* Ω, falls sie in jedem Punkt $x \in \Omega$ nach unten halbstetig ist.

Bemerkung 2.2.4 Gilt für alle Folgen $(x_k)_{k\in\mathbb{N}} \subset \Omega$ mit Grenzwert $x \in \Omega$ die Ungleichung

$$h(x) \leq \liminf_{k\to\infty} h(x_k),$$

so ist das eine äquivalente Bedingung dafür, daß h in x nach unten halbstetig ist. Für den Beweis, daß diese Bedingung hinreichend ist, nehmen wir außer dieser Bedingung an, es existiert ein $\varepsilon > 0$, so daß für jedes $\delta > 0$ ein $y \in \Omega$ existiert, für daß zwar $\|y - x\| < \delta$, aber $h(y) \leq h(x) - \varepsilon$ gilt. Wähle dann zu jedem $k \in \mathbb{N}$ und $\delta := \frac{1}{k+1}$ ein entsprechendes $x_k \in \Omega$ mit $h(x_k) \leq h(x) - \varepsilon$. Damit folgt $\lim_{k\to\infty} x_k = x$, aber $\inf_{k\geq i} h(x_k) \leq h(x) - \varepsilon$ und schließlich

$$\liminf_{i\to\infty} \inf_{k\geq i} h(x_k) \leq h(x) - \varepsilon < h(x).$$

Das ist aber ein Widerspruch zur Voraussetzung. Für den Beweis der Notwendigkeit habe die Folge $(x_k)_{k\in\mathbb{N}} \subset \Omega$ den Limes $x \in \Omega$ und es sei h in x nach unten halbstetig. Dann existiert zu jedem $j \in \mathbb{N}$ und $\varepsilon := \frac{1}{j+1}$ ein $k_j \in \mathbb{N}$, so daß $h(x_k) > h(x) - \frac{1}{j+1}$ für alle $k \geq k_j$ gilt. Damit ist aber auch $\inf_{k\geq k_j} h(x_k) \geq h(x) - \frac{1}{j+1}$, also

$$\liminf_{i\to\infty} \inf_{k\geq i} h(x_k) = \lim_{j\to\infty} \inf_{k\geq k_j} h(x_k) \geq \lim_{j\to\infty} (h(x) - \frac{1}{j+1}) = h(x).$$

Nun aber kommt das angekündigte Lemma.

Lemma 2.2.5 *Sei* $\Omega \subseteq \mathbb{R}^n$ *eine nichtleere, abgeschlossene und konvexe Menge und* $f : \mathbb{R}^n \to \mathbb{R}$ *eine auf einer offenen Obermenge von* Ω *stetig differenzierbare Abbildung. Dann ist die Abbildung* $\|\nabla_\Omega f(\cdot)\| : \Omega \to \mathbb{R}$*, definiert durch* $x \mapsto \|\nabla_\Omega f(x)\|$*, auf* Ω *nach unten halbstetig.*

Beweis. Sei $(x_k)_{k\in\mathbb{N}} \subset \Omega$ eine beliebige Folge mit Grenzwert $x \in \Omega$. Nach Bemerkung 2.2.4 reicht es,

$$\|\nabla_\Omega f(x)\| \leq \liminf_{k\to\infty} \|\nabla_\Omega f(x_k)\|$$

zu zeigen. Satz 2.2.3, Teil (ii) impliziert

$$\nabla f(x_k)^T \left(\frac{y - x_k}{\|y - x_k\|} \right) \geq - \|\nabla_\Omega f(x_k)\| \qquad \text{für alle } y \in \Omega$$

und so

$$\nabla f(x_k)^T(x_k - y) \leq \|\nabla_\Omega f(x_k)\| \, \|y - x_k\| \qquad \text{für alle } y \in \Omega.$$

Da f stetig differenzierbar ist, bekommen wir nach Übergang zum Limes inferior

$$\begin{aligned}
\nabla f(x)^T(x - y) &\leq \liminf_{i\to\infty} \inf_{k\geq i} (\|\nabla_\Omega f(x_k)\| \, \|y - x_k\|) \\
&\leq \liminf_{i\to\infty} (\inf_{k\geq i} \|\nabla_\Omega f(x_k)\| \sup_{k\geq i} \|y - x_k\|) \\
&= \liminf_{i\to\infty} \inf_{k\geq i} \|\nabla_\Omega f(x_k)\| \limsup_{i\to\infty} \sup_{k\geq i} \|y - x_k\| \\
&= \liminf_{i\to\infty} \inf_{k\geq i} \|\nabla_\Omega f(x_k)\| \, \|y - x\| \qquad \text{für alle } y \in \Omega.
\end{aligned}$$

Sei $p \in \mathbb{R}^n$ eine zulässige Richtung an Ω in x. Dann existiert ein $t > 0$, so daß $x + tp \in \Omega$ ist. Setzt man $y := x + tp$, so erhält man aus der obigen Gleichung

$$-\nabla f(x)^T p \leq \liminf_{k\to\infty} \|\nabla_\Omega f(x_k)\| \, \|p\|$$

und da $T(\Omega; x)$ aufgrund der Konvexität von Ω auch der Abschluß der zulässigen Richtungen an Ω in x ist, gilt diese Ungleichung auch für alle $p \in T(\Omega; x)$. Aus Satz 2.2.3, Teil (ii) ergibt sich dann

$$\|\nabla_\Omega f(x)\| \leq \liminf_{k\to\infty} \|\nabla_\Omega f(x_k)\| \, .$$

■

Kapitel 3

Trust-Region-Verfahren

In diesem Kapitel wird ein Trust-Region-Verfahren für linear restringierte Optimierungsaufgaben vorgestellt und das Konvergenzverhalten untersucht. Das anschließend in Kapitel 5 untersuchte Newton-Verfahren ist das Trust-Region-Verfahren, in dem aber die symmetrische Matrix in der Modellfunktion nicht mehr frei gewählt werden darf, sondern diese die Hessesche Matrix am jeweiligen Iterationspunkt des Verfahrens ist, liefert dazu noch stärkere Konvergenzresultate. Das Trust-Region-Verfahren wird anhand von linearen Restriktionen, das heißt anhand der Optimierungsaufgabe (P) erläutert, da sich die Resultate nahtlos vom Box-restringierten Fall übertragen. Für eine Implementation eignen sich allerdings Box-Restriktionen, das heißt die Restriktionen der Optimierungsaufgabe (Q), besser, da diese technisch einfacher zu handhaben sind.

3.1 Verfahren

Als erstes werden wir die Idee beschreiben, die hinter der Klasse der Trust-Region-Verfahren steht, da sicherlich nicht jeder Leser damit vertraut ist. Der erste Schritt besteht im wesentlichen darin, die Zielfunktion $f : \mathbb{R}^n \to \mathbb{R}$ auf einer Kugel um den aktuellen Iterationspunkt $x_k \in M$ mit $k \in \mathbb{N}$ zu approximieren. Der Radius dieser Kugel wird mit Δ_k, wobei dieser teilweise mit einer Konstante multipliziert wird, die approximierende Funktion mit $f_k : \mathbb{R}^n \to \mathbb{R}$ bezeichnet. Diese Modellfunktion kann von jeder beliebigen Gestalt sein, in diesem Verfahren wird allerdings eine allgemeine quadratische Approximation gewählt, das heißt, daß der rein quadratische Anteil durch eine beliebig gewählte symmetrische Matrix $B_k \in \mathbb{R}^{n\times n}$ bestimmt wird. Im Newton-Verfahren ersetzt man dann für zweimal stetig differenzierbare Funktionen die Matrix B_k durch die Hessesche Matrix $\nabla^2 f(x_k)$ am aktuellen Iterationspunkt x_k . Bei anderen Trust-Region-Verfahren wird dann die Modellfunktion auf der gewählten Kugel minimiert, bei diesem Verfahren wird

der Cauchy-Schritt p_k^C berechnet, der eine gewisse Minderung in der Modellfunktion garantiert und der stets eine Abstiegsrichtung darstellt, sollte man sich nicht an einem stationären Punkt befinden. Der Cauchy-Schritt wird in Abschnitt 3.3 noch ausführlich erläutert. Bei stationären Punkten bricht das Verfahren aus diesem Grund ab und der Test darauf ist in der Realität bei der Optimierungsaufgabe (Q) signifikant einfacher als bei Optimierungsaufgabe (P), daher wird die Abbruchbedingung in Abschnitt 3.2 auch nur für den spezielleren Fall behandelt. Wahlweise kann auch noch ein weiterer Schritt p_k berechnet werden, der aber bezüglich der Minderung in der Modellfunktion nicht wesentlich schlechter als der Cauchy-Schritt sein darf. Als nächstes wird die Minderung in der Zielfunktion f mit der in der Modellfunktion f_k verglichen. Abhängig vom Ergebnis dieses Tests, ausgedrückt durch den Quotienten χ_k, wird also der Radius Δ_k verkleinert oder vergrößert. Verkleinert wird der Radius, wenn die tatsächliche Minderung in der Zielfunktion nicht annähernd dem entspricht, was man sich durch die Ergebnisse bei der Modellfunktion erhofft hat, denn offenbar hat die Modellfunktion die Zielfunktion auf der aktuellen Kugel nicht genug approximiert. Er wird dagegen vergrößert, wenn das Ergebnis den Erwartungen entspricht. In einem dazwischenliegenden Intervall von Testergebnissen darf der neue Radius entweder verkleinert oder vergrößert werden. Auch abhängig davon wird der Iterationspunkt x_k um den Schritt p_k bewegt, wenn das Ergebnis des Tests zufriedenstellend ist und andernfalls nicht bewegt. Anschließend wird auch eine neue symmetrische Matrix für das quadratische Modell gewählt. In der Menge $\mathbb{N}$ schließen wir der Einfachheit halber die Null mit ein, es erleichtert die Notation. Es sei nun selbstvestädlich $n \in \mathbb{N} \setminus \{0\}$.

Verfahren 3.1.1 (Trust-Region-Verfahren) Sei die Funktion
$f : \mathbb{R}^n \to \mathbb{R}$ eine auf einer offenen Obermenge von M stetig differenzierbare Abbildung. Dann lautet das *Trust-Region-Verfahren*

- Gegeben seien die Konstanten $0 < \mu_0 < \frac{1}{2}$, $\mu_1 > 0$, $0 < \rho_0 < \rho_1 < \rho_2 < 1$ und $0 < \sigma_1 < \sigma_2 < 1 < \sigma_3$.

- Sei $x_0 \in M$ gegeben. Berechne $\nabla f(x_0)$, bestimme eine symmetrische Matrix $B_0 \in \mathbb{R}^{n \times n}$ und ein $\Delta_0 > 0$.

- Für $k \in \mathbb{N}$

  - Ist x_k ein stationärer Punkt der Optimierungsaufgabe (P), dann breche das Verfahren hier ab. Siehe hierzu Abschnitt 3.2.

  - Andernfalls

* Berechne den Cauchy-Schritt p_k^C zur Optimierungsaufgabe

$$\text{Minimiere } f_k(p) := f(x_k) + \nabla f(x_k)^T p + \frac{1}{2} p^T B_k p$$
$$\text{auf } \{p \in \mathbb{R}^n \mid \|p\| < \Delta_k\}.$$

Der Cauchy-Schritt ist allerdings nicht notwendigerweise eine Lösung dieser Optimierungsaufgabe, siehe hierzu Abschnitt 3.3.

* Berechne einen Schritt p_k, der

$$f_k(p_k) - f(x_k) \leq \mu_0 (f_k(p_k^C) - f(x_k)) \tag{3.1.1}$$

erfüllt, wobei $x_k + p_k \in M$ und $\|p_k\| \leq \mu_1 \Delta_k$ sei.

* Berechne

$$\chi_k := \frac{f(x_k) - f(x_k + p_k)}{f(x_k) - f_k(p_k)}$$

und setze

$$x_{k+1} := \begin{cases} x_k & \text{falls } \chi_k \leq \rho_0, \\ x_k + p_k & \text{falls } \chi_k > \rho_0. \end{cases} \tag{3.1.2}$$

Ein Iterationsindex k, bei dem $\chi_k > \rho_0$ erfüllt ist, wird *erfolgreich* genannt. Setze

$$\Delta_{k+1} \begin{cases} \in [\sigma_1 \min\{\|p_k\|, \Delta_k\}, \sigma_2 \Delta_k] & \text{falls } \chi_k \leq \rho_1, \\ \in [\sigma_1 \Delta_k, \sigma_3 \Delta_k] & \text{falls } \rho_1 < \chi_k < \rho_2, \\ \in [\Delta_k, \sigma_3 \Delta_k] & \text{falls } \chi_k \geq \rho_2. \end{cases}$$

Wähle außerdem eine symmetrische Matrix $B_{k+1} \in \mathbb{R}^{n \times n}$.

3.2 Abbruchbedingung

Nun werden wir die Abbruchbedingung des Verfahrens erläutern, da nicht auf den ersten Blick ersichtlich ist, wie man feststellt, ob man sich an einem stationären Punkt befindet. Nach Satz 2.2.3, Teil (iii) ist $x_k \in M$ für ein $k \in \mathbb{N}$ genau dann ein stationärer Punkt der Optimierungsaufgabe (P), wenn $\nabla_M f(x_k) = 0$ ist. Bei der Optimierungsaufgabe (Q) kann man offenbar $\nabla_{[l,u]} f(x_k)$ leicht bestimmen. Für den Tangentialkegel eines beliebigen Punktes $x \in [l, u]$ gilt nämlich

$$T([l,u]; x) = \left\{ \sum_{i=1}^n \lambda_i e_i \;\middle|\; \lambda_i \begin{cases} \geq 0 & \text{falls } x_i = l_i, \\ \in \mathbb{R} & \text{falls } l_i < x_i < u_i, \\ \leq 0 & \text{falls } x_i = u_i \end{cases} \right\},$$

wobei e_i der i-te Einheitsvektor im $\mathbb{R}^n$ sei. Damit ergibt sich für den projizierten Gradienten $\nabla_{[l,u]} f(x) = \Pi_{T([l,u];x)}(-\nabla f(x))$ als i-te Komponente

$$[\nabla_{[l,u]} f(x)]_i = \begin{cases} \max\{-\partial_i f(x), 0\} & \text{falls } x_i = l_i, \\ -\partial_i f(x) & \text{falls } l_i < x_i < u_i, \\ \min\{-\partial_i f(x), 0\} & \text{falls } x_i = u_i, \end{cases}$$

wobei $\partial_i f(x) := [\nabla f(x)]_i$ sei, das heißt die partielle Ableitung von f nach x_i bezeichne. Man erhält $\nabla_{[l,u]} f(x) = 0$ genau dann, wenn für die i-te Komponente des Gradienten $\nabla f(x)$

$$\partial_i f(x) \begin{cases} \geq 0 & \text{falls } x_i = l_i, \\ = 0 & \text{falls } l_i < x_i < u_i, \\ \leq 0 & \text{falls } x_i = u_i \end{cases}$$

gilt. Die Abbruchbedingung des Trust-Region-Verfahrens läßt sich also im Fall der Optimierungsaufgabe (Q) leicht verifizieren.

3.3 Cauchy-Schritt

Für die Konvergenzanalyse des Trust-Region-Verfahrens spielt der Cauchy-Schritt p_k^C eine zentrale Rolle, deshalb soll er in diesem Abschnitt ausführlich erläutert werden. Sollte ein weiterer Schritt p_k berechnet werden, so ist es deshalb wichtig, daß dieser bezüglich der Minderung in der Modellfunktion nicht wesentlich schlechter als der Cauchy-Schritt ist. Bei der Berechnung des Cauchy-Schritts wird im Prinzip vom aktuellen Iterationspunkt $x_k \in M$ mit $k \in \mathbb{N}$ in Richtung des negativen Gradienten $-\nabla f(x_k)$ gegangen, wobei dieser durch einen Parameter $\alpha > 0$ skaliert ist. Da nicht von vornherein klar ist, daß sich dieser dann Cauchy-Punkt genannte Punkt in der Menge M befindet, wird der Punkt auf M projiziert. Durch Subtrahieren von x_k erhält man so den Cauchy-Schritt. Der Cauchy-Schritt ist damit eine „projizierte Version des steilsten Abstiegs“. Der Parameter α wird dann so gewählt, daß die Minderung in der Modellfunktion mit der Minderung, die durch den Gradienten erwartet wird, vergleichbar bleibt, was durch Forderung (3.3.3) ausgedrückt wird. Forderung (3.3.5) stellt sicher, daß der Schritt nicht zu klein gewählt wird. Der gewählte Parameter α trägt dann die Bezeichnung α_k. Ein Zahlenbeispiel findet man in Burke et al., 1990.

Definition 3.3.1 (Cauchy-Schritt) Sei die Funktion $f : \mathbb{R}^n \to \mathbb{R}$ eine auf einer offenen Obermenge von M stetig differenzierbare Abbildung und $x_k \in M$ mit $k \in \mathbb{N}$ ein Iterationspunkt des Trust-Region-Verfahrens. Die Funktion $\pi_k : \mathbb{R}_+ \cup \{0\} \to \mathbb{R}^n$ sei durch

$$\pi_k(\alpha) := \Pi_M(x_k - \alpha \nabla f(x_k)) - x_k$$

definiert. Seien weiterhin die Konstanten $\gamma_1 > 0$, $0 < \gamma_2 < 1$ und $\gamma_3 > 0$ gegeben. Wähle ein $\alpha_k > 0$, so daß für die Konstanten μ_0 und μ_1 aus dem Trust-Region-Verfahren

$$f_k(\pi_k(\alpha_k)) - f(x_k) \leq \mu_0 \nabla f(x_k)^T \pi_k(\alpha_k) \quad \text{und} \quad \|\pi_k(\alpha_k)\| \leq \mu_1 \Delta_k \tag{3.3.3}$$

mit

$$\alpha_k \in [\gamma_1, \gamma_3] \quad \text{oder} \quad \alpha_k \in [\gamma_2 \hat{\alpha}_k, \gamma_3] \tag{3.3.4}$$

gilt, wobei für $\hat{\alpha}_k$

$$f_k(\pi_k(\hat{\alpha}_k)) - f(x_k) > \mu_0 \nabla f(x_k)^T \pi_k(\hat{\alpha}_k) \quad \text{oder} \quad \|\pi_k(\hat{\alpha}_k)\| > \mu_1 \Delta_k \tag{3.3.5}$$

gelte.

$$p_k^C := \pi_k(\alpha_k)$$

heißt dann ein *Cauchy-Schritt* zum Iterationspunkt x_k. Der *Cauchy-Punkt* ist durch

$$x_k^C := x_k + p_k^C$$

definiert.

Als nächstes zeigen wir die Existenz des Cauchy-Schritts, da diese nicht unmittelbar einzusehen ist.

Satz 3.3.2 *Sei $f : \mathbb{R}^n \to \mathbb{R}$ eine auf einer offenen Obermenge von M stetig differenzierbare Abbildung und $x_k \in M$ mit $k \in \mathbb{N}$ ein Iterationspunkt des Trust-Region-Verfahrens und kein stationärer Punkt der Optimierungsaufgabe (P). Dann existiert ein $\alpha_k > 0$, so daß die Forderungen (3.3.3) bis (3.3.5) erfüllt sind. α_k kann durch eine endliche Anzahl an Auswertungen der Funktion π_k gewonnen werden.*

Beweis. Zuerst zeigen wir, daß die Forderung (3.3.3) an den Cauchy-Schritt für alle hinreichend kleinen $\alpha > 0$ gilt. Vorher bemerken wir, daß aus Korollar 2.1.13

$$\begin{aligned} \lim_{0<\alpha\to 0} \frac{\pi_k(\alpha)}{\alpha} &= \lim_{0<\alpha\to 0} \frac{\Pi_M(x_k - \alpha \nabla f(x_k)) - x_k}{\alpha} \\ &= \Pi_{T(M;x_k)}(-\nabla f(x_k)) \\ &= \nabla_M f(x_k) \end{aligned}$$

folgt. Mit Satz 2.2.3, Teil (i) bekommt man

$$\begin{aligned}
&\lim_{0<\alpha\to 0} \frac{f_k(\pi_k(\alpha)) - f_k(0)}{\alpha} \\
=&\lim_{0<\alpha\to 0} \left(\nabla f(x_k)^T \left(\frac{\pi_k(\alpha)}{\alpha} \right) + \frac{1}{2} \left(\frac{\pi_k(\alpha)}{\alpha} \right)^T B_k \pi_k(\alpha) \right) \\
=&\nabla f(x_k)^T \left(\lim_{0<\alpha\to 0} \frac{\pi_k(\alpha)}{\alpha} \right) + \frac{1}{2} \left(\lim_{0<\alpha\to 0} \frac{\pi_k(\alpha)}{\alpha} \right)^T B_k \underbrace{\left(\lim_{0<\alpha\to 0} \pi_k(\alpha) \right)}_{=0} \\
=&\nabla f(x_k)^T \nabla_M f(x_k) \\
=&-\left\| \nabla_M f(x_k) \right\|^2 \\
<&0.
\end{aligned}$$

Da nach Voraussetzung $0 < \mu_0 < \frac{1}{2}$ ist, existiert ein $\varepsilon > 0$, so daß für alle α mit $0 < \alpha < \varepsilon$

$$\frac{f_k(\pi_k(\alpha)) - f_k(0)}{\alpha} < 2\mu_0 \nabla f(x_k)^T \nabla_M f(x_k)$$

gilt. Außerdem existiert wegen der oben gezeigten Darstellung für $\lim_{0<\alpha\to 0} \frac{\pi_k(\alpha)}{\alpha}$ ein $\delta > 0$, so daß für alle α mit $0 < \alpha < \delta$

$$\left| \nabla f(x_k)^T \nabla_M f(x_k) - \nabla f(x_k)^T \left(\frac{\pi_k(\alpha)}{\alpha} \right) \right| < -\nabla f(x_k)^T \nabla_M f(x_k)$$

ist. Insbesondere gilt also für diese α die Ungleichung

$$2\nabla f(x_k)^T \nabla_M f(x_k) < \nabla f(x_k)^T \left(\frac{\pi_k(\alpha)}{\alpha} \right)$$

und zusammen mit der ersten Ungleichung erhält man für alle α mit $0 < \alpha < \min\{\varepsilon, \delta\}$

$$\begin{aligned}
f_k(\pi_k(\alpha)) - f(x_k) <& 2\alpha\mu_0 \nabla f(x_k)^T \nabla_M f(x_k) \\
<& \mu_0 \nabla f(x_k)^T \pi_k(\alpha),
\end{aligned}$$

womit die Behauptung gezeigt ist. Erfüllt $\alpha_k := \gamma_1$ die Forderung (3.3.3), so sind wir fertig. Andernfalls wähle $\alpha_k := \gamma_1 \gamma_2{}^r$, wobei r die kleinste natürliche Zahl sei, für die (3.3.3) gilt. Setze nun $\hat{\alpha}_k := \gamma_1 \gamma_2{}^{r-1}$. Damit ist klar, daß α_k Forderung (3.3.4) und $\hat{\alpha}_k$ Forderung (3.3.5) erfüllt. ■

Nun werden wir noch zeigen, daß der Cauchy-Schritt tatsächlich eine Abstiegsrichtung ist, sollte es sich beim Iterationspunkt nicht um einen stationären Punkt handeln. Es gilt sogar die später häufiger benötigte Ungleichung

$$-\nabla f(x_k)^T \pi_k(\alpha) \geq \frac{\left\| \pi_k(\alpha) \right\|^2}{\alpha} > 0 \quad \text{für alle } \alpha > 0, \tag{3.3.6}$$

denn aus Lemma 2.1.2 ergibt sich die Ungleichungskette

$$\begin{aligned} 0 \leq & (\Pi_M(x_k - \alpha\nabla f(x_k)) - (x_k - \alpha\nabla f(x_k)))^T (x_k - \Pi_M(x_k - \alpha\nabla f(x_k))) \\ = & (\pi_k(\alpha) + \alpha\nabla f(x_k))^T (-\pi_k(\alpha)) \quad \text{für alle } \alpha > 0. \end{aligned}$$

Da x_k nicht stationär ist, haben wir $\|\nabla_M f(x_k)\| > 0$. Daher ist auch $\|\pi_k(\alpha)\| > 0$ für alle hinreichend kleinen $\alpha > 0$. Nach Satz 2.1.4, Teil (i) ist $\|\pi_k(\alpha)\|$ für $\alpha \geq 0$ monoton nichtfallend und daher

$$\frac{\|\pi_k(\alpha)\|^2}{\alpha} > 0 \quad \text{für alle } \alpha > 0.$$

Damit ist gezeigt, daß der Cauchy-Schritt eine Abstiegsrichtung und das Verfahren insgesamt durchführbar ist. Insbesondere kann man $p_k := p_k^C$ setzen.

3.4 Konvergenzanalyse ohne projizierten Gradienten

In diesem Abschnitt werden die einfachsten Ergebnisse der Konvergenzanalyse vorgestellt, wobei der projizierte Gradient nicht verwendet wird. Die Ergebnisse in diesem Abschnitt sind daher nicht sehr anschaulich, für die weitere Konvergenzanalyse aber unerläßlich. Der erste Schritt besteht darin, Abschätzungen über die Minderung in der Modellfunktion und in der Zielfunktion zu bekommen, die durch den Cauchy-Schritt erzielt werden. Das dazu bereitgestellte Lemma stammt aus Moré, 1988.

Lemma 3.4.1 *Sei $f : \mathbb{R}^n \to \mathbb{R}$ eine auf einer offenen Obermenge von M stetig differenzierbare Abbildung. Das Trust-Region-Verfahren breche nicht vorzeitig ab und sei $(x_k)_{k\in\mathbb{N}}$ die Folge, die durch dieses Verfahren erzeugt wird. Dann existiert ein $\zeta > 0$ derart, daß*

(3.4.7)

$$\begin{aligned} & -\nabla f(x_k)^T \pi_k(\alpha_k) \\ \geq & \zeta \left(\frac{\|\pi_k(\alpha_k)\|}{\alpha_k} \right) \min \left\{ \Delta_k, \frac{1}{1 + \|B_k\|} \left(\frac{\|\pi_k(\alpha_k)\|}{\alpha_k} \right) \right\} \quad \textit{für alle } k \in \mathbb{N} \end{aligned}$$

gilt.

Beweis. Der Beweis besteht darin, für einen fest vorgegebenen Punkt x_k die Forderungen 3.3.3 bis 3.3.5 an den Cauchy-Schritt zu untersuchen und zu Abschätzungen zu gelangen, die von k unabhängig sind. Sei also $k \in \mathbb{N}$ beliebig gewählt. Da das Trust-Region-Verfahren nach Voraussetzung nicht abbricht ist x_k nicht stationär und wir können Ungleichung (3.3.6) benutzen.

- Zunächst untersuchen wir die erste Möglichkeit der Forderung (3.3.4), und zwar, daß $\alpha_k \in [\gamma_1, \gamma_3]$ ist. In diesem Fall sieht man durch eine einfache Umformung von Ungleichung (3.3.6)

$$-\nabla f(x_k)^T \pi_k(\alpha_k) \geq \left(\frac{\|\pi_k(\alpha_k)\|}{\alpha_k} \right)^2 \alpha_k,$$

  so daß die gewünschte Ungleichung mit $0 < \zeta \leq \gamma_1$ erfüllt wird. $\surd$

- Als nächstes betrachten wir die zweite Möglichkeit der Forderung (3.3.4), und zwar, daß $\alpha_k \in [\gamma_2 \hat{\alpha}_k, \gamma_3]$ ist.

  – Dann ist die eine Möglichkeit aus (3.3.5), daß

$$f_k(\pi_k(\hat{\alpha}_k)) - f(x_k) > \mu_0 \nabla f(x_k)^T \pi_k(\hat{\alpha}_k)$$

    gilt. Daraus erhält man

$$\begin{aligned}(1-\mu_0)(-\nabla f(x_k)^T \pi_k(\hat{\alpha}_k)) <& f_k(\pi_k(\hat{\alpha}_k)) - f(x_k) - \nabla f(x_k)^T \pi_k(\hat{\alpha}_k) \\ =& \frac{1}{2} \pi_k(\hat{\alpha}_k)^T B_k \pi_k(\hat{\alpha}_k) \\ \leq& \frac{1}{2} \|B_k\| \, \|\pi_k(\hat{\alpha}_k)\|^2.\end{aligned}$$

    Schließlich ist dann

$$\frac{\|\pi_k(\hat{\alpha}_k)\|^2}{-\nabla f(x_k)^T \pi_k(\hat{\alpha}_k)} \geq \frac{2(1-\mu_0)}{\|B_k\|},$$

    wobei man beachte, daß $\nabla f(x_k)^T \pi_k(\alpha) < 0$ für alle $\alpha > 0$ nach (3.3.6) ist. Nach Wahl von α_k und (3.3.6) gilt außerdem die Ungleichungskette

$$\alpha_k \geq \gamma_2 \hat{\alpha}_k \geq \gamma_2 \frac{\|\pi_k(\hat{\alpha}_k)\|^2}{-\nabla f(x_k)^T \pi_k(\hat{\alpha}_k)}.$$

    Ebenfalls nach (3.3.6) und dem eben Gezeigten haben wir

$$\begin{aligned}-\nabla f(x_k)^T \pi_k(\alpha_k) \geq& \left(\frac{\|\pi_k(\alpha_k)\|}{\alpha_k} \right)^2 \alpha_k \\ \geq& \left(\frac{\|\pi_k(\alpha_k)\|}{\alpha_k} \right)^2 \gamma_2 \frac{\|\pi_k(\hat{\alpha}_k)\|^2}{-\nabla f(x_k)^T \pi_k(\hat{\alpha}_k)} \\ \geq& \left(\frac{\|\pi_k(\alpha_k)\|}{\alpha_k} \right)^2 \gamma_2 \frac{2(1-\mu_0)}{\|B_k\|}.\end{aligned}$$

    Wählt man $0 < \zeta \leq 2\gamma_2(1-\mu_0)$, so erhält man das gewünschte Ergebnis. $\surd$

– Wir nehmen nun an, daß die andere Forderung in (3.3.5), nämlich $\|\pi_k(\hat{\alpha}_k)\| > \mu_1\Delta_k$ erfüllt ist. Wir nehmen weiterhin ohne Beschränkung der Allgemeinheit an, daß $\gamma_2 \leq 1$ ist, denn (3.3.4) würde auch mit $\min\{\gamma_2, 1\}$ statt mit γ_2 gelten. Satz 2.1.4, Teil (ii) impliziert

$$\frac{\|\pi_k(\gamma_2\hat{\alpha}_k)\|}{\gamma_2\hat{\alpha}_k} \geq \frac{\|\pi_k(\hat{\alpha}_k)\|}{\hat{\alpha}_k}$$

und Teil (i) des gleichen Satzes

$$\|\pi_k(\alpha_k)\| \geq \|\pi_k(\gamma_2\hat{\alpha}_k)\| \geq \gamma_2 \|\pi_k(\hat{\alpha}_k)\| > \gamma_2\mu_1\Delta_k.$$

Ungleichung (3.3.6) ergibt nun

$$-\nabla f(x_k)^T\pi_k(\alpha_k) \geq \frac{\|\pi_k(\alpha_k)\|}{\alpha_k} \|\pi_k(\alpha_k)\| > \frac{\|\pi_k(\alpha_k)\|}{\alpha_k}\gamma_2\mu_1\Delta_k,$$

so daß die geforderte Ungleichung mit $0 < \zeta \leq \gamma_2\mu_1$ erfüllt ist. $\surd$

■

Diese Abschätzung ist in den meisten Fällen nicht sehr praktisch, da sie weder die Minderung in der Modellfunktion noch in der Zielfunktion ausdrückt. Zusammen mit der Forderung an den Cauchy-Schritt p_k^C in (3.3.3)

$$f(x_k) - f_k(\pi_k(\alpha_k)) \geq -\mu_0\nabla f(x_k)^T\pi_k(\alpha_k) \quad \text{und} \quad \|\pi_k(\alpha_k)\| \leq \mu_1\Delta_k$$

und der Forderung an den Schritt p_k in (3.1.1)

$$f(x_k) - f_k(p_k) \geq \mu_0(f(x_k) - f_k(p_k^C))$$

erhält man jedoch sofort die wesentlich handlichere Ungleichung über die Minderung in der Modellfunktion

(3.4.8)
$$\begin{aligned} &f(x_k) - f_k(p_k) \\ \geq &{\mu_0}^2\zeta\left(\frac{\|\pi_k(\alpha_k)\|}{\alpha_k}\right)\min\left\{\Delta_k, \frac{1}{1+\|B_k\|}\left(\frac{\|\pi_k(\alpha_k)\|}{\alpha_k}\right)\right\} \quad \text{für alle } k \in \mathbb{N}. \end{aligned}$$

Drückt man diese Ungleichung mit Hilfe des Quotienten χ_k aus, so bekommt man eine Abschätzung über die Minderung in der Zielfunktion, und zwar

(3.4.9)
$$\begin{aligned} &f(x_k) - f(x_k + p_k) \\ \geq &\chi_k{\mu_0}^2\zeta\left(\frac{\|\pi_k(\alpha_k)\|}{\alpha_k}\right)\min\left\{\Delta_k, \frac{1}{1+\|B_k\|}\left(\frac{\|\pi_k(\alpha_k)\|}{\alpha_k}\right)\right\} \quad \text{für alle } k \in \mathbb{N}. \end{aligned}$$

Die jetzt folgenden Konvergenzaussagen benötigen alle die Voraussetzung, daß die Folge $(f(x_k))_{k\in\mathbb{N}}$ nach unten beschränkt ist. Da wir später immer davon ausgehen werden, daß die Folge $(x_k)_{k\in\mathbb{N}}$ einen Häufungspunkt besitzt, ist dann diese Forderung automatisch erfüllt, da die Funktion f immer als stetig vorausgesetzt wird und das Verfahren eine monoton fallende Folge $(f(x_k))_{k\in\mathbb{N}}$ produziert. Der entscheidende Parameter, der die Qualität der Konvergenzaussagen beeinflußt, ist die Matrix B_k. Im weiteren ändern sich daher hauptsächlich die Voraussetzungen, die mit dieser Matrix verknüpft sind oder mit der ganzen Folge $(B_k)_{k\in\mathbb{N}}$. Die jetzt folgenden Aussagen in diesem Abschnitt findet man alle im vierten Abschnitt von Burke et al., 1990.

Satz 3.4.2 *Sei $f : \mathbb{R}^n \to \mathbb{R}$ eine auf einer offenen Obermenge von M stetig differenzierbare Abbildung. Das Trust-Region-Verfahren breche nicht vorzeitig ab und sei $(x_k)_{k\in\mathbb{N}}$ die Folge, die durch dieses Verfahren erzeugt wird. Sei weiterhin $(f(x_k))_{k\in\mathbb{N}}$ nach unten beschränkt und $(B_k)_{k\in\mathbb{N}}$ beschränkt. Dann gilt*

$$\liminf_{k\to\infty} \frac{\|\pi_k(\alpha_k)\|}{\alpha_k} = 0.$$

Beweis. Der Beweis erfolgt durch Widerspruch. Wir nehmen also an, daß ein $\varepsilon > 0$ existiert, so daß

$$\frac{\|\pi_k(\alpha_k)\|}{\alpha_k} \geq \varepsilon \quad \text{für alle } k \in \mathbb{N}$$

gilt. Wir behaupten, daß dann

$$\sum_{k=0}^{\infty} \Delta_k < \infty$$

ist. Um dies zu zeigen, betrachte man die Indizes $k \in \mathbb{N}$, bei denen $\chi_k > \rho_1$ ist. Gibt es nur eine endliche Anzahl von Iterationsindizes mit dieser Eigenschaft, dann ist $\Delta_{k+1} \leq \sigma_2 \Delta_k$ für alle hinreichend großen $k \in \mathbb{N}$, so daß man $\sum_{k=0}^{\infty} \Delta_k$ durch eine geometrische Reihe plus eine Konstante abschätzen kann, woraus zusammen mit $0 < \sigma_2 < 1$ deren Konvergenz folgt. Andernfalls sei $(k_i)_{i\in\mathbb{N}} \subseteq \mathbb{N}$ die Folge aller Iterationsindizes $k \in \mathbb{N}$ mit der Eigenschaft $\chi_k > \rho_1$. Dann ergibt sich aus Abschätzung (3.4.9)

$$f(x_{k_i}) - f(x_{k_i+1}) > \rho_1 {\mu_0}^2 \zeta \varepsilon \min\left\{\Delta_{k_i}, \frac{\varepsilon}{1 + \sup_{k\in\mathbb{N}} \|B_k\|}\right\} \quad \text{für alle } i \in \mathbb{N}.$$

Da weiterhin $(f(x_k))_{k\in\mathbb{N}}$ nach unten beschränkt ist, haben wir also

$$\sum_{i=0}^{\infty} \Delta_{k_i} < \infty.$$

Die Iterationsvorschrift für Δ_k impliziert weiter

$$\sum_{k=k_i}^{k_{i+1}-1} \Delta_k \leq \sum_{k=k_i}^{k_{i+1}-1} \sigma_3 {\sigma_2}^{k-k_i-1} \Delta_{k_i} \leq \frac{\sigma_3}{\sigma_2(1-\sigma_2)} \Delta_{k_i} \quad \text{für alle } i \in \mathbb{N},$$

womit man die Abschätzung

$$\sum_{k=0}^{\infty} \Delta_k = \sum_{k=0}^{k_0-1} \Delta_k + \sum_{i=0}^{\infty} \sum_{k=k_i}^{k_{i+1}-1} \Delta_k \leq \sum_{k=0}^{k_0-1} \Delta_k + \left(\frac{\sigma_3}{\sigma_2(1-\sigma_2)} \right) \sum_{i=0}^{\infty} \Delta_{k_i}$$

bekommt und daraus die Konvergenz von dessen Reihe. Wir zeigen nun, daß $\sum_{k=0}^{\infty} \Delta_k < \infty$

$$\lim_{k \to \infty} \chi_k = 1$$

impliziert. Aus der Ungleichung

$$\|x_{k+1} - x_k\| \leq \|p_k\| \leq \mu_1 \Delta_k \quad \text{für alle } k \in \mathbb{N}$$

erhält man zusammen mit der Dreiecksungleichung und $\sum_{k=0}^{\infty} \Delta_k < \infty$, daß $(x_k)_{k \in \mathbb{N}}$ eine Cauchy-Folge ist und damit konvergiert. Sei also

$$x^* := \lim_{k \to \infty} x_k.$$

Weiterhin sieht man sofort ein, daß $(p_k)_{k \in \mathbb{N}}$ gegen Null konvergiert. Wir definieren die Folge $(\varepsilon_k)_{k \in \mathbb{N}}$ durch

$$\varepsilon_k := \frac{\left| f(x_k + p_k) - f(x_k) - \nabla f(x_k)^T p_k \right|}{\|p_k\|}.$$

Dann erhält man aus der stetigen Differenzierbarkeit von f auf einer offenen Obermenge von M

$$\begin{aligned}
\varepsilon_k &= \left| \int_0^1 \left(\nabla f(x_k + t p_k) - \nabla f(x_k) \right)^T \left(\frac{p_k}{\|p_k\|} \right) \, dt \right| \\
&\leq \int_0^1 \|\nabla f(x_k + t p_k) - \nabla f(x_k)\| \, dt \\
&\leq \int_0^1 \left(\|\nabla f(x_k + t p_k) - \nabla f(x^*)\| + \|\nabla f(x^*) - \nabla f(x_k)\| \right) dt \quad \text{für alle } k \in \mathbb{N}
\end{aligned}$$

und damit die Konvergenz von $(\varepsilon_k)_{k \in \mathbb{N}}$ gegen Null. Hieraus ergibt sich

$$\begin{aligned}
|f(x_k + p_k) - f_k(p_k))| &= |f(x_k + p_k) - f(x_k) - (f_k(p_k) - f(x_k))| \\
&= \left| f(x_k + p_k) - f(x_k) - (\nabla f(x_k)^T p_k + \frac{1}{2} {p_k}^T B_k p_k) \right| \\
&\leq \left| f(x_k + p_k) - f(x_k) - \nabla f(x_k)^T p_k \right| + \frac{1}{2} \left| {p_k}^T B_k p_k \right| \\
&\leq \varepsilon_k \|p_k\| + \frac{1}{2} \|B_k\| \, \|p_k\|^2 \quad \text{für alle } k \in \mathbb{N}.
\end{aligned}$$

Aus der Ungleichung (3.4.8) und da aufgrund der Konvergenz von $(\Delta_k)_{k\in\mathbb{N}}$ gegen Null

$$\min\left\{\Delta_k, \frac{\varepsilon}{1+\sup_{k\in\mathbb{N}}\|B_k\|}\right\} = \Delta_k$$

für alle hinreichend großen $k \in \mathbb{N}$ ist, erhält man die Abschätzung

$$f(x_k) - f_k(p_k) \geq {\mu_0}^2 \zeta \varepsilon \Delta_k$$

für alle hinreichend großen $k \in \mathbb{N}$. Nun ist

$$\begin{aligned}
|\chi_k - 1| &= \left|\frac{f(x_k) - f(x_k+p_k) - (f(x_k) - f_k(p_k))}{f(x_k) - f_k(p_k)}\right| \\
&= \left|\frac{f(x_k+p_k) - f_k(p_k))}{f(x_k) - f_k(p_k)}\right| \\
&\leq \frac{\varepsilon_k \|p_k\| + \frac{1}{2}\|B_k\|\|p_k\|^2}{{\mu_0}^2\zeta\varepsilon\Delta_k} \\
&\leq \frac{\varepsilon_k\mu_2 + \frac{1}{2}\|B_k\|{\mu_2}^2\Delta_k}{{\mu_0}^2\zeta\varepsilon}
\end{aligned}$$

für alle hinreichend großen $k \in \mathbb{N}$ und man sieht, daß $\lim_{k\to\infty}\chi_k = 1$ ist. Die Iterationsvorschrift für Δ_k zeigt jedoch, daß Δ_k nicht verkleinert wird, falls $\chi_k \geq \rho_2$ ist. Daher kann $(\Delta_k)_{k\in\mathbb{N}}$ nicht gegen Null konvergieren und man erhält einen Widerspruch zur Voraussetzung, daß $\sum_{k=0}^{\infty}\Delta_k < \infty$ ist. Daher kann auch die anfangs gemachte Annahme nicht gelten und der Satz ist damit bewiesen. ■

Als nächstes benötigen wir zwei Definitionen.

Definition 3.4.3 (Niveaumenge) Sei $G \subseteq \mathbb{R}^n$ eine nichtleere Menge und $x \in G$. Sei weiterhin $f : \mathbb{R}^n \to \mathbb{R}$ eine Abbildung. Die *Niveaumenge* von f in x bezüglich G ist durch

$$L(G;x) := \{y \in G \mid f(y) \leq f(x)\}$$

definiert.

Weiterhin definieren wir die Folge $(\beta_k)_{k\in\mathbb{N}}$ durch

$$\beta_k := 1 + \max_{0\leq i\leq k}\|B_i\| .$$

Das nächste Lemma dient als Vorbereitung für den darauffolgenden Satz.

Lemma 3.4.4 *Sei $f : \mathbb{R}^n \to \mathbb{R}$ eine auf einer offenen Obermenge von M stetig differenzierbare Abbildung. Das Trust-Region-Verfahren breche nicht vorzeitig ab und sei $(x_k)_{k\in\mathbb{N}}$ die Folge, die durch dieses Verfahren erzeugt wird. Sei weiterhin ∇f auf der Niveaumenge $L(M; x_0)$ gleichmäßig stetig und die Folge $(f(x_k))_{k\in\mathbb{N}}$ nach unten beschränkt. Es existiere ein $\varepsilon > 0$, so daß*

$$\frac{\|\pi_k(\alpha_k)\|}{\alpha_k} \geq \varepsilon \quad \text{für alle } k \in \mathbb{N}$$

gilt. Dann existiert ein $\delta > 0$, so daß

$$\beta_k \Delta_k \geq \delta \quad \text{für alle } k \in \mathbb{N}$$

gilt.

Beweis. Zuerst zeigen wir, daß ein $\hat{\delta} > 0$ existiert, so daß

$$(1 + \|B_k\|)\,\|p_k\| \geq \hat{\delta} \quad \text{für alle } k \in \mathbb{N} \quad \text{mit } \chi_k < \rho_2 \tag{3.4.10}$$

ist. Angenommen, es existiert eine Folge $(k_i)_{i\in\mathbb{N}} \subseteq \mathbb{N}$ von Iterationsindizes mit der Eigenschaft $\chi_{k_i} < \rho_2$ für alle $i \in \mathbb{N}$ und $\lim_{i\to\infty}((1 + \|B_{k_i}\|)\,\|p_{k_i}\|) = 0$. Da $1 + \|B_k\| \geq 1$ für alle $k \in \mathbb{N}$ ist, gilt sogar $\lim_{i\to\infty} \|p_{k_i}\| = 0$. Wir definieren die Folge $(\varepsilon_i)_{i\in\mathbb{N}}$ durch

$$\varepsilon_i := \frac{\left| f(x_{k_i} + p_{k_i}) - f(x_{k_i}) - \nabla f(x_{k_i})^T p_{k_i} \right|}{\|p_{k_i}\|}.$$

Dann erhält man aus der stetigen Differenzierbarkeit von f auf einer offenen Obermenge von M

$$\begin{aligned}
\varepsilon_i &= \left| \int_0^1 (\nabla f(x_{k_i} + t p_{k_i}) - \nabla f(x_{k_i}))^T \left(\frac{p_{k_i}}{\|p_{k_i}\|} \right) dt \right| \\
&\leq \int_0^1 \|\nabla f(x_{k_i} + t p_{k_i}) - \nabla f(x_{k_i})\|\, dt \quad \text{für alle } i \in \mathbb{N}.
\end{aligned}$$

Da ∇f auf der Niveaumenge $L(M; x_0)$ gleichmäßig stetig und die Folge $(f(x_k))_{k\in\mathbb{N}}$ monoton fallend ist, erhält man, obwohl $(x_{k_i})_{i\in\mathbb{N}}$ nicht notwendigerweise konvergieren muß, die Konvergenz von $(\varepsilon_i)_{i\in\mathbb{N}}$ gegen Null. Hieraus ergibt sich

$$\begin{aligned}
&|f(x_{k_i} + p_{k_i}) - f_{k_i}(p_{k_i}))| \\
=&\,|f(x_{k_i} + p_{k_i}) - f(x_{k_i}) - (f_{k_i}(p_{k_i}) - f(x_{k_i}))| \\
=&\left| f(x_{k_i} + p_{k_i}) - f(x_{k_i}) - (\nabla f(x_{k_i})^T p_{k_i} + \frac{1}{2} {p_{k_i}}^T B_{k_i} p_{k_i}) \right| \\
\leq&\left| f(x_{k_i} + p_{k_i}) - f(x_{k_i}) - \nabla f(x_{k_i})^T p_{k_i} \right| + \frac{1}{2} \left| {p_{k_i}}^T B_{k_i} p_{k_i} \right| \\
\leq&\,\varepsilon_i \|p_{k_i}\| + \frac{1}{2} \|B_{k_i}\| \|p_{k_i}\|^2 \quad \text{für alle } i \in \mathbb{N}.
\end{aligned}$$

Aus der Ungleichung (3.4.8) zusammen mit der Voraussetzung dieses Satzes bekommt man

$$f(x_{k_i}) - f_{k_i}(p_{k_i}) \geq {\mu_0}^2 \zeta \varepsilon \min\left\{\Delta_{k_i}, \frac{\varepsilon}{1 + \|B_{k_i}\|}\right\} \quad \text{für alle } i \in \mathbb{N}.$$

Da $\lim_{i\to\infty}(1 + \|B_{k_i}\|)\,\|p_{k_i}\| = 0$ und $\|p_k\| \leq \mu_1 \Delta_k$ für alle $k \in \mathbb{N}$ ist, erhalten wir die Abschätzung

$$f(x_{k_i}) - f_{k_i}(p_{k_i}) \geq {\mu_0}^2 \zeta \varepsilon \frac{\|p_{k_i}\|}{\mu_1}$$

für alle hinreichend großen $i \in \mathbb{N}$. Nun gilt

$$\begin{aligned}
|\chi_{k_i} - 1| &= \left| \frac{f(x_{k_i}) - f(x_{k_i} + p_{k_i}) - (f(x_{k_i}) - f_{k_i}(p_{k_i}))}{f(x_{k_i}) - f_{k_i}(p_{k_i})} \right| \\
&= \left| \frac{f(x_{k_i} + p_{k_i}) - f_{k_i}(p_{k_i})}{f(x_{k_i}) - f_{k_i}(p_{k_i})} \right| \\
&\leq \frac{\varepsilon_i \|p_{k_i}\| + \frac{1}{2}\|B_{k_i}\|\,\|p_{k_i}\|^2}{{\mu_0}^2 \zeta \varepsilon \left(\frac{\|p_{k_i}\|}{\mu_1}\right)} \\
&\leq \frac{\mu_1 \varepsilon_i + \frac{1}{2}\mu_1 \|B_{k_i}\|\,\|p_{k_i}\|}{{\mu_0}^2 \zeta \varepsilon}
\end{aligned}$$

für alle hinreichend großen $i \in \mathbb{N}$ und man sieht, daß $\lim_{i\to\infty} \chi_{k_i} = 1$ ist. Dieser Widerspruch zeigt, daß (3.4.10) gilt. Wir zeigen die Behauptung des Satzes für

$$\delta := \min\{\beta_0 \Delta_0, \sigma_1 \hat{\delta}\}$$

jetzt mit Induktion über k. Offenbar ist die Behauptung für $k := 0$ erfüllt. Sei deshalb die Behauptung jetzt für ein beliebiges $k \in \mathbb{N}$ erfüllt. Ist $\chi_k \geq \rho_2$, so folgt $\beta_{k+1}\Delta_{k+1} \geq \beta_k \Delta_k$ aufgrund der Definition von β_k und der Iterationsvorschrift für Δ_k. Ist andernfalls $\chi_k < \rho_2$, so gilt

$$\beta_{k+1}\Delta_{k+1} \geq (1 + \|B_k\|)\sigma_1 \|p_k\| \geq \sigma_1 \hat{\delta}$$

und damit ist das Lemma bewiesen. ■

Im folgenden Satz werden die Voraussetzungen an die Folge $(B_k)_{k\in\mathbb{N}}$ gelockert, daher mußte $(\beta_k)_{k\in\mathbb{N}}$ definiert werden. So wird jetzt nur noch gefordert, daß $\sum_{k=0}^{\infty} \frac{1}{\beta_k} = \infty$ ist und nicht mehr wie in Satz 3.4.2 die Beschränktheit von $(B_k)_{k\in\mathbb{N}}$ vorausgesetzt. Dafür muß ∇f jetzt auf der Niveaumenge $L(M; x_0)$ gleichmäßig stetig sein.

Satz 3.4.5 *Sei $f : \mathbb{R}^n \to \mathbb{R}$ eine auf einer offenen Obermenge von M stetig differenzierbare Abbildung. Das Trust-Region-Verfahren breche nicht vorzeitig ab und sei $(x_k)_{k\in\mathbb{N}}$ die Folge, die durch dieses Verfahren erzeugt wird. Sei weiterhin ∇f auf der Niveaumenge $L(M; x_0)$ gleichmäßig stetig, die Folge $(f(x_k))_{k\in\mathbb{N}}$ nach unten beschränkt und*

$$\sum_{k=0}^{\infty} \frac{1}{\beta_k} = \infty.$$

Dann gilt

$$\liminf_{k\to\infty} \frac{\|\pi_k(\alpha_k)\|}{\alpha_k} = 0.$$

Beweis. Wir beweisen diesen Satz durch Widerspruch. Es existiere also ein $\varepsilon > 0$, so daß

$$\frac{\|\pi_k(\alpha_k)\|}{\alpha_k} \geq \varepsilon \quad \text{für alle } k \in \mathbb{N}$$

gilt. Damit ist gleichzeitig die Voraussetzung von Satz 3.4.4 erfüllt. Wir werden nun zeigen, daß $\sum_{k=0}^{\infty} \frac{1}{\beta_k} < \infty$ ist. Betrachte dazu die Menge

$$J := \left\{ k \in \mathbb{N} \mid \varphi(k) \geq \frac{k}{\nu} \right\},$$

wobei die Funktion $\varphi : \mathbb{N} \to \mathbb{N}$ dadurch definiert sei, daß $\varphi(k)$ die Anzahl der erfolgreichen Iterationsindizes sei, die nicht größer als k sind. Dabei sei $\nu \in \mathbb{N}$ mit $\nu > 0$ so gewählt, daß ${\sigma_3\sigma_2}^{\nu-1} < 1$ ist. Im ersten Teil des Beweises werden wir zeigen, daß

$$\sum_{k\notin J} \frac{1}{\beta_k} < \infty$$

ist und im zweiten Teil, daß

$$\sum_{k\in J} \frac{1}{\beta_k} < \infty$$

ist, woraus man das gewünschte Resultat erhält. Zum Beweis des ersten Teils überlegt man sich leicht, daß aus der Iterationsvorschrift für Δ_k und der Definition der Funktion φ

$$\Delta_k \leq {\sigma_3}^{\varphi(k)} {\sigma_2}^{k-\varphi(k)} \Delta_0 \quad \text{für alle } k \in \mathbb{N}$$

folgt und damit, da sich aus $k \notin J$ die Abschätzung $k - \varphi(k) > (\nu - 1)\frac{k}{\nu}$ ergibt,

$$\Delta_k \leq ({\sigma_3\sigma_2}^{\nu-1})^{\frac{k}{\nu}} \Delta_0 \quad \text{für alle } k \notin J.$$

Lemma 3.4.4 zeigt, daß ein $\delta > 0$ existiert, so daß $\frac{\delta}{\beta_k} \leq \Delta_k$ für alle $k \in \mathbb{N}$ ist und zusammen mit ${\sigma_3\sigma_2}^{\nu-1} < 1$ erhält man $\sum_{k \notin J} \frac{1}{\beta_k} < \infty$. Zum Beweis des zweiten Teils sei $\{k_i \in \mathbb{N} \mid i \in S\}$ die Menge der erfolgreichen Iterationsindizes, wobei S eine endliche Teilmenge von $\mathbb{N}$ oder $\mathbb{N}$ selber sein kann, je nachdem, ob die Anzahl der erfolgreichen Indizes endlich oder unendlich ist. Wir werden zuerst zeigen, daß

$$\sum_{i \in S} \frac{1}{\beta_{k_i}} < \infty$$

gilt. Man beachte, daß aus Abschätzung (3.4.9)

$$f(x_{k_i}) - f(x_{k_i+1}) > \rho_0 {\mu_0}^2 \zeta\varepsilon \min\left\{\Delta_{k_i}, \frac{\varepsilon}{1 + \|B_{k_i}\|}\right\} \qquad \text{für alle } i \in S$$

folgt und Lemma 3.4.4 zeigt, daß ein $\delta > 0$ existiert, so daß

$$\begin{aligned}\rho_0 {\mu_0}^2 \zeta\varepsilon \min\left\{\Delta_{k_i}, \frac{\varepsilon}{1 + \|B_{k_i}\|}\right\} &\geq \rho_0 {\mu_0}^2 \zeta\varepsilon \min\left\{\frac{\delta}{\beta_{k_i}}, \frac{\varepsilon}{1 + \|B_{k_i}\|}\right\} \\ &\geq \rho_0 {\mu_0}^2 \zeta\varepsilon \min\{\delta, \varepsilon\} \frac{1}{\beta_{k_i}} \qquad \text{für alle } i \in S\end{aligned}$$

ist, womit wir $\sum_{i \in S} \frac{1}{\beta_{k_i}} < \infty$ bewiesen haben. Daher gilt $\sum_{k \in J} \frac{1}{\beta_k} < \infty$, wenn wir

$$\sum_{k \in J} \frac{1}{\beta_k} \leq \nu \sum_{i \in S} \frac{1}{\beta_{k_i}}$$

zeigen. Sei dazu $i \in S$ beliebig vorgegeben. Wir behaupten, daß $\beta_k \geq \beta_{k_i}$ für alle $k \in J_i := J \cap [i\nu, (i+1)\nu - 1]$ ist. Dies ist der Fall, falls $k \geq k_i$ ist, da die Folge $(\beta_k)_{k \in \mathbb{N}}$ nichtfallend ist. Im anderen Fall nehmen wir $i\nu \leq k < k_i$ an, woraus man

$$\varphi(k) < \varphi(k_i) = i \leq \frac{k}{\nu}$$

erhält und damit $k \notin J$ und erst recht $k \notin J_i$. Da also nun $\beta_k \geq \beta_{k_i}$ für alle $k \in J_i$ gezeigt ist, gilt

$$\sum_{k \in J_i} \frac{1}{\beta_k} \leq \frac{\nu}{\beta_{k_i}}$$

und so

$$\sum_{k \in J} \frac{1}{\beta_k} \leq \nu \sum_{i \in S} \frac{1}{\beta_{k_i}},$$

womit der Satz bewiesen ist. ■

Aus Satz 3.4.5 kann man direkt folgern, daß die Anzahl der erfolgreichen Iterationsindizes unendlich ist, sollte das Verfahren nicht vorzeitig abbrechen.

Satz 3.4.6 *Sei $f : \mathbb{R}^n \to \mathbb{R}$ eine auf einer offenen Obermenge von M stetig differenzierbare Abbildung. Das Trust-Region-Verfahren breche nicht vorzeitig ab und sei $(x_k)_{k\in\mathbb{N}}$ die Folge, die durch dieses Verfahren erzeugt wird. Sei weiterhin ∇f auf der Niveaumenge $L(M; x_0)$ gleichmäßig stetig, die Folge $(f(x_k))_{k\in\mathbb{N}}$ nach unten beschränkt und*

$$\sum_{k=0}^{\infty} \frac{1}{\beta_k} = \infty.$$

Dann ist die Menge S der erfolgreichen Iterationsindizes unendlich.

Beweis. Wir nehmen an, es existiere ein $\hat{k} \in \mathbb{N}$, so daß alle Iterationsindizes $k \geq \hat{k}$ nicht erfolgreich sind. Aus $x_k = x_{\hat{k}}$ für alle $k \geq \hat{k}$ ergibt sich

$$\pi_k(\alpha) = \pi_{\hat{k}}(\alpha) \quad \text{für alle } \alpha > 0 \quad \text{und} \quad \text{alle } k \geq \hat{k}.$$

Mit $\alpha_k \leq \gamma_3$ für alle $k \in \mathbb{N}$ erhalten wir aus Satz 2.1.4, Teil (ii)

$$\frac{\|\pi_k(\alpha_k)\|}{\alpha_k} \geq \frac{\|\pi_k(\gamma_3)\|}{\gamma_3} = \frac{\|\pi_{\hat{k}}(\gamma_3)\|}{\gamma_3} \quad \text{für alle } k \geq \hat{k}.$$

Satz 3.4.5 impliziert $\|\pi_{\hat{k}}(\gamma_3)\| = 0$ und Satz 2.1.4, Teil (i) ergibt $\|\pi_{\hat{k}}(\alpha)\| = 0$ für alle $\alpha \leq \gamma_3$ und damit

$$\nabla_M f(x_{\hat{k}}) = \lim_{0<\alpha\to 0} \frac{\pi_{\hat{k}}(\alpha)}{\alpha} = 0.$$

Nach Satz 2.2.3, Teil (iii) ist damit $x_{\hat{k}}$ ein stationärer Punkt der Aufgabe, f auf M zu minimieren, womit ein Widerspruch zur Voraussetzung, daß das Verfahren nicht vorzeitig abbreche, vorliegt. ■

3.5 Konvergenzanalyse mit projiziertem Gradienten

Bezieht man den projizierten Gradienten in die Konvergenzanalyse mit ein, so werden die eher technisch anmutenden Resultate des letzten Abschnitts für den Leser anschaulich. So sagen die Hauptresultate dieses Abschnitts, das sind die Sätze 3.5.4 und 3.5.5, aus, daß unter geeigneten Voraussetzungen und der Annahme der Existenz eines Häufungspunktes der Folge $(x_k)_{k\in\mathbb{N}}$ dieser Häufungspunkt stationär ist. Der Unterschied zwischen den beiden Sätzen besteht darin, daß unter stärkeren Voraussetzungen als in Satz 3.5.4 in Satz 3.5.5 sogar eine Teilfolge von erfolgreichen Iterationsindizes konstruiert werden kann, deren Iterationspunkte gegen den stationären Punkt konvergieren.

Als erstes jedoch benötigen wir eine Abschätzung über den projizierten Gradienten, die diesen mit den Sätzen des letzten Abschnitts in Verbindung bringt. Die Sätze dieses Abschnitts stammen alle aus dem fünften Abschnitt von Burke et al., 1990.

Lemma 3.5.1 *Sei $f : \mathbb{R}^n \to \mathbb{R}$ eine auf einer offenen Obermenge von M differenzierbare Abbildung und $x_k \in M$ mit $k \in \mathbb{N}$ ein Iterationspunkt des Trust-Region-Verfahrens. Dann gilt*

$$\left\|\nabla_M f(x_k^C)\right\| \leq \left\|\nabla f(x_k^C) - \nabla f(x_k)\right\| + \frac{\left\|x_k^C - x_k\right\|}{\alpha_k}.$$

Beweis. Aus Lemma 2.1.2 folgt

$$(\Pi_M(x_k - \alpha_k \nabla f(x_k)) - (x_k - \alpha_k \nabla f(x_k)))^T (\Pi_M(x_k - \alpha_k \nabla f(x_k)) - y) \leq 0$$

für alle $y \in M$ und mit der Definition des Cauchy-Punktes x_k^C durch

$$x_k^C = \Pi_M(x_k - \alpha_k \nabla f(x_k))$$

ergibt sich die Ungleichung

$$\begin{aligned} \alpha_k \nabla f(x_k)^T (x_k^C - y) &\leq -\left(x_k^C - x_k\right)^T (x_k^C - y) \\ &\leq \left\|x_k^C - x_k\right\| \left\|x_k^C - y\right\| \qquad \text{für alle } y \in M. \end{aligned}$$

Sei nun $p \in \mathbb{R}^n$ eine zulässige Richtung an M in x_k^C mit $\|p\| \leq 1$. Dann ist $x_k^C + tp \in M$ für ein $t > 0$ und setzt man $y := x_k^C + tp$, so ergibt sich

$$-\nabla f(x_k)^T p \leq \frac{\left\|x_k^C - x_k\right\|}{\alpha_k}$$

und damit

$$\begin{aligned} -\nabla f(x_k^C)^T p &= -\left(\nabla f(x_k^C) - \nabla f(x_k) + \nabla f(x_k)\right)^T p \\ &\leq \left\|\nabla f(x_k^C) - \nabla f(x_k)\right\| + \frac{\left\|x_k^C - x_k\right\|}{\alpha_k}. \end{aligned}$$

Da M seiner Definition nach konvex ist, ist der Tangentialkegel $T(M; x_k^C)$ gleichzeitig der Abschluß der zulässigen Richtungen an M in x_k^C, womit Satz 2.2.3, Teil (ii) das gewünschte Resultat liefert. ■

Der nächste Satz ist unter Einbeziehung von Lemma 3.5.1 eine direkte Konsequenz aus Satz 3.4.5.

Satz 3.5.2 *Sei $f : \mathbb{R}^n \to \mathbb{R}$ eine auf einer offenen Obermenge von M stetig differenzierbare Abbildung. Das Trust-Region-Verfahren breche nicht vorzeitig ab und sei $(x_k)_{k\in\mathbb{N}}$ die Folge, die durch dieses Verfahren erzeugt wird. Sei weiterhin ∇f auf der Niveaumenge $L(M; x_0)$ gleichmäßig stetig, die Folge $(f(x_k))_{k\in\mathbb{N}}$ nach unten beschränkt und*

$$\sum_{k=0}^{\infty} \frac{1}{\beta_k} = \infty.$$

Dann gilt

$$\liminf_{k\to\infty} \left\|\nabla_M f(x_k^C)\right\| = 0.$$

Beweis. Dieses Resultat folgt mit Lemma 3.5.1 aus Satz 3.4.5. Nach letztgenanntem Satz haben wir

$$\lim_{k\to\infty} \frac{\|\pi_k(\alpha_k)\|}{\alpha_k} = 0.$$

Aufgrund der Identität $x_k^C - x_k = \pi_k(\alpha_k)$ für alle $k \in \mathbb{N}$ existiert also eine Teilfolge $(k_i)_{i\in\mathbb{N}} \subseteq \mathbb{N}$ von Iterationsindizes, so daß

$$\lim_{i\to\infty} \frac{\left\|x_{k_i}^C - x_{k_i}\right\|}{\alpha_{k_i}} = 0$$

gilt. Da $\alpha_k \leq \gamma_3$ für alle $k \in \mathbb{N}$ ist, erhält man für diese Teilfolge auch

$$\lim_{i\to\infty} \left\|x_{k_i}^C - x_{k_i}\right\| = 0.$$

Wegen der gleichmäßigen Stetigkeit von ∇f auf $L(M; x_0)$ bekommen wir

$$\lim_{i\to\infty} \left\|\nabla f(x_{k_i}^C) - \nabla f(x_{k_i})\right\| = 0.$$

Die Abschätzung

$$\left\|\nabla_M f(x_{k_i}^C)\right\| \leq \left\|\nabla f(x_{k_i}^C) - \nabla f(x_{k_i})\right\| + \frac{\left\|x_{k_i}^C - x_{k_i}\right\|}{\alpha_{k_i}} \quad \text{für alle } i \in \mathbb{N}$$

aus Lemma 3.5.1 vollendet schließlich den Beweis. ∎

Im folgenden Satz werden Voraussetzungen an die Modellfunktion f_k und damit erneut an die Matrix B_k gemacht, da das der einzige Parameter ist, der in jedem Schleifendurchlauf neu gewählt werden darf. So wird gefordert, daß die Implikation

$$\sum_{k=0}^{\infty} \|p_k\| < \infty \quad \Longrightarrow \quad \lim_{k\to\infty} \frac{f(x_k + p_k) - f_k(p_k)}{\|p_k\|^2} = 0 \tag{3.5.11}$$

gilt. Ist f auf einer offenen Obermenge von M zweimal stetig differenzierbar, so existiert nach dem Satz von Taylor für jedes $k \in \mathbb{N}$ ein $0 < t_k < 1$, so daß

$$f(x_k + p_k) = f(x_k) + \nabla f(x_k)^T p_k + \frac{1}{2} {p_k}^T \nabla^2 f(x_k + t_k p_k) p_k$$

ist. Die Forderung (3.5.11) ist dann im Fall der Konvergenz von $(x_k)_{k\in\mathbb{N}}$ äquivalent zu

$$\sum_{k=0}^{\infty} \|p_k\| < \infty \quad \Longrightarrow \quad \lim_{k\to\infty} \frac{{p_k}^T (\nabla^2 f(x_k) - B_k) p_k}{\|p_k\|^2} = 0.$$

Der Nachteil von Satz 3.5.2, daß er keine Aussage über das Verhalten von $\left\|\nabla_M f(x_k^C)\right\|$ bezüglich der Menge der erfolgreichen Iterationsindizes macht, wird damit behoben. Im Gegensatz zu Theorem 5.3 aus Burke et al., 1990 muß meiner Ansicht nach auch die gleichmäßige Stetigkeit von ∇f auf der Niveaumenge $L(M; x_0)$ vorausgesetzt werden, da man andernfalls nicht die für die Anwendung von Lemma 3.5.1 notwendige Gleichung (3.5.12) folgern könnte.

Satz 3.5.3 *Sei $f : \mathbb{R}^n \to \mathbb{R}$ eine auf einer offenen Obermenge von M stetig differenzierbare Abbildung. Das Trust-Region-Verfahren breche nicht vorzeitig ab und sei $(x_k)_{k\in\mathbb{N}}$ die Folge, die durch dieses Verfahren erzeugt wird. Sei weiterhin ∇f auf der Niveaumenge $L(M; x_0)$ gleichmäßig stetig, die Folge $(f(x_k))_{k\in\mathbb{N}}$ nach unten beschränkt, die Folge $(B_k)_{k\in\mathbb{N}}$ beschränkt und es gelte die Implikation (3.5.11). Dann gilt*

$$\liminf_{\substack{k\to\infty \\ k\in S}} \left\|\nabla_M f(x_k^C)\right\| = 0,$$

wobei S die Menge der erfolgreichen Iterationssindizes sei.

Beweis. Der Beweis erfolgt in sieben Schritten.

- Es genügt,
$$\lim_{\substack{k\to\infty \\ k\in S}} \frac{\|\pi_k(\alpha_k)\|}{\alpha_k} = 0$$
zu zeigen. Denn aufgrund der Identität $x_k^C - x_k = \pi_k(\alpha_k)$ existiert eine Teilfolge $(k_i)_{i\in\mathbb{N}} \subseteq S$, so daß
$$\lim_{i\to\infty} \frac{\left\|x_{k_i}^C - x_{k_i}\right\|}{\alpha_{k_i}} = 0$$
ist. Da $\alpha_k \leq \gamma_3$ für alle $k \in \mathbb{N}$ ist, erhält man für diese Teilfolge auch
$$\lim_{i\to\infty} \left\|x_{k_i}^C - x_{k_i}\right\| = 0.$$

Aufgrund der gleichmäßigen Stetigkeit von ∇f auf $L(M; x_0)$ bekommen wir

$$\lim_{i\to\infty} \left\|\nabla f(x_{k_i}^C) - \nabla f(x_{k_i})\right\| = 0. \tag{3.5.12}$$

Die Abschätzung

$$\left\|\nabla_M f(x_{k_i}^C)\right\| \le \left\|\nabla f(x_{k_i}^C) - \nabla f(x_{k_i})\right\| + \frac{\left\|x_{k_i}^C - x_{k_i}\right\|}{\alpha_{k_i}} \quad \text{für alle } i \in \mathbb{N}$$

aus Lemma 3.5.1 schließt dann den Beweis. √

- Der Beweis der anfangs aufgestellten Behauptung erfolgt durch Widerspruch. Wir nehmen also an, es existiert ein $\varepsilon > 0$, so daß

$$\frac{\|\pi_k(\alpha_k)\|}{\alpha_k} \ge \varepsilon \quad \text{für alle } k \in S \tag{3.5.13}$$

ist. Wir werden zeigen, daß diese Annahme die Aussagen

$$\liminf_{\substack{k\to\infty \\ k\in S}} \frac{\|\pi_k(\alpha_k)\|}{\Delta_k} > 0, \tag{3.5.14}$$

$$\lim_{\substack{k\to\infty \\ k\notin S}} \frac{\|\pi_k(\alpha_k)\|}{\alpha_k} = 0, \tag{3.5.15}$$

$$\lim_{\substack{k\to\infty \\ k\notin S}} \frac{\|\pi_k(\alpha_k)\|}{\|p_k\|} = 0 \tag{3.5.16}$$

impliziert. √

- Als erstes aber zeigen wir, daß die Annahme (3.5.13) und die sich daraus ergebenen Folgerungen (3.5.14) bis (3.5.16) zu einem Widerspruch führen, bevor wir die Folgerungen selbst beweisen. Betrachte dazu die Menge der erfolgreichen Iterationsindizes S. Nach Satz 3.4.6 ist diese Menge unendlich. Sei Z eine unendliche Teilmenge von Iterationsindizes, so daß $Z \cap S = \varnothing$ ist und daß $k \in Z$ impliziert, daß $k+1 \in S$ ist. Wäre nur eine endliche Menge an Iterationsindizes nicht erfolgreich, erhielte man mit Satz 3.4.5 einen Widerspruch direkt zur Annahme (3.5.13). Da jeder Iterationsindex $k \in Z$ nicht erfolgreich ist, hat man $\pi_{k+1}(\alpha) = \pi_k(\alpha)$ für alle $\alpha > 0$ und alle $k \in Z$. Wäre $\alpha_{k+1} \ge \alpha_k$ für unendlich viele $k \in Z$, bekäme man mit Satz 2.1.4, Teil (ii) und (3.5.13)

$$\varepsilon \le \frac{\|\pi_{k+1}(\alpha_{k+1})\|}{\alpha_{k+1}} \le \frac{\|\pi_{k+1}(\alpha_k)\|}{\alpha_k} = \frac{\|\pi_k(\alpha_k)\|}{\alpha_k} \quad \text{für alle } k \in Z.$$

Nach (3.5.15) kann das nicht unendlich oft passieren und daher gilt $\alpha_{k+1} < \alpha_k$ für alle hinreichend großen $k \in Z$. Mit Satz 2.1.4, Teil (i) erhält man daraus

$$\|\pi_{k+1}(\alpha_{k+1})\| \leq \|\pi_{k+1}(\alpha_k)\| = \|\pi_k(\alpha_k)\|$$

für alle hinreichend großen $k \in Z$. Nach der Iterationsvorschrift für Δ_k ist $\sigma_1 \|p_k\| \leq \Delta_{k+1}$ für alle $k \in \mathbb{N}$, so daß man

$$\frac{\|\pi_{k+1}(\alpha_{k+1})\|}{\Delta_{k+1}} \leq \frac{\|\pi_{k+1}(\alpha_{k+1})\|}{\sigma_1 \|p_k\|} \leq \frac{\|\pi_k(\alpha_k)\|}{\sigma_1 \|p_k\|}$$

für alle hinreichend großen $k \in Z$ bekommt. Da $k+1 \in S$ für alle $k \in Z$ ist, erhält man mit Hilfe von (3.5.14) und (3.5.16)

$$\begin{aligned}
0 &< \lim_{i\to\infty} \inf_{\{k\in S|k\geq i\}} \frac{\|\pi_k(\alpha_k)\|}{\Delta_k} \\
&\leq \lim_{i\to\infty} \sup_{\{k\in Z|k\geq i\}} \frac{\|\pi_{k+1}(\alpha_{k+1})\|}{\Delta_{k+1}} \\
&\leq \lim_{i\to\infty} \sup_{\{k\in Z|k\geq i\}} \frac{\|\pi_k(\alpha_k)\|}{\sigma_1 \|p_k\|} \\
&= \lim_{\substack{k\to\infty \\ k\notin S}} \frac{\|\pi_k(\alpha_k)\|}{\sigma_1 \|p_k\|} \\
&= 0,
\end{aligned}$$

womit der Widerspruch, der sich aus (3.5.13) und seinen Folgerungen ergibt, ersichtlich wird. $\surd$

- Als nächstes werden wir als Vorbereitung zeigen, daß aus Annahme (3.5.13)

$$\sum_{k=0}^{\infty} \|p_k\| < \infty$$

folgt und daher die implizierte Aussage in (3.5.11) gültig ist. Die Abschätzung (3.4.9) vereinfacht sich unter den gemachten Annahmen zu

$$f(x_k) - f(x_{k+1}) \geq \rho_0 {\mu_0}^2 \zeta\varepsilon \min\left\{\Delta_k, \frac{\varepsilon}{1+\|B_k\|}\right\} \qquad \text{für alle } k \in S.$$

Da $(f(x_k))_{k\in\mathbb{N}}$ nach unten beschränkt und $(B_k)_{k\in\mathbb{N}}$ beschränkt ist, konvergiert $(\Delta_k)_{k\in S}$ gegen Null. Speziell ist

$$f(x_k) - f(x_{k+1}) \geq \rho_0 {\mu_0}^2 \zeta\varepsilon\Delta_k$$

für alle hinreichend großen $k \in S$ und man bekommt auch wegen der Beschränktheit von $(f(x_k))_{k\in\mathbb{N}}$ nach unten

$$\sum_{k\in S} \Delta_k < \infty.$$

Um $\sum_{k=0}^{\infty} \Delta_k < \infty$ zu zeigen, sei der Einfachheit halber $(k_i)_{i\in\mathbb{N}} \subseteq \mathbb{N}$ die Folge der erfolgreichen Iterationsindizes, also $(k_i)_{i\in\mathbb{N}} = S$. Die Iterationsvorschrift für Δ_k impliziert

$$\sum_{k=k_i}^{k_{i+1}-1} \Delta_k \leq \sum_{k=k_i}^{k_{i+1}-1} \sigma_3 {\sigma_2}^{k-k_i-1} \Delta_{k_i} \leq \frac{\sigma_3}{\sigma_2(1-\sigma_2)} \Delta_{k_i} \quad \text{für alle } i \in \mathbb{N},$$

womit man die Abschätzung

$$\sum_{k=0}^{\infty} \Delta_k = \sum_{k=0}^{k_0-1} \Delta_k + \sum_{i=0}^{\infty} \sum_{k=k_i}^{k_{i+1}-1} \Delta_k \leq \sum_{k=0}^{k_0-1} \Delta_k + \left(\frac{\sigma_3}{\sigma_2(1-\sigma_2)}\right) \sum_{i=0}^{\infty} \Delta_{k_i}$$

bekommt und daraus die Konvergenz von dessen Reihe. Aufgrund von $\sigma_1 \|p_k\| \leq \Delta_{k+1}$ für alle $k \in \mathbb{N}$ gilt $\sum_{k=0}^{\infty} \|p_k\| < \infty$ und damit die implizierte Aussage in (3.5.11). $\surd$

- Jetzt zeigen wir, daß aus (3.5.13) Aussage (3.5.14) folgt. Aufgrund der Definition des Cauchy-Schritts gilt $\|\pi_k(\alpha_k)\| \leq \mu_1 \Delta_k$ für alle $k \in \mathbb{N}$ und da $(\Delta_k)_{k\in\mathbb{N}}$ gegen Null konvergiert muß wegen unserer Annahme (3.5.13) die Folge $(\alpha_k)_{k\in S}$ auch gegen Null konvergieren. Für alle hinreichend großen $k \in S$ trifft daher die zweite Möglichkeit der Wahl von α_k in (3.3.4) zu, also $\alpha_k \geq \gamma_2 \hat{\alpha}_k$, wobei $\hat{\alpha}_k$ (3.3.5) erfüllt. Wir nehmen nun an, daß $\hat{\alpha}_k$ für unendlich viele $k \in S$ Forderung (3.3.5) erfüllt, das heißt es gelte

$$f_k(\pi_k(\hat{\alpha}_k)) - f(x_k) > \mu_0 \nabla f(x_k)^T \pi_k(\hat{\alpha}_k)$$

für unendlich viele $k \in S$. Aus der Charakterisierung der Projektion folgt

$$-\nabla f(x_k)^T \pi_k(\alpha) \geq \frac{\|\pi_k(\alpha)\|^2}{\alpha} \quad \text{für alle } \alpha > 0,$$

wie in (3.3.6) schon gezeigt wurde. Zusammen erhält man dann

$$\begin{aligned}
\|B_k\| &> \frac{1}{2}\frac{\pi_k(\hat{\alpha}_k)^T B_k \pi_k(\hat{\alpha}_k)}{\|\pi_k(\hat{\alpha}_k)\|^2} \\
&= \frac{f_k(\pi_k(\hat{\alpha}_k)) - f(x_k) - \nabla f(x_k)^T \pi_k(\hat{\alpha}_k)}{\|\pi_k(\hat{\alpha}_k)\|^2} \\
&> -(1-\mu_0)\frac{\nabla f(x_k)^T \pi_k(\hat{\alpha}_k)}{\|\pi_k(\hat{\alpha}_k)\|^2} \\
&> -\frac{1}{2}\frac{\nabla f(x_k)^T \pi_k(\hat{\alpha}_k)}{\|\pi_k(\hat{\alpha}_k)\|^2} \\
&> -\mu_0\frac{\nabla f(x_k)^T \pi_k(\hat{\alpha}_k)}{\|\pi_k(\hat{\alpha}_k)\|^2} \\
&\geq \frac{\mu_0}{\hat{\alpha}_k}
\end{aligned}$$

für unendlich viele $k \in S$. Da aber die Folge $(B_k)_{k\in\mathbb{N}}$ beschränkt ist und $\alpha_k \geq \gamma_2\hat{\alpha}_k$ gilt, ist eine Teilfolge von $(\alpha_k)_{k\in S}$ durch eine Schranke, die größer als Null ist, nach unten beschränkt. Das widerspricht unserer früheren Beobachtung, daß $(\alpha_k)_{k\in S}$ gegen Null konvergiert. Daher muß $\hat{\alpha}_k$ für alle hinreichend großen $k \in S$ die zweite Forderung in (3.3.5) erfüllen. Korollar 2.1.5 impliziert nun

$$\|\pi_k(\alpha_k)\| \geq \min\{\gamma_2, 1\}\,\|\pi_k(\hat{\alpha}_k)\| \geq \min\{\gamma_2, 1\}\mu_1\Delta_k$$

für alle hinreichend großen $k \in S$ und damit ist (3.5.14) gezeigt. $\surd$

- Als nächstes zeigen wir, daß aus (3.5.13) die Aussage (3.5.15) folgt. Wir nehmen an, es existiert eine Folge $(k_i)_{i\in\mathbb{N}} \subseteq \mathbb{N}$ von nicht erfolgreichen Iterationsindizes und ein $\delta > 0$, so daß

$$\frac{\|\pi_{k_i}(\alpha_{k_i})\|}{\alpha_{k_i}} \geq \delta \qquad \text{für alle } i \in \mathbb{N}$$

ist. Aus Abschätzung (3.4.8) ergibt sich zusammen mit den Tatsachen, daß die Folge $(B_k)_{k\in\mathbb{N}}$ beschränkt ist und die Folge $(\Delta_k)_{k\in\mathbb{N}}$ gegen Null konvergiert, daß

$$f(x_{k_i}) - f_{k_i}(p_{k_i}) \geq {\mu_0}^2\zeta\delta\Delta_{k_i} \geq {\mu_0}^2\zeta\delta\frac{\|p_{k_i}\|}{\mu_1}$$

für alle hinreichend großen $i \in \mathbb{N}$ ist. Aufgrund von $\sum_{k=0}^{\infty}\|p_k\| < \infty$ konvergiert die Folge $(x_k)_{k\in\mathbb{N}}$ und die Folge $(p_k)_{k\in\mathbb{N}}$ konvergiert gegen

Null. Wir definieren die Folge $(\varepsilon_i)_{i\in\mathbb{N}}$ durch

$$\varepsilon_i := \frac{\left| f(x_{k_i}+p_{k_i}) - f(x_{k_i}) - \nabla f(x_{k_i})^T p_{k_i} \right|}{\|p_{k_i}\|}.$$

Dann erhält man aus der stetigen Differenzierbarkeit von ∇f auf einer offenen Obermenge von M

$$\begin{aligned}
\varepsilon_i &= \left| \int_0^1 (\nabla f(x_{k_i}+tp_{k_i}) - \nabla f(x_{k_i}))^T \left(\frac{p_{k_i}}{\|p_{k_i}\|}\right) dt \right| \\
&\leq \int_0^1 \|\nabla f(x_{k_i}+tp_{k_i}) - \nabla f(x_{k_i})\| \, dt \\
&\leq \int_0^1 (\|\nabla f(x_{k_i}+tp_{k_i}) - \nabla f(x^*)\| + \|\nabla f(x^*) - \nabla f(x_{k_i})\|) \, dt
\end{aligned}$$

für alle $i \in \mathbb{N}$ und damit die Konvergenz von $(\varepsilon_i)_{i\in\mathbb{N}}$ gegen Null. Hieraus ergibt sich

$$\begin{aligned}
&|f(x_{k_i}+p_{k_i}) - f_{k_i}(p_{k_i}))| \\
=&|f(x_{k_i}+p_{k_i}) - f(x_{k_i}) - (f_{k_i}(p_{k_i}) - f(x_{k_i}))| \\
=&\left| f(x_{k_i}+p_{k_i}) - f(x_{k_i}) - (\nabla f(x_{k_i})^T p_{k_i} + \frac{1}{2} {p_{k_i}}^T B_{k_i} p_{k_i}) \right| \\
\leq&\left| f(x_{k_i}+p_{k_i}) - f(x_{k_i}) - \nabla f(x_{k_i})^T p_{k_i} \right| + \frac{1}{2} \left| {p_{k_i}}^T B_{k_i} p_{k_i} \right| \\
\leq&\varepsilon_i \|p_{k_i}\| + \frac{1}{2} \|B_{k_i}\| \, \|p_{k_i}\|^2 \qquad \text{für alle } i \in \mathbb{N}.
\end{aligned}$$

Nun ist

$$\begin{aligned}
|\chi_{k_i} - 1| &= \left| \frac{f(x_{k_i}) - f(x_{k_i}+p_{k_i}) - (f(x_{k_i}) - f_{k_i}(p_{k_i}))}{f(x_{k_i}) - f_{k_i}(p_{k_i})} \right| \\
&= \left| \frac{f(x_{k_i}+p_{k_i}) - f_{k_i}(p_{k_i})}{f(x_{k_i}) - f_{k_i}(p_{k_i})} \right| \\
&\leq \frac{\varepsilon_i \|p_{k_i}\| + \frac{1}{2} \|B_{k_i}\| \, \|p_{k_i}\|^2}{{\mu_0}^2 \zeta\delta \left(\frac{\|p_{k_i}\|}{\mu_1} \right)} \\
&\leq \frac{\mu_1\varepsilon_i + \frac{1}{2}\mu_1 \|B_{k_i}\| \, \|p_{k_i}\|}{{\mu_0}^2 \zeta\delta}
\end{aligned}$$

für alle hinreichend großen $i \in \mathbb{N}$ und man sieht, daß $\lim_{i\to\infty} \chi_{k_i} = 1$ ist. Das heißt, daß der Iterationsindex k_i für alle hinreichend großen $i \in \mathbb{N}$ erfolgreich ist, was unserer Wahl der Folge $(k_i)_{i\in\mathbb{N}}$ widerspricht. $\surd$

- Der Beweis, daß aus (3.5.13) Aussage (3.5.15) folgt, verläuft ähnlich. Wir nehmen an, es existiert eine Folge $(k_i)_{i\in\mathbb{N}} \subseteq \mathbb{N}$ von nicht erfolgreichen Iterationsindizes und ein $\delta > 0$, so daß

$$\frac{\|\pi_{k_i}(\alpha_{k_i})\|}{\|p_{k_i}\|} \geq \delta \qquad \text{für alle } i \in \mathbb{N}$$

ist. Beachtet man

$$\frac{\|\pi_{k_i}(\alpha_{k_i})\|}{\alpha_{k_i}} \geq \frac{\|\pi_{k_i}(\alpha_{k_i})\|}{\|p_{k_i}\|}\frac{\|p_{k_i}\|}{\gamma_3} \geq \frac{\delta}{\gamma_3}\|p_{k_i}\| \qquad \text{für alle } i \in \mathbb{N},$$

so erhält man zusammen mit (3.4.8)

$$\begin{aligned}
&f(x_{k_i}) - f_{k_i}(p_{k_i}) \\
\geq&{\mu_0}^2\zeta\left(\frac{\|\pi_{k_i}(\alpha_{k_i})\|}{\alpha_{k_i}}\right)\min\left\{\Delta_{k_i}, \frac{1}{1+\|B_{k_i}\|}\left(\frac{\|\pi_{k_i}(\alpha_{k_i})\|}{\alpha_{k_i}}\right)\right\} \\
\geq&{\mu_0}^2\zeta\left(\frac{\delta}{\gamma_3}\|p_{k_i}\|\right)\min\left\{\frac{\|p_{k_i}\|}{\mu_1}, \frac{1}{1+\sup_{i\in\mathbb{N}}\|B_{k_i}\|}\left(\frac{\delta}{\gamma_3}\|p_{k_i}\|\right)\right\} \\
\geq&{\mu_0}^2\zeta\frac{\delta}{\gamma_3}\min\left\{\frac{1}{\mu_1}, \frac{\delta}{\gamma_3(1+\sup_{i\in\mathbb{N}}\|B_{k_i}\|)}\right\}\|p_{k_i}\|^2 \\
\geq&\hat{\delta}\|p_{k_i}\|^2 \qquad \text{für alle } i \in \mathbb{N},
\end{aligned}$$

wobei

$$0 < \hat{\delta} \leq {\mu_0}^2\zeta\frac{\delta}{\gamma_3}\min\left\{\frac{1}{\mu_1}, \frac{\delta}{\gamma_3(1+\sup_{i\in\mathbb{N}}\|B_{k_i}\|)}\right\}$$

gilt. Da außerdem die Gültigkeit von (3.5.11) gefordert wird, erhält man aus

$$\begin{aligned}
|\chi_{k_i} - 1| &= \left|\frac{f(x_{k_i}) - f(x_{k_i}+p_{k_i}) - (f(x_{k_i}) - f_{k_i}(p_{k_i}))}{f(x_{k_i}) - f_{k_i}(p_{k_i})}\right| \\
&= \left|\frac{f(x_{k_i}+p_{k_i}) - f_{k_i}(p_{k_i})}{f(x_{k_i}) - f_{k_i}(p_{k_i})}\right| \\
&\leq \left|\frac{f(x_{k_i}+p_{k_i}) - f_{k_i}(p_{k_i})}{\hat{\delta}\|p_{k_i}\|^2}\right|,
\end{aligned}$$

daß $\lim_{i\to\infty}\chi_{k_i} = 1$ gilt. Wie im Beweisabschnitt davor heißt das, daß der Iterationsindex k_i für alle hinreichend großen $i \in \mathbb{N}$ erfolgreich ist, was unserer Wahl der Folge $(k_i)_{i\in\mathbb{N}}$ widerspricht. Mit diesem Teilabschnitt ist auch der ganze Beweis beendet. $\surd$

■

Jetzt kommen wir zu den Hauptresultaten der Konvergenzanalyse des Trust-Region-Verfahrens. Der erste der beiden Sätze liefert unter Annahme der Existenz eines Häufungspunktes der Folge $(x_k)_{k\in\mathbb{N}}$, die durch das Trust-Region-Verfahren erzeugt wird, das Resultat, daß dieser Häufungspunkt stationär ist. Man findet allerdings unter schwächeren Voraussetzungen im Gegensatz zum zweiten der beiden Sätze nur eine Teilfolge von $(x_k)_{k\in\mathbb{N}}$, deren Iterationsindizes nicht notwendigerweise erfolgreich sind.

Satz 3.5.4 *Sei $f : \mathbb{R}^n \to \mathbb{R}$ eine auf einer offenen Obermenge von M stetig differenzierbare Abbildung. Das Trust-Region-Verfahren breche nicht vorzeitig ab und sei $(x_k)_{k\in\mathbb{N}}$ die Folge, die durch dieses Verfahren erzeugt wird. Sei weiterhin $(B_k)_{k\in\mathbb{N}}$ beschränkt und $(x_k)_{k\in\mathbb{N}}$ besitze einen Häufungspunkt $x^* \in M$. Dann existiert eine Teilfolge $(x_{k_i})_{i\in\mathbb{N}}$ von Iterationspunkten, die gegen x^* konvergiert und für die*

$$\lim_{i\to\infty} \left\|\nabla_M f(x_{k_i}^C)\right\| = 0$$

gilt. Ferner konvergiert auch $(x_{k_i}^C)_{i\in\mathbb{N}}$ gegen x^* und daher ist x^* ein stationärer Punkt der Optimierungsaufgabe (P).

Beweis. Sei $(g_j)_{j\in\mathbb{N}} \subseteq \mathbb{N}$ irgendeine Folge von Indizes, so daß $(x_{g_j})_{j\in\mathbb{N}}$ gegen x^* konvergiert. Wenn

$$\liminf_{j\to\infty} \frac{\left\|\pi_{g_j}(\alpha_{g_j})\right\|}{\alpha_{g_j}} = 0$$

ist, sind wir fertig. Denn da $x_k^C - x_k = \pi_k(\alpha_k)$ und $\alpha_k \le \gamma_3$ für alle $k \in \mathbb{N}$ gilt, haben wir auch

$$\liminf_{j\to\infty} \left\|x_{g_j}^C - x_{g_j}\right\| = 0.$$

Daher konvergiert eine Teilfolge von $(x_{g_j}^C)_{j\in\mathbb{N}}$ ebenfalls gegen x^* und mit Lemma 3.5.1 erhält man das gewünschte Resultat. Wir nehmen also an, es existiert ein $\varepsilon > 0$, so daß

$$\frac{\left\|\pi_{g_j}(\alpha_{g_j})\right\|}{\alpha_{g_j}} \ge \varepsilon \quad \text{für alle } j \in \mathbb{N}$$

ist. Satz 3.4.2 garantiert, daß es für jedes $0 < \delta < \varepsilon$ eine Teilfolge $(h_j)_{j\in\mathbb{N}} \subseteq \mathbb{N}$ von Iterationsindizes gibt, für die

$$\begin{aligned} \frac{\|\pi_k(\alpha_k)\|}{\alpha_k} &\ge \delta \quad \text{für alle } k \in Z := \{\nu \in \mathbb{N} \mid g_j \le \nu < h_j \text{ für alle } j \in \mathbb{N}\} \\ \text{und } \frac{\left\|\pi_{h_j}(\alpha_{h_j})\right\|}{\alpha_{h_j}} &< \delta \quad \text{für alle } j \in \mathbb{N} \end{aligned}$$

ist. Daher impliziert Abschätzung (3.4.9)

$$f(x_k) - f(x_{k+1}) \geq \rho_0 {\mu_0}^2 \zeta \delta \min \left\{ \Delta_k, \frac{\delta}{1 + \|B_k\|} \right\}$$

für alle erfolgreichen Indizes $k \in Z$. Da weiterhin $(B_k)_{k \in \mathbb{N}}$ beschränkt ist und die Folge $(f(x_k))_{k \in \mathbb{N}}$ konvergiert, konvergiert $(\Delta_k)_{k \in \{\nu \in Z | \nu \text{ erfolgreich }\}}$ gegen Null. Insbesondere ist

$$\Delta_k < \frac{\delta}{\|B_k\|}$$

für alle hinreichend großen erfolgreichen Iterationsindizes $k \in Z$. Da darüberhinaus $\|p_k\| \leq \mu_1 \Delta_k$ für alle $k \in \mathbb{N}$ ist und bei einem nicht erfolgreichen Iterationsindex k der Punkt x_k nicht verändert wird, gilt

$$f(x_k) - f(x_{k+1}) \geq \frac{\rho_0 {\mu_0}^2 \zeta \delta}{\mu_1} \|x_{k+1} - x_k\|$$

sogar für alle hinreichend großen $k \in Z$. Zusammengesetzt ergibt das

$$f(x_{g_j}) - f(x_{h_j}) \geq \frac{\rho_0 {\mu_0}^2 \zeta \delta}{\mu_1} \left\|x_{g_j} - x_{h_j}\right\| \qquad \text{für alle } j \in \mathbb{N}$$

und damit erhält man die Konvergenz von $(x_{h_j})_{j \in \mathbb{N}}$ gegen x^*. Insbesondere haben wir gerade gezeigt, daß es für jedes $0 < \delta < \varepsilon$ eine Teilfolge $(h_j)_{j \in \mathbb{N}} \subseteq \mathbb{N}$ von Iterationsindizes gibt, für die

$$\frac{\left\|\pi_{h_j}(\alpha_{h_j})\right\|}{\alpha_{h_j}} < \delta \quad \text{und} \quad \left\|x_{h_j} - x^*\right\| < \delta \quad \text{für alle } j \in \mathbb{N}$$

gilt, wobei man die eben konstruierte Folge $(h_j)_{j \in \mathbb{N}}$ möglicherweise um eine endliche Anzahl an Folgengliedern kürzen muß. Daher gibt es eine Teilfolge $(k_j)_{j \in \mathbb{N}} \subseteq \mathbb{N}$ von Iterationsindizes, so daß $(x_{k_j})_{j \in \mathbb{N}}$ gegen x^* konvergiert und

$$\lim_{j \to \infty} \frac{\left\|\pi_{k_j}(\alpha_{k_j})\right\|}{\alpha_{k_j}} = 0$$

ist. Wie oben konvergiert aufgrund der Identität $x_k^C - x_k = \pi_k(\alpha_k)$ und der Schranke $\alpha_k \leq \gamma_3$ für alle $k \in \mathbb{N}$ die Folge $(x_{k_j}^C)_{j \in \mathbb{N}}$ ebenfalls gegen x^*, so daß Lemma 3.5.1 die Konvergenz von $(\nabla_M f(x_{k_j}^C))_{j \in \mathbb{N}}$ gegen Null liefert. Nach Lemma 2.2.5 ist die Funktion $\|\nabla_M f(\cdot)\| : M \to \mathbb{R}$, definiert durch $x \mapsto \|\nabla_M f(x)\|$, nach unten halbstetig. Aus diesem Grund gilt $\|\nabla_M f(x^*)\| = 0$ und aus Satz 2.2.3, Teil (iii) folgt, daß x^* ein stationärer Punkt der Optimierungsaufgabe (P) ist. ■

Das zweite Hauptresultat dagegen liefert eine Teilfolge von $(x_k)_{k\in\mathbb{N}}$, deren dazugehörige Iterationsindizes alle erfolgreich sind. Allerdings muß die Implikation (3.5.11) vorausgesetzt werden und im Gegensatz zu Theorem 5.5 aus Burke et al., 1990 meiner Ansicht nach auch die gleichmäßige Stetigkeit von ∇f auf der Niveaumenge $L(M; x_0)$. Das hat den Grund, daß diese Voraussetzung auch in Satz 3.5.3 gemacht wird, auf den im Beweis zu diesem Satz zurückgegriffen wird. Der Beweis ist leider in weiten Teilen identisch zum Beweis von Satz 3.5.4, er ließe sich aber auch nicht durch Ausgliederung eines Lemmas vereinfachen.

Satz 3.5.5 *Sei $f : \mathbb{R}^n \to \mathbb{R}$ eine auf einer offenen Obermenge von M stetig differenzierbare Abbildung. Das Trust-Region-Verfahren breche nicht vorzeitig ab und sei $(x_k)_{k\in\mathbb{N}}$ die Folge, die durch dieses Verfahren erzeugt wird. Sei weiterhin ∇f auf der Niveaumenge $L(M; x_0)$ gleichmäßig stetig, die Folge $(B_k)_{k\in\mathbb{N}}$ beschränkt und es gelte die Implikation (3.5.11). Sei $x^* \in M$ ein Häufungspunkt der Folge $(x_k)_{k\in\mathbb{N}}$. Dann existiert eine Teilfolge $(k_i)_{i\in\mathbb{N}} \subseteq \mathbb{N}$ von erfolgreichen Iterationsindizes, so daß $(x_{k_i})_{i\in\mathbb{N}}$ gegen x^* konvergiert und für die*

$$\lim_{i\to\infty} \left\|\nabla_M f(x^C_{k_i})\right\| = 0$$

gilt. Ferner konvergiert auch $(x^C_{k_i})_{i\in\mathbb{N}}$ gegen x^* und daher ist x^* ein stationärer Punkt der Optimierungsaufgabe (P).

Beweis. Sei $(g_j)_{j\in\mathbb{N}} \subseteq \mathbb{N}$ irgendeine Folge von Iterationsindizes, so daß $(x_{g_j})_{j\in\mathbb{N}}$ gegen x^* konvergiert. Die nicht erfolgreichen Iterationsindizes g_j erhöhe man solange um 1, bis man entweder auf einen erfolgreichen Iterationsindex oder auf g_{j+1} trifft, wobei in diesem Fall g_j aus der Folge der Indizes entfernt werden kann, da $x_{g_{j+1}} = x_{g_j}$ ist. Da nach Satz 3.4.6 die Anzahl der erfolgreichen Iterationsindizes unendlich ist, ist diese Vorgehensweise möglich. Auf diese Weise erhält man eine Folge $(x_{g_j})_{j\in\mathbb{N}}$, aus der alle aufeinanderfolgenden Duplikate von Punkten x_{g_j} entfernt wurden, die Indizes der restlichen Punkte aber soweit erhöht worden sind, daß all diese erfolgreich sind, aber ohne die Punkte selbst zu verändern. Wenn

$$\lim_{j\to\infty} \frac{\left\|\pi_{g_j}(\alpha_{g_j})\right\|}{\alpha_{g_j}} = 0$$

ist, sind wir fertig. Denn da $x^C_k - x_k = \pi_k(\alpha_k)$ und $\alpha_k \leq \gamma_3$ für alle $k \in \mathbb{N}$ gilt, haben wir auch

$$\lim_{j\to\infty} \left\|x^C_{g_j} - x_{g_j}\right\| = 0.$$

Daher konvergiert eine Teilfolge von $(x^C_{g_j})_{j\in\mathbb{N}}$ ebenfalls gegen x^* und mit Lemma 3.5.1 erhält man das gewünschte Resultat. Wir nehmen also an, es existiert

ein $\varepsilon > 0$, so daß

$$\frac{\left\|\pi_{g_j}(\alpha_{g_j})\right\|}{\alpha_{g_j}} \geq \varepsilon \quad \text{für alle } j \in \mathbb{N}$$

ist. Satz 3.5.3 garantiert, daß es für jedes $\delta \in (0, \varepsilon)$ eine Teilfolge $(h_j)_{j\in\mathbb{N}} \subseteq \mathbb{N}$ von erfolgreichen Iterationsindizes gibt, für die

$$\begin{aligned} \frac{\|\pi_k(\alpha_k)\|}{\alpha_k} &\geq \delta \quad \text{für alle } k \in Z := \{\nu \in \mathbb{N} \mid g_j \leq \nu < h_j \text{ für alle } j \in \mathbb{N}\} \\ \text{und } \frac{\left\|\pi_{h_j}(\alpha_{h_j})\right\|}{\alpha_{h_j}} &< \delta \quad \text{für alle } j \in \mathbb{N} \end{aligned}$$

ist. Daher impliziert Abschätzung (3.4.9)

$$f(x_k) - f(x_{k+1}) \geq \rho_0 {\mu_0}^2 \zeta \delta \min\left\{\Delta_k, \frac{\delta}{1 + \|B_k\|}\right\}$$

für alle erfolgreichen Indizes $k \in Z$. Da weiterhin $(B_k)_{k\in\mathbb{N}}$ beschränkt ist und die Folge $(f(x_k))_{k\in\mathbb{N}}$ konvergiert, konvergiert $(\Delta_k)_{k\in\{\nu\in Z \mid \nu \text{ erfolgreich }\}}$ gegen Null. Insbesondere ist

$$\Delta_k < \frac{\delta}{\|B_k\|}$$

für alle hinreichend großen erfolgreichen Iterationsindizes $k \in Z$. Da darüberhinaus $\|p_k\| \leq \mu_1 \Delta_k$ für alle $k \in \mathbb{N}$ ist und bei einem nicht erfolgreichen Iterationsindex k der Punkt x_k nicht verändert wird, gilt

$$f(x_k) - f(x_{k+1}) \geq \frac{\rho_0 {\mu_0}^2 \zeta \delta}{\mu_1} \|x_{k+1} - x_k\|$$

sogar für alle hinreichend großen $k \in Z$. Zusammengesetzt ergibt das

$$f(x_{g_j}) - f(x_{h_j}) \geq \frac{\rho_0 {\mu_0}^2 \zeta \delta}{\mu_1} \left\|x_{g_j} - x_{h_j}\right\| \quad \text{für alle } j \in \mathbb{N}$$

und damit erhält man die Konvergenz von $(x_{h_j})_{j\in\mathbb{N}}$ gegen x^*. Insbesondere haben wir gerade gezeigt, daß es für jedes $\delta \in (0, \varepsilon)$ eine Teilfolge $(h_j)_{j\in\mathbb{N}} \subseteq \mathbb{N}$ von erfolgreichen Iterationsindizes gibt, für die

$$\frac{\left\|\pi_{h_j}(\alpha_{h_j})\right\|}{\alpha_{h_j}} < \delta \quad \text{und} \quad \left\|x_{h_j} - x^*\right\| < \delta \quad \text{für alle } j \in \mathbb{N}$$

gilt, wobei man die eben konstruierte Folge $(h_j)_{j\in\mathbb{N}}$ möglicherweise um eine endliche Anzahl an Folgengliedern kürzen muß. Daher gibt es eine Teilfolge $(k_j)_{j\in\mathbb{N}} \subseteq \mathbb{N}$ von erfolgreichen Iterationsindizes, so daß $(x_{k_j})_{j\in\mathbb{N}}$ gegen x^* konvergiert und

$$\lim_{j\to\infty} \frac{\left\|\pi_{k_j}(\alpha_{k_j})\right\|}{\alpha_{k_j}} = 0$$

ist. Wie oben konvergiert aufgrund der Identität $x_k^C - x_k = \pi_k(\alpha_k)$ und der Schranke $\alpha_k \leq \gamma_3$ für alle $k \in \mathbb{N}$ die Folge $(x_{k_j}^C)_{j\in\mathbb{N}}$ ebenfalls gegen x^*, so daß Lemma 3.5.1 die Konvergenz von $(\nabla_M f(x_{k_j}^C))_{j\in\mathbb{N}}$ gegen Null liefert. Nach Lemma 2.2.5 ist die Funktion $\|\nabla_M f(\cdot)\| : M \to \mathbb{R}$, definiert durch $x \mapsto \|\nabla_M f(x)\|$, nach unten halbstetig. Aus diesem Grund gilt $\|\nabla_M f(x^*)\| = 0$ und aus Satz 2.2.3, Teil (iii) folgt, daß x^* ein stationärer Punkt der Optimierungsaufgabe (P) ist. ■

Kapitel 4

Geometrische Aspekte

In diesem Kapitel werden gewisse geometrische Aspekte des in Kapitel 3 behandelten Trust-Region-Verfahrens dargestellt. Im ersten Abschnitt werden die exponierte Seitenfläche und der Normalkegel eingeführt, die in einer gewissen Beziehung zueinander und zum Tangentialkegel stehen. Wir werden schließlich zeigen, daß unter den Voraussetzungen, die wir auch bei den Konvergenzsätzen des Trust-Region-Verfahrens gemacht haben, letztendlich alle Iterationspunkte des Trust-Region-Verfahrens in der durch den negativen Gradienten an einem stationären Punkt exponierten Seitenfläche liegen. Dieses Resultat wird benötigt, da wir keine strikte Komplementarität an einem stationären Punkt fordern. Im zweiten Abschnitt behandeln wir Zwischenschritte und das Verfahren, mit dem man theoretisch bestimmte Bedingungen an diese erfüllen kann. Zwischenschritte sind für die Konvergenzanalyse des Newton-Verfahrens wichtig, da man mit ihnen weitere Forderungen an die Iterationsschritte p_k für $k \in \mathbb{N}$ stellen kann. So gelingt es, mit dem Newton-Verfahren hochdimensionale Probleme zu lösen und gleichzeitig lineare oder superlineare Konvergenz zu gewährleisten.

4.1 Exponierte Seitenfläche und Normalkegel

Als erstes definieren wir die exponierte Seitenfläche.

Definition 4.1.1 (Exponierte Seitenfläche) Sei $\Omega \subseteq \mathbb{R}^n$ eine nichtleere, abgeschlossene und konvexe Menge und $d \in \mathbb{R}^n$. Die durch d *exponierte Seitenfläche* von Ω ist durch

$$E(\Omega; d) := \{z \in \Omega \mid d^T z \geq d^T y \text{ für alle } y \in \Omega\}$$

definiert.

Eine anschauliche Darstellung der exponierten Seitenfläche leiten wir jetzt für die Fälle her, die uns im weiteren Verlauf tatsächlich interessieren werden. Bezüglich des Trust-Region-Verfahrens und des Newton-Verfahrens ist nämlich nur der Fall $d := -\nabla f(x)$ für $x \in M$ interessant. Wir unterscheiden nun die Optimierungsaufgaben (P) und (Q) und betrachten zuerst den spezielleren Fall, das heißt die Optimierungsaufgabe (Q). Sei dazu $x \in [l,u]$ beliebig gewählt. Durch einfaches Ausrechnen des Skalarprodukts $-\nabla f(x)^T z$ für ein beliebiges $z \in [l,u]$ erhält man

$$E([l,u]; -\nabla f(x)) = \left\{ z \in [l,u] \;\middle|\; \begin{array}{l} z_i = l_i \text{ falls } \partial_i f(x) > 0 \text{ und} \\ z_i = u_i \text{ falls } \partial_i f(x) < 0 \text{ für alle } 1 \leq i \leq n \end{array} \right\}$$

und man sieht, daß es sich um eine Hyperfläche des Quaders $[l,u]$ handelt. Ein Bildbeispiel dazu findet man in Lin and Moré, 1999. Hinsichtlich des allgemeineren Falls sei $x^* \in M$ nun ein stationärer Punkt der Optimierungsaufgabe (P). Dann existieren nach dem Satz von Karush-Kuhn-Tucker Lagrange-Multiplikatoren $\{\lambda_1, \ldots, \lambda_m\} \subset \mathbb{R}$, für die

$$\nabla f(x^*) = \sum_{i=1}^{m} \lambda_i c_i$$

gilt, wobei

$$\lambda_i \begin{cases} \geq 0 & \text{falls } {c_i}^T x^* = l_i, \\ = 0 & \text{falls } l_i < {c_i}^T x^* < u_i, \\ \leq 0 & \text{falls } {c_i}^T x^* = u_i \end{cases}$$

ist. Aus der Gleichung $-\nabla f(x^*)^T z = \sum_{i=1}^{m} -\lambda_i ({c_i}^T z)$ läßt sich dann

$$E(M; -\nabla f(x^*)) = \left\{ z \in M \;\middle|\; \begin{array}{l} {c_i}^T z = l_i \text{ falls } \lambda_i > 0 \text{ und} \\ {c_i}^T z = u_i \text{ falls } \lambda_i < 0 \text{ für alle } 1 \leq i \leq m \end{array} \right\}$$

ablesen und man sieht auch hier, daß es sich um eine Hyperfläche des Polyeders M handelt. Eine für uns wichtige Eigenschaft der durch den negativen Gradienten exponierten Seitenfläche ergibt sich sofort im Zusammenhang mit stationären Punkten.

Satz 4.1.2 *Sei $\Omega \subseteq \mathbb{R}^n$ eine nichtleere, abgeschlossene und konvexe Menge und $f : \mathbb{R}^n \to \mathbb{R}$ eine auf einer offenen Obermenge von Ω stetig differenzierbare Abbildung. Dann ist $x \in \Omega$ genau dann ein stationärer Punkt der Aufgabe, f auf Ω zu minimieren, wenn $x \in E(\Omega; -\nabla f(x))$ ist.*

Beweis. Sei $x \in \Omega$ ein stationärer Punkt der Aufgabe, f auf Ω zu minimieren. Das ist genau dann der Fall, wenn

$$\nabla f(x)^T (y - x) \geq 0 \quad \text{für alle } y \in \Omega$$

ist. Das heißt aber, daß

$$-\nabla f(x)^T x \geq -\nabla f(x)^T y \quad \text{für alle } y \in \Omega$$

gilt und diese Ungleichung ist äquivalent zu der Aussage

$$x \in E(\Omega; -\nabla f(x)).$$

■

In Abschnitt 3.2 wurde die Abbruchbedingung für das Trust-Region-Verfahren beschrieben, die auf stationäre Punkte testet. Bei der Optimierungsaufgabe (Q) ist ein beliebiger Punkt $x \in [l, u]$ genau dann stationär, wenn $\nabla_{[l,u]} f(x) = 0$ ist und das ist genau dann der Fall, wenn für die i-te Komponente des Gradienten $\nabla f(x)$

$$\partial_i f(x) := [\nabla f(x)]_i \begin{cases} \geq 0 & \text{falls } x_i = l_i, \\ = 0 & \text{falls } l_i < x_i < u_i, \\ \leq 0 & \text{falls } x_i = u_i \end{cases}$$

gilt. Das ist aber nach der oben gewonnenen Charakterisierung von $E([l, u]; -\nabla f(x))$ genau dann der Fall, wenn

$$x \in E([l, u]; -\nabla f(x))$$

ist. Satz 4.1.2 stellt somit ein bereits bekanntes Ergebnis in neuer Form dar. Den Beweis des nächsten Lemmas können wir hier leider nicht präsentieren, da er zu viele geometrische Aspekte beinhaltet, die den Rahmen dieser Arbeit sprengen würden. Er basiert auf der Tatsache, daß ein Polyeder nur eine endliche Anzahl an Seitenflächen besitzt, deren genaue Definition dann ebenfalls noch ausstehen würde.

Lemma 4.1.3 *Sei $d^* \in \mathbb{R}^n$. Dann existiert eine Umgebung $U(d^*) \subseteq \mathbb{R}^n$ von d^*, so daß*

$$E(M; d) \subseteq E(M; d^*) \quad \textit{für alle } d \in U(d^*)$$

gilt.

Beweis. Siehe den Beweis von Theorem 3.1 in Burke and Moré, 1994. ■

Als nächstes folgt die Definition des Normalkegels.

Definition 4.1.4 (Normalkegel) Sei $\Omega \subseteq \mathbb{R}^n$ eine nichtleere, abgeschlossene und konvexe Menge und $x \in \mathbb{R}^n$. Der *Normalkegel* an Ω in x ist durch

$$N(\Omega; x) := \{z \in \mathbb{R}^n \mid z^T(y - x) \leq 0 \text{ für alle } y \in \Omega\}$$

definiert.

Bemerkung 4.1.5

- Der Normalkegel ist der Polarkegel des Tangentialkegels. Um dies zu sehen, beachte man, daß der Normalkegel auch in der Form
$$N(\Omega; x) = \{z \in \mathbb{R}^n \mid z^T p \leq 0 \text{ für alle } p \in T(\Omega; x)\}$$
geschrieben werden kann. Um das zu zeigen, muß nur noch die Richtung
$$\begin{aligned}&\{z \in \mathbb{R}^n \mid z^T(y - x) \leq 0 \text{ für alle } y \in \Omega\}\\ \subseteq &\{z \in \mathbb{R}^n \mid z^T p \leq 0 \text{ für alle } p \in T(\Omega; x)\}\end{aligned}$$
gezeigt werden, da die andere Richtung offensichtlich ist. Sei also $z \in \mathbb{R}^n$ so gewählt, daß $z^T(y-x) \leq 0$ für alle $y \in \Omega$ ist und sei $p \in T(\Omega; x)$ beliebig vorgegeben. Da $T(\Omega; x)$ auch gleichzeitig der Abschluß der zulässigen Richtungen an Ω in x ist, existiert eine Folge $(p_k)_{k\in\mathbb{N}} \subset \Omega$ von zulässigen Richtungen an Ω in x, die gegen p konvergiert. Zu jedem p_k existiert ein $t_k > 0$, so daß $x + t_k p_k \in \Omega$ ist und damit gilt
$$z^T(t_k p_k) = z^T(x + t_k p_k - x) \leq 0.$$
Division durch t_k und Übergang zum Grenzwert für $k \to \infty$ liefern das gewünschte Resultat. So ergibt sich die Beziehung
$$N(\Omega; x) = T(\Omega; x)^{\perp}$$
nach der Definition des Polarkegels.

- Weiterhin ergibt sich eine Beziehung zur exponierten Seitenfläche. Seien dazu $x, d \in \mathbb{R}^n$ beliebig vorgegeben. Dann gilt
$$x \in E(\Omega; d)$$
genau dann, wenn
$$d \in N(\Omega; x)$$
ist. Zum Beweis dazu sei $x \in E(\Omega; d)$. Das ist genau dann der Fall, wenn $d^T x \geq d^T y$ für alle $y \in \Omega$ ist, also $d^T(y - x) \leq 0$ für alle $y \in \Omega$ gilt. Das ist aber nach der Definition des Normalkegels genau dann wahr, wenn $d \in N(\Omega; x)$ ist.

Sei $x^* \in M$ jetzt ein stationärer Punkt der Optimierungsaufgabe (P). Nach Dunn, 1987 ist x^* ein *regulärer stationärer Punkt*, falls
$$-\nabla f(x^*) \in \mathrm{ri}(N(M; x^*))$$
gilt, wobei $\mathrm{ri}(G)$ einer beliebigen Menge $G \subseteq \mathbb{R}^n$ das relative Innere von G sei. In Burke and Moré, 1994 wird gezeigt, daß x^* genau dann regulär ist, wenn
$$x^* \in \mathrm{ri}(E(M; -\nabla f(x^*)))$$

gilt. Sei $x^* \in M$ nun ein regulärer stationärer Punkt der Optimierungsaufgabe (P). Mit Hilfe der Darstellung von $\nabla f(x^*)$ durch Lagrange-Multiplikatoren $\{\lambda_1, \ldots, \lambda_m\} \subset \mathbb{R}$ als

$$\nabla f(x^*) = \sum_{i=1}^{m} \lambda_i c_i,$$

wobei

$$\lambda_i \begin{cases} \geq 0 & \text{falls } {c_i}^T x^* = l_i, \\ = 0 & \text{falls } l_i < {c_i}^T x^* < u_i, \\ \leq 0 & \text{falls } {c_i}^T x^* = u_i \end{cases}$$

gilt, und der Gleichung

$$E(M; -\nabla f(x^*)) = \left\{ z \in M \;\middle|\; \begin{array}{l} {c_i}^T z = l_i \text{ falls } \lambda_i > 0 \text{ und} \\ {c_i}^T z = u_i \text{ falls } \lambda_i < 0 \text{ für alle } 1 \leq i \leq m \end{array} \right\}$$

erhalten wir $l_i < {c_i}^T x^* < u_i$, falls $\lambda_i = 0$ und $1 \leq i \leq m$ ist. Ließt man diese Aussage anders, so können wir die Bedingung an die Lagrange-Multiplikatoren zu

$$\lambda_i \begin{cases} > 0 & \text{falls } {c_i}^T x^* = l_i, \\ = 0 & \text{falls } l_i < {c_i}^T x^* < u_i, \\ < 0 & \text{falls } {c_i}^T x^* = u_i \end{cases}$$

verschärfen und das entspricht der Definition der strikten Komplementarität. Ein stationärer Punkt $x^* \in M$ der Optimierungsaufgabe (P) ist also genau dann ein regulärer stationärer Punkt, wenn an diesem Punkt die Bedingung der strikten Komplementarität erfüllt ist. Als nächstes benötigen wir eine weitere Definition.

Definition 4.1.6 (Aktive Restriktionen) Seien $x, y \in M$. Die Menge der *aktiven Restriktionen* an M in x ist definiert durch

$$I(M; x) := \{i \in \{1, \ldots, m\} \mid {c_i}^T x = l_i \text{ oder } {c_i}^T x = u_i\}.$$

Die Inklusion

$$I(M; x) \sqsubset I(M; y)$$

ist definiert durch

$$I_l(M; x) \subseteq I_l(M; y) \quad \text{und} \quad I_u(M; x) \subseteq I_u(M; y),$$

wobei

$$\begin{aligned} I_l(M; x) :=& \{i \in \{1, \ldots, m\} \mid {c_i}^T x_i = l_i\} \text{ und} \\ I_u(M; x) :=& \{i \in \{1, \ldots, m\} \mid {c_i}^T x_i = u_i\} \end{aligned}$$

sei.

Bemerkung 4.1.7 Die Definition einer neuen Inklusion $I(x) \sqsubset I(y)$ ist deshalb nötig, da andernfalls nicht zwischen unterer und oberer Grenze unterschieden würde. Diese Unterscheidung wird im weiteren Verlauf noch wichtig sein.

Bisher wurde globale und superlineare Konvergenz von Verfahren zur Lösung der Optimierungsaufgabe (P) nur unter der Voraussetzung der strikten Komplementarität an einem stationären Punkt $x^* \in M$ erzielt und die Beweise beruhen darauf, daß sich die Menge der aktiven Restriktionen letztendlich nicht mehr verändert, wenn die Folge der Iterationspunkte gegen x^* konvergiert. Diese Vorgehensweise ist in unserem Fall nicht möglich, da wir strikte Komplementarität nicht voraussetzen. Wir werden jetzt das Resultat entwickeln, daß unter bestimmten Voraussetzungen

$$x_k, x_k^C, x_k + p_k \in E(M; -\nabla f(x^*))$$

für alle hinreichend großen $k \in \mathbb{N}$ sind. Dieses Resultat impliziert im Fall von strikter Komplementarität

$$x_k, x_k^C, x_k + p_k \in \mathrm{ri}(E(M; -\nabla f(x^*)))$$

für alle hinreichend großen $k \in \mathbb{N}$ und damit auch, daß sich die Menge der aktiven Restriktionen letztendlich nicht mehr verändert. Mit Hilfe der Definition des Normalkegels und der dazugehörigen Bemerkung 4.1.5 erhält man sofort aus Lemma 4.1.3 das nächste Resultat, es stammt aus dem dritten Abschnitt von Burke and Moré, 1994.

Korollar 4.1.8 *Seien $(x_k)_{k\in\mathbb{N}} \subset M$ eine Folge und $(d_k)_{k\in\mathbb{N}} \subset \mathbb{R}^n$ eine gegen $d^* \in \mathbb{R}^n$ konvergente Folge mit $d_k \in N(M; x_k)$ für alle $k \in \mathbb{N}$. Dann existiert ein $\hat{k} \in \mathbb{N}$, so daß*

$$x_k \in E(M; d^*) \quad \text{für alle } k \geq \hat{k}$$

ist.

Beweis. Nach Lemma 4.1.3 existiert eine Umgebung $U(d^*) \subseteq \mathbb{R}^n$ von d^*, so daß

$$E(M; d) \subseteq E(M; d^*) \quad \text{für alle } d \in U(d^*)$$

gilt. Da weiterhin $(d_k)_{k\in\mathbb{N}}$ gegen d^* konvergiert, existiert ein $\hat{k} \in \mathbb{N}$, so daß $d_k \in U(d^*)$ für alle $k \geq \hat{k}$ ist. Sei nun $k \in \mathbb{N}$ beliebig gewählt. Dann ist nach Bemerkung 4.1.5

$$d_k \in N(M; x_k)$$

genau dann, wenn

$$x_k \in E(M; d_k)$$

ist. Daraus erhält man insgesamt

$$x_k \in E(M; d_k) \subseteq E(M; d^*) \quad \text{für alle } k \geq \hat{k}.$$

■

Das nächste Lemma findet man im vierten Abschnitt von Burke and Moré, 1994.

Lemma 4.1.9 *Sei $f : \mathbb{R}^n \to \mathbb{R}$ eine auf einer offenen Obermenge von M stetig differenzierbare Abbildung und $(x_k)_{k\in\mathbb{N}} \subset M$ eine gegen $x^* \in M$ konvergente Folge. Dann gilt*

$$\lim_{k\to\infty} \|\nabla_M f(x_k)\| = 0$$

genau dann, wenn ein $\hat{k} \in \mathbb{N}$ existiert, so daß $x_k \in E(M; -\nabla f(x^*))$ für alle $k \geq \hat{k}$ ist.

Beweis.

$\Rightarrow$ **:** Es gelte

$$\lim_{k\to\infty} \|\nabla_M f(x_k)\| = 0.$$

Da der Normalkegel der Polarkegel des Tangentialkegels ist, gilt mit Lemma 2.1.9

$$\begin{aligned} -\nabla f(x_k) =& \Pi_{T(M;x_k)}(-\nabla f(x_k)) + \Pi_{N(M;x_k)}(-\nabla f(x_k)) \\ =& \nabla_M f(x_k) + \Pi_{N(M;x_k)}(-\nabla f(x_k)) \quad \text{für alle } k \in \mathbb{N}. \end{aligned}$$

Daher konvergiert $(\Pi_{N(M;x_k)}(-\nabla f(x_k)))_{k\in\mathbb{N}}$ gegen $-\nabla f(x^*)$ und mit Korollar 4.1.8 erhält man ein $\hat{k} \in \mathbb{N}$, so daß

$$x_k \in E(M; -\nabla f(x^*)) \quad \text{für alle } k \geq \hat{k}$$

ist. √

$\Leftarrow$ **:** Es existiere ein $\hat{k} \in \mathbb{N}$, so daß $x_k \in E(M; -\nabla f(x^*))$ für alle $k \geq \hat{k}$ ist. Wieder nach Bemerkung 4.1.5 gilt

$$-\nabla f(x^*) \in N(M; x_k) \quad \text{für alle } k \geq \hat{k}.$$

Nach Lemma 2.1.9 und da der Normalkegel der Polarkegel des Tangentialkegels ist, haben wir wie eben

$$-\Pi_{T(M;x_k)}(-\nabla f(x_k)) = \Pi_{N(M;x_k)}(-\nabla f(x_k)) + \nabla f(x_k) \quad \text{für alle } k \in \mathbb{N}$$

und somit

$$\begin{aligned}\left\|\Pi_{T(M;x_k)}(-\nabla f(x_k))\right\| &= \left\|\Pi_{N(M;x_k)}(-\nabla f(x_k)) + \nabla f(x_k)\right\| \\ &= \inf_{y\in N(M;x_k)} \|y + \nabla f(x_k)\| \qquad \text{für alle } k \in \mathbb{N}.\end{aligned}$$

Zusammen erhalten wir

$$\left\|\Pi_{T(M;x_k)}(-\nabla f(x_k))\right\| \leq \|-\nabla f(x^*) + \nabla f(x_k)\| \qquad \text{für alle } k \geq \hat{k}$$

und der Limes für $k \to \infty$ liefert das gewünschte Resultat. √

■

Es gilt

$$\Pi_\Omega(x) \in E(\Omega; x - \Pi_\Omega(x)) \qquad \text{für alle } x \in \mathbb{R}^n, \tag{4.1.1}$$

denn nach Lemma 2.1.2 ist $(\Pi_\Omega(x) - x)^T(y - \Pi_\Omega(x)) \geq 0$ für alle $x \in \mathbb{R}^n$ und alle $y \in \Omega$ und so

$$(x - \Pi_\Omega(x))^T \Pi_\Omega(x) \geq (x - \Pi_\Omega(x))^T y$$

für alle $x \in \mathbb{R}^n$ und alle $y \in \Omega$. Diese Aussage wird im nächsten Satz benutzt.

Satz 4.1.10 *Sei $f : \mathbb{R}^n \to \mathbb{R}$ eine auf einer offenen Obermenge von M stetig differenzierbare Abbildung und $(x_k)_{k\in\mathbb{N}} \subset M$ eine gegen $x^* \in M$ konvergente Folge. Weiterhin existiere ein $\hat{k} \in \mathbb{N}$, so daß $x_k \in E(M; -\nabla f(x^*))$ für alle $k \geq \hat{k}$ ist. Dann existiert ein $\tilde{k} \in \mathbb{N}$, so daß*

$$\Pi_M(x_k - \alpha\nabla f(x_k)) \in E(M; -\nabla f(x^*)) \quad \textit{für alle } \alpha > 0 \quad \textit{und} \quad \textit{alle } k \geq \tilde{k}$$

ist.

Beweis.

- Sei $\alpha > 0$ beliebig gewählt. Wir behaupten, daß

$$\left\|\frac{\Pi_M(x_k - \alpha\nabla f(x_k)) - x_k}{\alpha}\right\| \leq \|\nabla_M f(x_k)\| \qquad \text{für alle } k \in \mathbb{N} \tag{4.1.2}$$

gilt. Sei dazu $k \in \mathbb{N}$ beliebig vorgegeben. Aus Ungleichung (3.3.6) ergibt sich

$$\frac{\|\Pi_M(x_k - \alpha\nabla f(x_k)) - x_k\|^2}{\alpha} \leq -\nabla f(x_k)^T(\Pi_M(x_k - \alpha\nabla f(x_k)) - x_k)$$

und aus Satz 2.2.3, Teil (ii) folgt

$$-\nabla f(x_k)^T p \leq \|\nabla_M f(x_k)\| \, \|p\|$$

für alle zulässigen Richtungen $p \in \mathbb{R}^n$ an M in x_k, speziell also

$$\begin{aligned} &-\nabla f(x_k)^T(\Pi_M(x_k - \alpha\nabla f(x_k)) - x_k) \\ &\leq \|\nabla_M f(x_k)\| \, \|\Pi_M(x_k - \alpha\nabla f(x_k)) - x_k\| \, . \end{aligned}$$

Aus beiden Ungleichungen zusammen erhält man

$$\frac{\|\Pi_M(x_k - \alpha\nabla f(x_k)) - x_k\|^2}{\alpha} \leq \|\nabla_M f(x_k)\| \, \|\Pi_M(x_k - \alpha\nabla f(x_k)) - x_k\|$$

und damit das gewünschte Ergebnis. $\surd$

- Da die Folge $(x_k)_{k\in\mathbb{N}}$ gegen $x^* \in M$ konvergiert und $x_k \in E(M; -\nabla f(x^*))$ für alle $k \geq \hat{k}$ ist, implizieren Ungleichung (4.1.2) und Lemma 4.1.9

$$0 \leq \lim_{k\to\infty} \left\| \frac{\Pi_M(x_k - \alpha\nabla f(x_k)) - x_k}{\alpha} \right\| \leq \lim_{k\to\infty} \|\nabla_M f(x_k)\| = 0.$$

für alle $\alpha > 0$. Definiert man nun die Folge $(d_k)_{k\in\mathbb{N}}$ von Funktionen $d_k : \mathbb{R}_+ \to \mathbb{R}^n$ mit $k \in \mathbb{N}$ durch

$$d_k(\alpha) := -\nabla f(x_k) - \frac{\Pi_M(x_k - \alpha\nabla f(x_k)) - x_k}{\alpha},$$

so bekommt man

$$\lim_{k\to\infty} d_k(\alpha) = -\nabla f(x^*) \quad \text{für alle } \alpha > 0,$$

wobei die Konvergenz gleichmäßig in α ist. Unabhängig davon haben wir nach Aussage (4.1.1)

$$\begin{aligned} \Pi_M(x_k - \alpha\nabla f(x_k)) \in& E(M; x_k - \alpha\nabla f(x_k) - \Pi_M(x_k - \alpha\nabla f(x_k))) \\ =& E(M; \alpha d_k(\alpha)) \\ =& E(M; d_k(\alpha)) \quad \text{für alle } \alpha > 0 \quad \text{und alle } k \in \mathbb{N}. \end{aligned}$$

Nach Lemma 4.1.3 existiert ein $\tilde{k} \in \mathbb{N}$, so daß

$$E(M; d_k(\alpha)) \subseteq E(M; -\nabla f(x^*)) \quad \text{für alle } \alpha > 0 \quad \text{und alle } k \geq \tilde{k}$$

gilt und insgesamt erhalten wir

$$\Pi_M(x_k - \alpha\nabla f(x_k)) - x_k \in E(M; -\nabla f(x^*)) \quad \text{für alle } \alpha > 0$$

und alle $k \geq \tilde{k}$. $\surd$

■

Um die wahre Aussagekraft des Satzes deutlich zu machen und zu nutzen formulieren wir das folgende Korollar. Man benötigt nämlich keine ganze Folge $(x_k)_{k\in\mathbb{N}} \subset M$, von der alle hinreichend großen Folgenglieder in $E(M; -\nabla f(x^*))$ liegen, sondern lediglich einen Punkt, der in einer hinreichend kleinen Umgebung um x^* liegen muß.

Korollar 4.1.11 *Sei $f : \mathbb{R}^n \to \mathbb{R}$ eine auf einer offenen Obermenge von M stetig differenzierbare Abbildung und $x^* \in M$ beliebig vorgegeben. Dann existiert eine Umgebung $U(x^*) \subseteq \mathbb{R}^n$ von x^*, so daß für alle $x \in U(x^*)$ aus $x \in E(M; -\nabla f(x^*))$ folgt, daß*

$$\Pi_M(x - \alpha\nabla f(x)) \in E(M; -\nabla f(x^*)) \quad \textit{für alle } \alpha > 0$$

ist.

Beweis. Der Beweis erfolgt durch Widerspruch. Angenommen, für jede Umgebung $U(x^*) \subseteq \mathbb{R}^n$ von x^* existiert ein $x \in U(x^*)$ und ein $\alpha > 0$, für das zwar $x \in E(M; -\nabla f(x^*))$ ist, aber

$$\Pi_M(x - \alpha\nabla f(x)) \notin E(M; -\nabla f(x^*))$$

gilt. Durch die Wahl immer kleinerer Umgebungen von x^* erhält man so eine Folge $(x_k)_{k\in\mathbb{N}} \subset M$ mit Grenzwert x^* und eine Folge $(\alpha_k)_{k\in\mathbb{N}} \subset \mathbb{R}_+$, für die

$$x_k \in E(M; -\nabla f(x^*)) \quad \text{für alle } k \in \mathbb{N}$$

ist, aber für die auch

$$\Pi_M(x_k - \alpha_k\nabla f(x_k)) \notin E(M; -\nabla f(x^*)) \quad \text{für alle } k \in \mathbb{N}$$

gilt. Das ist offensichtlich ein Widerspruch zu Satz 4.1.10. ■

Sei $x^* \in M$ ein stationärer Punkt der Optimierungsaufgabe (P). Dann gilt die Implikation

(4.1.3)

$$x \in E(M; -\nabla f(x^*)) \text{ und } I(M; x) \sqsubset I(M; y) \quad \Longrightarrow \quad y \in E(M; -\nabla f(x^*))$$

für alle $x, y \in M$. Diese Aussage folgt direkt aus der am Anfang dieses Abschnitts gewonnenen Charakterisierung der exponierten Seitenfläche $E(M; -\nabla f(x^*))$. Der nächste Satz benutzt diese Tatsache. Wir wollen zeigen, daß unter gewissen Voraussetzungen letztendlich alle Iterationspunkte des Trust-Region-Verfahrens in der durch den negativen Gradienten eines stationären Punktes $x^* \in M$ exponierten Seitenfläche $E(M; -\nabla f(x^*))$ enthalten sind. Satz 3.5.5 und Satz 4.1.10 liefern mit einer weiteren Voraussetzung ein solches Ergebnis.

Satz 4.1.12 *Sei $f : \mathbb{R}^n \to \mathbb{R}$ eine auf einer offenen Obermenge von M stetig differenzierbare Abbildung. Das Trust-Region-Verfahren breche nicht vorzeitig ab und sei $(x_k)_{k\in\mathbb{N}}$ die Folge, die durch dieses Verfahren erzeugt wird. Sei weiterhin ∇f auf der Niveaumenge $L(M; x_0)$ gleichmäßig stetig, die Folge $(B_k)_{k\in\mathbb{N}}$ beschränkt und es gelte die Implikation (3.5.11). Die Folge $(x_k)_{k\in\mathbb{N}}$ konvergiere gegen $x^* \in M$ und es gelte die Inklusion*

$$I(M; x_k^C) \sqsubset I(M; x_k + p_k) \quad \text{für alle } k \in \mathbb{N}. \tag{4.1.4}$$

Dann existiert ein Iterationsindex $\hat{k} \in \mathbb{N}$, so daß

$$x_k, x_k^C, x_k + p_k \in E(M; -\nabla f(x^*)) \quad \text{für alle } k \geq \hat{k}$$

sind.

Beweis. Satz 3.5.5 zeigt, daß eine Teilfolge $(k_i)_{i\in\mathbb{N}} \subseteq \mathbb{N}$ von erfolgreichen Iterationsindizes existiert, so daß $(x_{k_i}^C)_{i\in\mathbb{N}}$ gegen x^* konvergiert und

$$\lim_{i\to\infty} \left\|\nabla_M f(x_{k_i}^C)\right\| = 0$$

ist. Lemma 4.1.9 sagt aus, daß ein $\hat{i} \in \mathbb{N}$ existiert, so daß

$$x_{k_i}^C \in E(M; -\nabla f(x^*)) \quad \text{für alle } i \geq \hat{i}$$

gilt. Da jeder Index k_i für $i \in \mathbb{N}$ erfolgreich ist, folgt aus der Forderung (4.1.4) und der Implikation (4.1.3)

$$x_{k_i+1} = x_{k_i} + p_{k_i} \in E(M; -\nabla f(x^*)) \quad \text{für alle } i \geq \hat{i}.$$

Nach Korollar 4.1.11 existiert eine Umgebung $U(x^*) \subseteq \mathbb{R}^n$ von x^*, so daß für alle $x \in U(x^*)$ aus $x \in E(M; -\nabla f(x^*))$ folgt, daß

$$\Pi_M(x - \alpha\nabla f(x)) \in E(M; -\nabla f(x^*)) \quad \text{für alle } \alpha > 0$$

ist. Da die Folge $(x_k)_{k\in\mathbb{N}}$ gegen x^* konvergiert, existiert ein $\tilde{k} \in \mathbb{N}$, so daß $x_k \in U(x^*)$ für alle $k \geq \tilde{k}$ ist. Wähle nun ein $k_i \geq \tilde{k}$ mit $i \geq \hat{i}$ und setze $\hat{k} := k_i + 1$. Wir zeigen jetzt die Aussage des Satzes mit Induktion über k. Es gilt offenbar

$$x_{\hat{k}} \in E(M; -\nabla f(x^*))$$

und da $\hat{k} \geq \tilde{k}$ ist, haben wir auch

$$x_{\hat{k}}^C \in E(M; -\nabla f(x^*))$$

und so wieder mit Forderung (4.1.4)

$$x_{\hat{k}} + p_{\hat{k}} \in E(M; -\nabla f(x^*)),$$

so daß mit dieser Wahl von $\hat{k}$ die Induktionsverankerung gezeigt ist. Sei nun

$$x_k, x_k^C, x_k + p_k \in E(M; -\nabla f(x^*))$$

für einen beliebigen Iterationsindex $k \geq \hat{k}$ erfüllt. Dann ist offensichtlich

$$x_{k+1} \in E(M; -\nabla f(x^*)),$$

denn es gilt nach der Iterationsvorschrift des Trust-Region-Verfahrens $x_{k+1} = x_k + p_k$, wenn der Iterationsindex k erfolgreich ist und $x_{k+1} = x_k$ andernfalls. Da $x_{k+1} \in U(x^*)$ ist, erhalten wir

$$x_{k+1}^C \in E(M; -\nabla f(x^*))$$

und erneut mit Forderung (4.1.4)

$$x_{k+1} + p_{k+1} \in E(M; -\nabla f(x^*)),$$

womit wir auch den Induktionsschritt gezeigt haben. ■

4.2 Zwischenschritte

Bisher war für die Konvergenzanalyse des Trust-Region-Verfahrens nicht wichtig, ob oder wie man einen weiteren Schritt p_k außer dem Cauchy-Schritt p_k^C zur Lösung des Trust-Region-Subproblems bestimmt. Um allerdings lineare oder superlineare Konvergenz des in Kapitel 3 vorgestellten Newton-Verfahrens zu gewährleisten, müssen für die Iterationsschritte p_k mit $k \in \mathbb{N}$ zusätzliche Bedingungen gelten, die der Cauchy-Schritt nicht notwendigerweise erfüllt. Da unser Augenmerk insbesondere auf der Lösung hochdimensionaler Optimierungsaufgaben liegt, ist es für uns wichtig, diese Forderungen iterativ zu erfüllen. Deshalb definieren wir als erstes den Begriff der Zwischenschritte.

Definition 4.2.1 (Zwischenschritte) Sei $f : \mathbb{R}^n \to \mathbb{R}$ eine auf einer offenen Obermenge von M stetig differenzierbare Abbildung und $x_k \in M$ mit $k \in \mathbb{N}$ ein Iterationspunkt des Trust-Region-Verfahrens. Sei weiterhin $r \in \mathbb{N} \setminus \{0\}$ eine Konstante. Dann definiert man die *Zwischenschritte*

$$\{x_k^1 := x_k^C, \ldots, x_k^{r+1}\} \subset M$$

zum Iterationsindex k dadurch, daß

$$\begin{aligned} \left\| x_k^j - x_k \right\| &\leq \mu_1 \Delta_k \quad \text{für alle } 1 \leq j \leq r+1 \text{ und} \\ I(M; x_k^j) &\sqsubset I(M; x_k^{j+1}) \quad \text{für alle } 1 \leq j \leq r \end{aligned} \tag{4.2.5}$$

sowie

(4.2.6)

$$q_k(x_k^{j+1}) \leq q_k(x_k^j) + \mu_0 \min\{\nabla q_k(x_k^j)^T (x_k^{j+1} - x_k^j), 0\} \quad \text{für alle } 1 \leq j \leq r$$

gelten, wobei die Funktion $q_k : \mathbb{R}^n \to \mathbb{R}$ durch

$$q_k(x) := \nabla f(x_k)^T (x - x_k) + \frac{1}{2}(x - x_k)^T B_k (x - x_k)$$

definiert ist.

Bemerkung 4.2.2

- Die Konstante r setzt nur eine obere Schranke für die Anzahl der Zwischenschritte. Nimmt man nämlich für ein beliebiges $1 \leq j \leq r$
$$x_k^{j+1} = x_k^j,$$
so sind die Forderungen (4.2.5) und (4.2.6) offenbar erfüllt. Wir werden aber sehen, daß die beiden Bedingungen nicht nur mit dieser trivialen Wahl der Zwischenschritte erfüllt sind.

- Bevor wir zeigen, daß tatsächlich auch nichttriviale Zwischenschritte existieren, sollten wir diese in das Trust-Region-Verfahren integrieren. Wir erinnern uns, daß die Forderung (3.1.1) im Trust-Region-Verfahren an den Schritt p_k
$$f_k(p_k) - f(x_k) \leq \mu_0 (f_k(p_k^C) - f(x_k))$$
mit $x_k + p_k \in M$ und $\|p_k\| \leq \mu_1 \Delta_k$ lautet. Setzen wir
$$p_k := x_k^{r+1} - x_k,$$
so ist aufgrund von Forderung (4.2.5) an die Zwischenschritte $x_k + p_k \in M$ und $\|p_k\| \leq \mu_1 \Delta_k$. Insbesondere aber folgt aus der Ungleichung (4.2.6)
$$q_k(x_k^{j+1}) \leq q_k(x_k^j) \quad \text{für alle } 1 \leq j \leq r$$
und da $q_k(x) = f_k(x - x_k) - f(x_k)$ ist, erhält man damit
$$f_k(x_k^{j+1} - x_k) - f(x_k) \leq f_k(x_k^j - x_k) - f(x_k) \quad \text{für alle } 1 \leq j \leq r$$
und schließlich
$$f_k(p_k) - f(x_k) = f_k(x_k^{r+1} - x_k) - f(x_k) \leq f_k(p_k^C) - f(x_k),$$
womit auch Forderung (3.1.1) erfüllt ist. Die Zwischenschritte stehen also im Einklang mit der angegebenen Wahl eines weiteren Schritts im Trust-Region-Verfahren.

- Der letzte Punkt zeigt, daß durch die Zwischenschritte der Funktionswert der Modellfunktion im Punkt p_k niedriger ist als in p_k^C, die Trust-Region-Grenze $\mu_1\Delta_k$ nicht verletzt wird und die schon im Cauchy-Punkt aktiven Restriktionen weiterhin aktiv sind. Die letzte dieser Eigenschaften des Schritts p_k ermöglicht erst die Anwendung von Satz 4.1.12.

Als nächstes zeigen wir die Existenz von nichttrivialen Zwischenschritten. Der Grund dafür ist, daß wir später noch eine weitere Forderung an den letzten Zwischenschritt stellen werden, die nichttriviale Zwischenschritte erfordert.

Satz 4.2.3 *Sei $f : \mathbb{R}^n \to \mathbb{R}$ eine auf einer offenen Obermenge von M stetig differenzierbare Abbildung und $x_k \in M$ mit $k \in \mathbb{N}$ ein Iterationspunkt des Trust-Region-Verfahrens. Seien bereits die Zwischenschritte*

$$\{x_k^1 := x_k^C, \ldots, x_k^j\} \subset M$$

mit $1 \leq j \leq r$ berechnet und sei x_k^j kein stationärer Punkt der Optimierungsaufgabe, q_k auf M zu minimieren. Es gelte weiterhin

$$\left\|x_k^j - x_k\right\| < \mu_1\Delta_k.$$

Dann existiert ein $x_k^j \neq x_k^{j+1} \in M$, so daß

$$\left\|x_k^{j+1} - x_k\right\| \leq \mu_1\Delta_k \quad \text{und} \quad I(M; x_k^j) \sqsubset I(M; x_k^{j+1})$$

sowie

$$q_k(x_k^{j+1}) \leq q_k(x_k^j) + \mu_0 \min\{\nabla q_k(x_k^j)^T(x_k^{j+1} - x_k^j), 0\}$$

gilt.

Beweis. Der Beweis erfolgt in zwei Teilen. Im ersten Teil werden wir eine zu (4.2.6) hinreichende Bedingung finden, die leichter zu handhaben ist und die auch Bedingung (4.2.5) vereinfacht. Im zweiten Teil werden wir dann sehen, wie man diese vereinfachten Forderungen erfüllen kann.

- Die Schwierigkeit des Beweises liegt einzig darin, in jedem Zwischenschritt die aktiven Restriktionen zu erhalten. Wir müssen daher einen Untervektorraum einführen, der nur Vektoren enthält, die senkrecht auf allen Richtungen aktiver Restriktionen stehen. Seien dazu $\{v_1, \ldots, v_s\} \subset \mathbb{R}^n$ mit $1 \leq s \leq n$ eine Orthonormalbasis des Vektorraums

  $$V_k^j := \{v \in \mathbb{R}^n \mid c_i{}^T v = 0 \text{ für alle } i \in I(M; x_k^j)\}$$

  und

  $$A_k^j := \begin{pmatrix} v_1 & \ldots & v_s \end{pmatrix} \in \mathbb{R}^{n \times s}$$

die Matrix, die die Vektoren dieser Orthonormalbasis als Spalten hat. Es ist dann s die Dimension dieses Vektorraums. Jetzt kann man die am Anfang gestellten Forderungen an x_k^{j+1} anders ausdrücken, und zwar suchen wir ein $w \in \mathbb{R}^s$, für das

$$\left\| x_k^j + A_k^j w - x_k \right\| \leq \mu_1 \Delta_k \tag{4.2.7}$$

und

$$q_k(x_k^j + A_k^j w) - q_k(x_k^j) \leq \mu_0 \min\{({A_k^j}^T \nabla q_k(x_k^j))^T w, 0\}$$

gilt. Denn durch die Wahl der Matrix A_k^j stellen wir sicher, daß die aktiven Restriktionen in x_k^j auch im nächsten Zwischenschritt aktiv sind. Definiert man jetzt noch die Funktion $q : \mathbb{R}^s \to \mathbb{R}$ durch

$$q(w) := q_k(x_k^j + A_k^j w) - q_k(x_k^j),$$

so formt man die zweite der beiden Forderungen in

$$q(w) \leq \mu_0 \min\{\nabla q(0)^T w, 0\} \tag{4.2.8}$$

um. Denn es ist

$$\begin{aligned}
q(w) =& q_k(x_k^j + A_k^j w) - q_k(x_k^j) \\
=& \nabla f(x_k)^T (x_k^j + A_k^j w - x_k - (x_k^j - x_k)) \\
&+ \frac{1}{2}(x_k^j + A_k^j w - x_k)^T B_k (x_k^j + A_k^j w - x_k) \\
&- \frac{1}{2}(x_k^j - x_k)^T B_k (x_k^j - x_k) \\
=& \nabla f(x_k)^T A_k^j w \\
&+ \frac{1}{2}((x_k^j - x_k)^T B_k A_k^j w + w^T {A_k^j}^T B_k (x_k^j - x_k) + w^T {A_k^j}^T B_k A_k^j w) \\
=& ({A_k^j}^T \nabla f(x_k))^T w + ({A_k^j}^T B_k (x_k^j - x_k))^T w + \frac{1}{2} w^T {A_k^j}^T B_k A_k^j w
\end{aligned}$$

für alle $w \in \mathbb{R}^s$. Daraus ergibt sich sofort

$$\begin{aligned}
\nabla q(w) =& {A_k^j}^T \nabla f(x_k) + {A_k^j}^T B_k (x_k^j - x_k) + {A_k^j}^T B_k A_k^j w \\
=& {A_k^j}^T (\nabla f(x_k) + B_k (x_k^j + A_k^j w - x_k)) \\
=& {A_k^j}^T \nabla q_k(x_k^j) + {A_k^j}^T B_k A_k^j w \quad \text{für alle } w \in \mathbb{R}^s
\end{aligned}$$

und so

$$\nabla q(0) = {A_k^j}^T \nabla q_k(x_k^j).$$

$\surd$

- Nun müssen wir nur noch zeigen, daß ein $0 \neq w \in \mathbb{R}^s$ existiert, das (4.2.7) und (4.2.8) erfüllt. Die Vorgehensweise entspricht dabei der, die wir beim Beweis der Existenz des Cauchy-Schritts kennengelernt haben, siehe dazu Abschnitt 3.3. Wir definieren die Menge

$$M_k^j := \{A_k^{j\,T}(x - x_k^j) \mid x \in M\} \subseteq \mathbb{R}^s$$

und die Funktion $\pi : \mathbb{R}_+ \cup \{0\} \to \mathbb{R}^s$ durch

$$\pi(\alpha) := \Pi_{M_k^j}(-\alpha \nabla q(0)).$$

Wir zeigen, daß die Forderungen (4.2.7) und (4.2.8) für alle hinreichend kleinen $\alpha > 0$ gelten, sollten wir uns bei x_k^j nicht an einem stationären Punkt der Optimierungsaufgabe, q_k auf M zu minimieren, befinden. Diese Forderung müssen wir jetzt noch umformen. Sei Null dazu ein stationärer Punkt der Optimierungsaufgabe, q auf M_k^j zu minimieren, das heißt, es gelte

$$\nabla q(0)^T y \geq 0 \quad \text{für alle } y \in M_k^j.$$

Daraus folgt mit der Definition der Funktion q_k sofort

$$(A_k^{j\,T} \nabla q_k(x_k^j))^T A_k^{j\,T}(x - x_k^j) \geq 0 \quad \text{für alle } x \in M$$

und da die Matrix A_k^j die Vektoren einer Orthonormalbasis von V_k^j als Spaltenvektoren besitzt, erhalten wir weiter

$$\nabla q_k(x_k^j)^T (x - x_k^j) \geq 0 \quad \text{für alle } x \in M.$$

Das heißt aber, daß x_k^j ein stationärer Punkt der Optimierungsaufgabe, q_k auf M zu minimieren, ist. Die Forderung des Satzes impliziert daher, daß Null kein stationärer Punkt der Optimierungsaufgabe, q auf M_k^j zu minimieren, ist und demnach erhalten wir mit Satz 2.2.3, Teil (iii)

$$\nabla_{M_k^j} q(0) \neq 0.$$

Als nächstes bemerken, daß aus Korollar 2.1.13

$$\begin{aligned}\lim_{0<\alpha\to 0} \frac{\pi(\alpha)}{\alpha} &= \lim_{0<\alpha\to 0} \frac{\Pi_{M_k^j}(-\alpha \nabla q(0))}{\alpha} \\ &= \Pi_{T(M_k^j;0)}(-\nabla q(0)) \\ &= \nabla_{M_k^j} q(0)\end{aligned}$$

folgt. Mit Satz 2.2.3, Teil (i) bekommt man

$$\begin{aligned}
&\lim_{0<\alpha\to 0} \frac{q(\pi(\alpha))}{\alpha} \\
=&\lim_{0<\alpha\to 0} \left(\left(A_k^{j\,T} \nabla f(x_k)\right)^T \left(\frac{\pi(\alpha)}{\alpha}\right) + \left(A_k^{j\,T} B_k (x_k^j - x_k)\right)^T \left(\frac{\pi(\alpha)}{\alpha}\right)\right) \\
&+ \lim_{0<\alpha\to 0} \left(\frac{1}{2} \left(\frac{\pi(\alpha)}{\alpha}\right)^T A_k^{j\,T} B_k A_k^j \pi(\alpha) \right) \\
=&\left(A_k^{j\,T} (\nabla f(x_k) + B_k(x_k^j - x_k))\right)^T \left(\lim_{0<\alpha\to 0} \frac{\pi(\alpha)}{\alpha}\right) \\
&+ \frac{1}{2} \left(\lim_{0<\alpha\to 0} \frac{\pi(\alpha)}{\alpha}\right)^T A_k^{j\,T} B_k A_k^j \underbrace{\left(\lim_{0<\alpha\to 0} \pi(\alpha)\right)}_{=0} \\
=&\nabla q(0)^T \nabla_{M_k^j} q(0) \\
=&-\left\| \nabla_{M_k^j} q(0) \right\|^2 \\
<&0.
\end{aligned}$$

Da nach Voraussetzung $0 < \mu_0 < \frac{1}{2}$ ist, existiert ein $\varepsilon > 0$, so daß für alle α mit $0 < \alpha < \varepsilon$

$$\frac{q(\pi(\alpha))}{\alpha} < 2\mu_0 \nabla q(0)^T \nabla_{M_k^j} q(0)$$

gilt. Außerdem existiert wegen der oben gezeigten Darstellung für $\lim_{0<\alpha\to 0} \frac{\pi(\alpha)}{\alpha}$ ein $\delta > 0$, so daß für alle α mit $0 < \alpha < \delta$

$$\left| \nabla q(0)^T \nabla_{M_k^j} q(0) - \nabla q(0)^T \left(\frac{\pi(\alpha)}{\alpha}\right) \right| < -\nabla q(0)^T \nabla_{M_k^j} q(0)$$

ist. Insbesondere gilt also für diese α die Ungleichung

$$2\nabla q(0)^T \nabla_{M_k^j} q(0) < \nabla q(0)^T \left(\frac{\pi(\alpha)}{\alpha}\right)$$

und zusammen mit der ersten Ungleichung erhält man für alle α mit $0 < \alpha < \min\{\varepsilon, \delta\}$

$$\begin{aligned}
q(\pi(\alpha)) <& 2\alpha\mu_0 \nabla q(0)^T \nabla_{M_k^j} q(0) \\
<& \mu_0 \nabla q(0)^T \pi(\alpha).
\end{aligned}$$

Setzt man nun $w := \pi(\alpha)$, so erfüllen alle hinreichend kleinen α mit $0 < \alpha < \min\{\varepsilon, \delta\}$ Forderung (4.2.7)

$$\left\| x_k^j + A_k^j \pi(\alpha) - x_k \right\| \le \mu_1 \Delta_k,$$

da wir $\left\|x_k^j - x_k\right\| < \mu_1 \Delta_k$ vorausgesetzt haben, sowie Forderung (4.2.8)

$$q(\pi(\alpha)) \leq \mu_0 \nabla q(0)^T \pi(\alpha) = \mu_0 \min\{\nabla q(0)^T \pi(\alpha), 0\},$$

da $\nabla q(0)^T \pi(\alpha) < 0$ für alle $\alpha > 0$ gilt, wie man analog zur Ungleichungskette (3.3.6) zeigt. $\surd$

■

Kapitel 5

Newton-Verfahren

In diesem Kapitel wird das Newton-Verfahren erläutert und dessen Konvergenzeigenschaften beschrieben. Das Newton-Verfahren benötigt als Zielfunktion eine zweimal stetig differenzierbare Funktion. Paßt man die Voraussetzungen der aussagekräftigsten Konvergenzsätze des Trust-Region-Verfahrens ein wenig an die zweimal stetige Differenzierbarkeit an und fügt man als wesentliche Voraussetzung hinzu, daß die Hessesche Matrix eines Häufungspunktes der Folge, die durch das Newton-Verfahren erzeugt wird, auf einem speziellen Untervektorraum des $\mathbb{R}^n$ positiv definit ist, so konvergiert die ganze Folge gegen einen stationären Punkt der Optimierungsaufgabe (P). Im weiteren Verlauf werden die Zwischenschritte aus Abschnitt 4.2 benutzt und eine weitere Bedingung an diese geknüpft. Mit diesen erhalten wir als weiteres Hauptresultat mindestens lineare oder superlineare Konvergenz der Folge. Unter Annahmen, die eher theoretischer Natur sind, können wir auch quadratische Konvergenz zeigen.

5.1 Verfahren

Das hier beschriebene Newton-Verfahren ist eine spezielle Version des in Kapitel 3 dargestellten Trust-Region-Verfahrens und läßt sich nur auf zweimal stetig differenzierbare Funktionen anwenden. Es wird nämlich statt der Matrix $B_k \in \mathbb{R}^{n\times n}$ die Hessesche Matrix $\nabla^2 f(x_k)$ am aktuellen Iterationspunkt x_k benutzt. Weiterhin begnügen wir uns nicht damit, den Cauchy-Punkt als nächsten Iterationspunkt zu wählen, sondern fügen an dieser Stelle die Zwischenschritte ein. In Abschnitt 4.2 haben wir ausführlich beschrieben, daß die Zwischenschritte Forderung (5.1.1) erfüllen und somit in das Newton-Verfahren eingebettet sind.

Verfahren 5.1.1 (Newton-Verfahren) Sei die Funktion $f : \mathbb{R}^n \to \mathbb{R}$ eine auf einer offenen Obermenge von M zweimal stetig differenzierbare Abbildung.

Dann lautet das *Newton-Verfahren*

- Gegeben seien die Konstanten $0 < \mu_0 < \frac{1}{2}$, $\mu_1 > 0$, $0 < \rho_0 < \rho_1 < \rho_2 < 1$ und $0 < \sigma_1 < \sigma_2 < 1 < \sigma_3$.
- Sei $x_0 \in M$ gegeben. Berechne $\nabla f(x_0)$ sowie $\nabla^2 f(x_0)$ und bestimme ein $\Delta_0 > 0$.
- Für $k \in \mathbb{N}$
  - Ist x_k ein stationärer Punkt der Optimierungsaufgabe (P), dann breche das Verfahren hier ab. Siehe hierzu Abschnitt 3.2.
  - Andernfalls
    - Berechne den Cauchy-Schritt p_k^C zur Optimierungsaufgabe

$$\begin{aligned} &\text{Minimiere } f_k(p) := f(x_k) + \nabla f(x_k)^T p + \frac{1}{2} p^T \nabla^2 f(x_k) p \\ &\text{auf } \{p \in \mathbb{R}^n \mid \|p\| < \Delta_k\}. \end{aligned}$$

      Der Cauchy-Schritt ist allerdings nicht notwendigerweise eine Lösung dieser Optimierungsaufgabe, siehe hierzu Abschnitt 3.3.
    - Berechne einen Schritt p_k, der

$$f_k(p_k) - f(x_k) \leq \mu_0 (f_k(p_k^C) - f(x_k)) \tag{5.1.1}$$

      erfüllt, wobei $\|p_k\| \leq \mu_1 \Delta_k$ und $x_k + p_k \in M$ sei. Benutze dazu die Zwischenschritte aus Abschnitt 4.2.
    - Berechne

$$\chi_k := \frac{f(x_k) - f(x_k + p_k)}{f(x_k) - f_k(p_k)}$$

      und setze

$$x_{k+1} := \begin{cases} x_k & \text{falls } \chi_k \leq \rho_0, \\ x_k + p_k & \text{falls } \chi_k > \rho_0. \end{cases} \tag{5.1.2}$$

      Ein Iterationsindex k, bei dem $\chi_k > \rho_0$ erfüllt ist, wird *erfolgreich* genannt. Setze

$$\Delta_{k+1} \begin{cases} \in [\sigma_1 \min\{\|p_k\|, \Delta_k\}, \sigma_2 \Delta_k] & \text{falls } \chi_k \leq \rho_1, \\ \in [\sigma_1 \Delta_k, \sigma_3 \Delta_k] & \text{falls } \rho_1 < \chi_k < \rho_2, \\ \in [\Delta_k, \sigma_3 \Delta_k] & \text{falls } \chi_k \geq \rho_2. \end{cases}$$

5.2 Konvergenz

In diesem Abschnitt untersuchen wir das Konvergenzverhalten des Newton-Verfahrens. Der zentrale Konvergenzsatz dieses Abschnitts sagt aus, daß bei positiver Definitheit der Hesseschen Matrix eines Häufungspunktes der Folge, die durch das Newton-Verfahren erzeugt wird, auf einem speziellen Untervektorraum des $\mathbb{R}^n$ die ganze Folge gegen einen stationären Punkt der Optimierungsaufgabe (P) konvergiert. Als erstes aber werden wir das letzte Resultat bei der Konvergenzanalyse des Trust-Region-Verfahrens, Satz 3.5.5, auf das Newton-Verfahren übertragen und feststellen, daß sich einige Voraussetzungen vereinfachen.

Korollar 5.2.1 *Sei $f : \mathbb{R}^n \to \mathbb{R}$ eine auf einer offenen Obermenge von M zweimal stetig differenzierbare Abbildung. Das Newton-Verfahren breche nicht vorzeitig ab und sei $(x_k)_{k\in\mathbb{N}}$ die Folge, die durch dieses Verfahren erzeugt wird. Sei weiterhin $\nabla^2 f$ auf der Niveaumenge $L(M; x_0)$ beschränkt und $x^* \in M$ ein Häufungspunkt der Folge $(x_k)_{k\in\mathbb{N}}$. Dann existiert eine Teilfolge $(k_i)_{i\in\mathbb{N}}$ von erfolgreichen Iterationsindizes, so daß $(x_{k_i})_{i\in\mathbb{N}}$ gegen x^* konvergiert und für die*

$$\lim_{i\to\infty} \left\|\nabla_M f(x^C_{k_i})\right\| = 0$$

gilt. Ferner konvergiert auch $(x^C_{k_i})_{i\in\mathbb{N}}$ gegen x^* und daher ist x^* ein stationärer Punkt der Optimierungsaufgabe (P).

Beweis. Dieses Resultat folgt mit Satz 3.5.5, wenn wir die noch fehlenden Voraussetzungen zur Anwendung dieses Satzes zeigen. Zuerst zeigen wir, daß ∇f auf $L(M; x_0)$ gleichmäßig stetig ist. Seien dazu $x, y \in L(M; x_0)$ beliebig gewählt. Es gilt

$$\begin{aligned}
\|\nabla f(y) - \nabla f(x)\| &= \left\| \int_0^1 \nabla^2 f(x + t(y-x))(y-x)\, dt \right\| \\
&\le \int_0^1 \left\|\nabla^2 f(x + t(y-x))(y-x)\right\| \, dt \\
&\le \int_0^1 \left\|\nabla^2 f(x + t(y-x))\right\| \|y-x\| \, dt
\end{aligned}$$

und da $\nabla^2 f$ auf $L(M; x_0)$ beschränkt ist, ergibt sich hieraus sofort die gleichmäßige Stetigkeit. Zu zeigen bleibt noch die Implikation (3.5.11)

$$\sum_{k=0}^{\infty} \|p_k\| < \infty \quad \Longrightarrow \quad \lim_{k\to\infty} \frac{f(x_k + p_k) - f_k(p_k)}{\|p_k\|^2} = 0.$$

Sei also $\sum_{k=0}^{\infty} \|p_k\| < \infty$. Trivialerweise konvergiert dann auch die Folge $(p_k)_{k\in\mathbb{N}}$ gegen Null. Aus der Ungleichung

$$\|x_{k+1} - x_k\| \le \|p_k\| \qquad \text{für alle } k \in \mathbb{N}$$

erhält man zusammen mit der Dreiecksungleichung, daß $(x_k)_{k\in\mathbb{N}}$ eine Cauchy-Folge ist und damit ebenfalls konvergiert. Da weiterhin f auf einer offenen Obermenge von M zweimal stetig differenzierbar ist, existiert nach dem Satz von Taylor für jedes $k \in \mathbb{N}$ ein $0 < t_k < 1$, so daß

$$f(x_k + p_k) = f(x_k) + \nabla f(x_k)^T p_k + \frac{1}{2} {p_k}^T \nabla^2 f(x_k + t_k p_k) p_k$$

ist. Aus der Konvergenz von $(p_k)_{k\in\mathbb{N}}$ gegen Null und der Konvergenz von $(x_k)_{k\in\mathbb{N}}$ ergibt sich

$$\lim_{k\to\infty} \frac{{p_k}^T (\nabla^2 f(x_k + t_k p_k) - \nabla^2 f(x_k)) p_k}{\|p_k\|^2} = 0$$

und so zusammen mit der Taylor-Entwicklung

$$\lim_{k\to\infty} \frac{f(x_k + p_k) - f_k(p_k)}{\|p_k\|^2} = 0.$$

■

Die weitere Konvergenzanalyse des Newton-Verfahrens erfordert, daß wir annehmen, daß eine Teilfolge $(x_{k_i})_{i\in\mathbb{N}}$ von Iterationspunkten gegen einen stationären Punkt $x^* \in M$ der Optimierungsaufgabe (P) konvergiert, der bestimmte Regularitätsvoraussetzungen erfüllt, woraus wir folgern werden, daß x^* ein isolierter stationärer Punkt ist.

Definition 5.2.2 (Isolierter stationärer Punkt) Sei $\Omega \subseteq \mathbb{R}^n$ eine nichtleere, abgeschlossene und konvexe Menge und $f : \mathbb{R}^n \to \mathbb{R}$ eine auf einer offenen Obermenge von Ω differenzierbare Abbildung. Sei weiterhin $x \in \Omega$ ein stationärer Punkt der Aufgabe, f auf Ω zu minimieren. Dann ist x ein *isolierter stationärer Punkt* der Aufgabe, f auf Ω zu minimieren, wenn eine Umgebung $U(x) \subseteq \mathbb{R}^n$ von x existiert, so daß $U(x)$ keinen stationären Punkt außer x enthält.

Die Regularitätsvoraussetzungen benötigen zwei weitere Definitionen.

Definition 5.2.3 (Aufgespannter Kegel, aufgespannter Vektorraum) Sei $G \subseteq \mathbb{R}^n$ eine nichtleere Menge. Dann ist

- der von G *aufgespannte Kegel* durch

$$\text{cone}(G) := \{tg \mid t \geq 0 \text{ und } g \in G\}$$

definiert,

- der von G *aufgespannte Vektorraum* durch

$$\operatorname{span}(G) := \left\{ \sum_{i=1}^{r} \lambda_i g_i \;\middle|\; \begin{array}{l} r \in \mathbb{N} \setminus \{0\} \text{ und } \lambda_i \in \mathbb{R} \text{ für alle } 1 \leq i \leq r \\ \text{und } g_i \in G \text{ für alle } 1 \leq i \leq r \end{array} \right\}$$

  definiert.

Wir werden fordern, daß die Hessesche Matrix $\nabla^2 f(x^*)$ eines stationären Punktes $x^* \in M$ auf der Menge

$$D(M; x^*) := \operatorname{span}(E(M; -\nabla f(x^*)) - x^*)$$

positiv definit ist. Da für eine beliebige Menge $G \subseteq \mathbb{R}^n$ die Inklusion

$$\operatorname{cone}(G) \subseteq \operatorname{span}(G)$$

gilt, haben wir auch

$$\operatorname{cone}(E(M; -\nabla f(x^*)) - x^*) \subseteq D(M; x^*). \tag{5.2.3}$$

Im nächsten Lemma werden wir eine Charakterisierung der Teilmenge

$$\operatorname{cone}(E(M; -\nabla f(x^*)) - x^*)$$

liefern, die wir dazu nutzen können, um diese Menge mit isolierten stationären Punkten in Verbindung zu bringen. Die nächsten beiden Lemmata stammen aus dem sechsten Abschnitt von Burke and Moré, 1994.

Lemma 5.2.4 *Sei $f : \mathbb{R}^n \to \mathbb{R}$ eine auf einer offenen Obermenge von M stetig differenzierbare Abbildung. Sei weiterhin $x^* \in M$ ein stationärer Punkt der Optimierungsaufgabe (P). Dann gilt*

$$\operatorname{cone}(E(M; -\nabla f(x^*)) - x^*) = \{p \in T(M; x^*) \mid \nabla f(x^*)^T p = 0\}.$$

Beweis.

$\subseteq$ **:** Sei $p \in \operatorname{cone}\{E(M; -\nabla f(x^*)) - x^*\}$ beliebig gewählt. Dann ist $p = t(x - x^*)$ für ein $t > 0$ und ein $x \in E(M; -\nabla f(x^*))$. Daraus ergibt sich durch Umstellen der Gleichung, daß $p \in T(M; x^*)$ ist. Aus $x \in E(M; -\nabla f(x^*))$ folgt

$$(-\nabla f(x^*))^T x \geq (-\nabla f(x^*))^T x^*$$

und da x^* ein stationärer Punkt der Optimierungsaufgabe (P) ist, erhalten wir mit Satz 4.1.2

$$(-\nabla f(x^*))^T x^* \geq (-\nabla f(x^*))^T x.$$

Mit $p = t(x - x^*)$ bekommen wir $\nabla f(x^*)^T p = 0$. $\surd$

$\supseteq$: Sei $p \in T(M;x^*)$ mit $\nabla f(x^*)^T p = 0$ beliebig gewählt. Nach Definition des Tangentialkegels existieren Folgen $(t_k)_{k\in\mathbb{N}} \subset \mathbb{R}_+$ und $(r_k)_{k\in\mathbb{N}} \subset \mathbb{R}^n$ mit

$$x_k := x^* + t_k p + r_k \in M \text{ für alle } k \in \mathbb{N}, \quad \lim_{k\to\infty} t_k = 0 \text{ und } \lim_{k\to\infty} \frac{r_k}{t_k} = 0.$$

Wir betrachten nun die Menge $I(M;x^*)$ der aktiven Restriktionen in x^*. In einem Fall ist $i \notin I(M;x^*)$ mit $1 \le i \le m$. Das heißt, daß $l_i < {c_i}^T x^* < u_i$ und damit auch

$$l_i < {c_i}^T (x^* + tp) < u_i$$

für alle hinreichend kleinen $t > 0$ ist. Haben wir stattdessen $i \in I(M;x^*)$ mit $1 \le i \le m$ und nehmen wir ${c_i}^T x^* = l_i$ an, so erhalten wir aus $x^* + t_k p + r_k \in M$ für alle $k \in \mathbb{N}$

$${c_i}^T (t_k p + r_k) \ge 0 \quad \text{für alle } k \in \mathbb{N}.$$

Dividiert man nun beide Seiten durch t_k, so erhält man durch Übergang zum Grenzwert für $k \to \infty$, daß ${c_i}^T p \ge 0$ und damit

$$l_i \le {c_i}^T (x^* + tp) \le u_i$$

für alle hinreichend kleinen $t > 0$ ist. Der Beweis dieser Ungleichung im Fall ${c_i}^T x^* = u_i$ erfolgt analog und so haben wir $x^* + tp \in E(M; -\nabla f(x^*))$ für alle hinreichend kleinen $t > 0$ gezeigt. Daraus erhalten wir dann $p \in \text{cone}(E(M; -\nabla f(x^*)) - x^*)$. $\surd$

■

Das nun folgende Lemma bringt die Menge $\text{cone}(E(M; -\nabla f(x^*)) - x^*)$ mit isolierten stationären Punkten in Beziehung. Da die Aussage aber auch allgemein für eine nichtleere, abgeschlossene und konvexe Menge $\Omega \subseteq \mathbb{R}^n$ gilt, benutzen wir nicht $\text{cone}(E(M; -\nabla f(x^*)) - x^*)$ selber, sondern die im letzten Lemma angegebene Charakterisierung.

Lemma 5.2.5 *Sei $\Omega \subseteq \mathbb{R}^n$ eine nichtleere, abgeschlossene und konvexe Menge und $f : \mathbb{R}^n \to \mathbb{R}$ eine auf einer offenen Obermenge von Ω stetig differenzierbare Abbildung und in $x^* \in \Omega$ zweimal stetig differenzierbar. Sei weiterhin x^* ein stationärer Punkt der Aufgabe, f auf Ω zu minimieren, und $\nabla^2 f(x^*)$ auf der Menge*

$$\{p \in T(\Omega;x^*) \mid \nabla f(x^*)^T p = 0\}$$

positiv definit. Dann ist x^* ein isolierter stationärer Punkt der Aufgabe, f auf Ω zu minimieren.

Beweis. Der Beweis erfolgt durch Widerspruch. Angenommen, x^* ist kein isolierter stationärer Punkt der Aufgabe, f auf Ω zu minimieren. Dann existiert eine Folge $(x_k)_{k\in\mathbb{N}} \subset M$ von stationären Punkten, die gegen x^* konvergiert. Sei nun die Folge $(v_k)_{k\in\mathbb{N}}$ durch

$$v_k := \frac{x_k - x^*}{\|x_k - x^*\|}$$

definiert und nach einem möglicherweise notwendigen Übergang zu Teilfolgen sei $v := \lim_{k\to\infty} v_k$. Es ist dann $\|v_k\| = 1$ für alle $k \in \mathbb{N}$, woraus wir $\|v\| = 1$ und damit $v \neq 0$ folgern. Außerdem ist $v_k \in T(M; x^*)$ für alle $k \in \mathbb{N}$ und aufgrund der Abgeschlossenheit von Tangentialkegeln gilt somit auch $v \in T(M; x^*)$. Da x_k für alle $k \in \mathbb{N}$ stationär ist, haben wir

$$\nabla f(x_k)^T (x^* - x_k) \geq 0 \quad \text{für alle } k \in \mathbb{N}$$

und da x^* selber stationär ist, haben wir auch

$$\nabla f(x^*)^T (x_k - x^*) \geq 0 \quad \text{für alle } k \in \mathbb{N},$$

was zusammengenommen

$$\nabla f(x^*)^T v = 0$$

und

$$(\nabla f(x_k) - \nabla f(x^*))^T (x_k - x^*) \leq 0 \quad \text{für alle } k \in \mathbb{N}$$

impliziert. Nach der Definition der Ableitung ist

$$\nabla f(x_k) - \nabla f(x^*) = \nabla^2 f(x^*)(x_k - x^*) + \xi(x_k - x^*) \quad \text{für alle } k \in \mathbb{N},$$

mit einer Funktion $\xi : \mathbb{R}^n \to \mathbb{R}^n$ mit der Eigenschaft

$$\lim_{h\to 0} \frac{\xi(h)}{\|h\|} = 0.$$

Daraus bekommen wir

$$(x_k - x^*)^T \nabla^2 f(x^*)(x_k - x^*) + \xi(x_k - x^*)^T (x_k - x^*) \leq 0 \quad \text{für alle } k \in \mathbb{N},$$

was man mit zweimaliger Division durch $\|x_k - x^*\|$ zu

$$v_k{}^T \nabla^2 f(x^*) v_k + \left(\frac{\xi(x_k - x^*)}{\|x_k - x^*\|} \right)^T v_k \leq 0 \quad \text{für alle } k \in \mathbb{N}$$

umformt. Der Übergang zum Limes für $k \to \infty$ liefert

$$v^T \nabla^2 f(x^*) v \leq 0,$$

womit ein Widerspruch zur Voraussetzung, daß $\nabla^2 f(x^*)$ auf der Menge

$$\{p \in T(M; x^*) \mid \nabla f(x^*)^T p = 0\}$$

positiv definit ist, vorliegt. ■

Als weiteren Begriff definieren wir den isolierten Häufungspunkt.

Definition 5.2.6 (Isolierter Häufungspunkt) Sei $(x_k)_{k\in\mathbb{N}} \subset \mathbb{R}^n$ eine Folge und $x \in \mathbb{R}^n$ ein Häufungspunkt von $(x_k)_{k\in\mathbb{N}}$. Dann ist x ein *isolierter Häufungspunkt* von $(x_k)_{k\in\mathbb{N}}$, wenn eine Umgebung $U(x) \subseteq \mathbb{R}^n$ von x existiert, so daß $U(x)$ keinen Häufungspunkt außer x enthält.

Das nächste Lemma ist rein technischer Natur und stammt aus dem sechsten Abschnitt von Burke et al., 1990.

Lemma 5.2.7 *Sei $(x_k)_{k\in\mathbb{N}} \subset \mathbb{R}^n$ eine Folge und $x \in \mathbb{R}^n$ ein isolierter Häufungspunkt von $(x_k)_{k\in\mathbb{N}}$. Dann konvergiert entweder $(x_k)_{k\in\mathbb{N}}$ gegen x oder es existieren eine Teilfolge $(k_i)_{i\in\mathbb{N}} \subseteq \mathbb{N}$ von Indizes, so daß $(x_{k_i})_{i\in\mathbb{N}}$ gegen x konvergiert und ein $\varepsilon > 0$, so daß*

$$\|x_{k_i+1} - x_{k_i}\| \geq \varepsilon \quad \text{für alle } i \in \mathbb{N}$$

gilt.

Beweis. Wir nehmen an, die Folge $(x_k)_{k\in\mathbb{N}}$ konvergiert nicht. Sei dann $U(x) \subseteq \mathbb{R}^n$ eine Umgebung von x, so daß $U(x)$ keinen Häufungspunkt außer x enthält. Wähle jetzt ein $\varepsilon > 0$, so daß

$$\{y \in \mathbb{R}^n \mid \|y - x\| \leq 2\varepsilon\} \subseteq U(x)$$

ist. Falls

$$\|x_k - x\| \leq 2\varepsilon$$

für alle hinreichend großen $k \in \mathbb{N}$ gilt, dann ist $(x_k)_{k\in\mathbb{N}}$ beschränkt und für jeden Häufungspunkt $y \in \mathbb{R}^n$ von $(x_k)_{k\in\mathbb{N}}$ folgt $y \in U(x)$, das $y = x$ und damit die Konvergenz von $(x_k)_{k\in\mathbb{N}}$ gegen x impliziert. Also haben wir gezeigt, daß

$$\|x_k - x\| > 2\varepsilon$$

für unendlich viele $k \in \mathbb{N}$ gilt und somit die Existenz einer Teilfolge $(k_i)_{i\in\mathbb{N}} \subseteq \mathbb{N}$ von Indizes nachgewiesen, für die

$$\|x_{k_i} - x^*\| \leq 2\varepsilon \quad \text{und} \quad \|x_{k_i+1} - x^*\| > 2\varepsilon \quad \text{für alle } i \in \mathbb{N}$$

gilt. Die Folge $(x_{k_i})_{i\in\mathbb{N}}$ ist demnach beschränkt und für jeden Häufungspunkt $y \in \mathbb{R}^n$ von $(x_{k_i})_{i\in\mathbb{N}}$ folgt wie eben $y \in U(x)$, das $y = x$ und damit die Konvergenz von $(x_{k_i})_{i\in\mathbb{N}}$ gegen x nach sich zieht. Wir haben sogar

$$\|x_{k_i} - x^*\| \leq \varepsilon$$

für alle hinreichend großen $i \in \mathbb{N}$ und kürzt man die Folge $(k_i)_{i\in\mathbb{N}}$ um die Folgenglieder, für die $\|x_{k_i} - x^*\| > \varepsilon$ ist, erhalten wir

$$\|x_{k_i+1} - x_{k_i}\| \geq \|x_{k_i+1} - x^*\| - \|x_{k_i} - x^*\| > 2\varepsilon - \varepsilon = \varepsilon,$$

womit die Behauptung des Lemmas gezeigt ist. ■

Der nun folgende Satz stellt das am Anfang angekündigte Konvergenzresultat dar, in dem die drei letzten Lemmata endlich zum Zuge kommen. Dieser Satz nennt Voraussetzungen, unter denen die gesamte Folge, die durch das Newton-Verfahren erzeugt wird, gegen einen stationären Punkt der Optimierungsaufgabe (P) konvergiert.

Satz 5.2.8 *Sei $f : \mathbb{R}^n \to \mathbb{R}$ eine auf einer offenen Obermenge von M zweimal stetig differenzierbare Abbildung. Das Newton-Verfahren breche nicht vorzeitig ab und sei $(x_k)_{k\in\mathbb{N}}$ die Folge, die durch dieses Verfahren erzeugt wird. Sei weiterhin $\nabla^2 f$ auf der Niveaumenge $L(M; x_0)$ beschränkt und es gelte die Inklusion (4.1.4). Sei $x^* \in M$ ein Häufungspunkt der Folge $(x_k)_{k\in\mathbb{N}}$ mit der Eigenschaft, daß die Hessesche Matrix $\nabla^2 f(x^*)$ auf der Menge $D(M; x^*)$ positiv definit ist. Dann ist x^* ist ein stationärer Punkt der Optimierungsaufgabe (P) und die ganze Folge $(x_k)_{k\in\mathbb{N}}$ konvergiert gegen x^*.*

Beweis. Da die Voraussetzungen aus Korollar 5.2.1 in diesem Satz erfüllt sind, folgt, daß x^* ein stationärer Punkt der Optimierungsaufgabe (P) ist. Der Beweis der Konvergenz von $(x_k)_{k\in\mathbb{N}}$ gegen x^* erfolgt in zwei Schritten.

- Als erstes zeigen wir, daß x^* ein isolierter Häufungspunkt der Folge $(x_k)_{k\in\mathbb{N}}$ ist. Da die Hessesche Matrix $\nabla^2 f(x^*)$ auf der Menge $D(M; x^*)$ positiv definit ist, ist sie wegen (5.2.3) auch auf

  $$\operatorname{cone}(E(M; -\nabla f(x^*)) - x^*)$$

  positiv definit und damit nach Lemma 5.2.4 auch auf der Menge

  $$\{p \in T(M; x^*) \mid \nabla f(x^*)^T p = 0\}.$$

  Lemma 5.2.5 impliziert nun, daß x^* ein isolierter stationärer Punkt der Optimierungsaufgabe (P) ist. Daher existiert eine Umgebung $U(x^*) \subseteq \mathbb{R}^n$ von x^*, so daß $U(x^*)$ keinen stationären Punkt außer x^*

enthält. $U(x^*)$ enthält auch keinen weiteren Häufungspunkt der Folge $(x_k)_{k\in\mathbb{N}}$ außer x^*, da jeder Häufungspunkt von $(x_k)_{k\in\mathbb{N}}$ nach Korollar 5.2.1 ein stationärer Punkt der Optimierungsaufgabe (P) ist. Damit haben wir gezeigt, daß x^* ein isolierter Häufungspunkt der Folge $(x_k)_{k\in\mathbb{N}}$ ist. $\surd$

- Der Beweis der Konvergenz von $(x_k)_{k\in\mathbb{N}}$ gegen x^* wird nun durch Widerspruch geführt. Wir nehmen an, die Folge $(x_k)_{k\in\mathbb{N}}$ konvergiert nicht gegen x^*. Nach Lemma 5.2.7 existieren eine Teilfolge $(k_i)_{i\in\mathbb{N}} \subseteq \mathbb{N}$ von Iterationsindizes, so daß $(x_{k_i})_{i\in\mathbb{N}}$ gegen x konvergiert und ein $\varepsilon > 0$, so daß

$$\|x_{k_i+1} - x_{k_i}\| \geq \varepsilon \quad \text{für alle } i \in \mathbb{N}$$

gilt. Insbesondere ist $\|p_{k_i}\| \geq \varepsilon$ für alle $i \in \mathbb{N}$ und damit $(k_i)_{i\in\mathbb{N}}$ eine Teilfolge von erfolgreichen Iterationsindizes. Wir definieren nun die Folge $(v_i)_{i\in\mathbb{N}}$ durch

$$v_i := \frac{p_{k_i}}{\|p_{k_i}\|}.$$

und nach einem möglicherweise notwendigen Übergang zu Teilfolgen sei $v := \lim_{i\to\infty} v_i$. Es gilt

$$x_{k_i} + tv_i \in M \quad \text{für alle } 0 \leq t \leq \varepsilon \quad \text{und alle } i \in \mathbb{N},$$

da $\|p_{k_i}\| \geq \varepsilon$ für alle $i \in \mathbb{N}$ ist und somit durch Übergang zum Grenzwert für $i \to \infty$

$$x^* + tv \in M \quad \text{für alle } 0 \leq t \leq \varepsilon.$$

Wir haben damit $v \in T(M; x^*)$ erhalten und werden nun $\nabla f(x^*)^T v = 0$ zeigen. Aufgrund der Beschränktheit von $\nabla^2 f$ auf der Niveaumenge $L(M; x_0)$ ist insbesondere die Folge $(\nabla^2 f(x_k))_{k\in\mathbb{N}}$ der Hesseschen Matrizen an den Iterationspunkten beschränkt. Da weiterhin

$$\varepsilon \leq \|p_{k_i}\| \leq \mu_1 \Delta_{k_i} \quad \text{für alle } i \in \mathbb{N}$$

gilt, ist die Teilfolge $(\Delta_{k_i})_{i\in\mathbb{N}}$ durch eine Konstante größer Null nach unten beschränkt. Da aber die Folge $(f(x_k))_{k\in\mathbb{N}}$ monoton fallend ist, gilt auch

$$\lim_{i\to\infty} (f(x_{k_i}) - f(x_{k_i} + p_{k_i})) = 0.$$

Zusammen mit Ungleichung (3.4.9) in der Form

$$\begin{aligned} &f(x_{k_i}) - f(x_{k_i} + p_{k_i}) \\ >&\rho_0 {\mu_0}^2 \zeta \left(\frac{\|\pi_{k_i}(\alpha_{k_i})\|}{\alpha_{k_i}} \right) \min \left\{ \Delta_{k_i}, \frac{1}{1 + \|\nabla^2 f(x_{k_i})\|} \left(\frac{\|\pi_{k_i}(\alpha_{k_i})\|}{\alpha_{k_i}} \right) \right\} \end{aligned}$$

für alle $i \in \mathbb{N}$ ergibt sich daraus

$$\lim_{i\to\infty} \frac{\|\pi_{k_i}(\alpha_{k_i})\|}{\alpha_{k_i}} = 0.$$

Aufgrund der Identität $x_k^C - x_k = \pi_k(\alpha_k)$ und $\alpha_k \leq \gamma_3$ für alle $k \in \mathbb{N}$ ist auch

$$\lim_{i\to\infty} \left\|x_{k_i}^C - x_{k_i}\right\| = 0$$

und Lemma 3.5.1 impliziert nun

$$\lim_{i\to\infty} \left\|\nabla_M f(x_{k_i}^C)\right\| = 0.$$

Aus Lemma 4.1.9 folgt weiter

$$x_{k_i}^C \in E(M; -\nabla f(x^*)) \quad \text{für alle } i \in \mathbb{N},$$

wobei man die Folge $(k_i)_{i\in\mathbb{N}}$ möglicherweise um eine endliche Anzahl an Folgengliedern kürzen muß. Voraussetzung (4.1.4) ergibt, daß

$$x_{k_i} + p_{k_i} \in E(M; -\nabla f(x^*)) \quad \text{für alle } i \in \mathbb{N}$$

ist, speziell folgt

$$\nabla f(x^*)^T (x_{k_i} + p_{k_i} - x^*) = 0 \quad \text{für alle } i \in \mathbb{N}.$$

Dividiert man nun durch $\|p_{k_i}\|$ und berücksichtigt man $\|p_{k_i}\| \geq \varepsilon$ für alle $i \in \mathbb{N}$, so erhalten wir nach Übergang zum Grenzwert für $i \to \infty$

$$\nabla f(x^*)^T v = 0.$$

Wir haben jetzt gezeigt, daß

$$v \in \{p \in T(M; x^*) \mid \nabla f(x^*)^T p = 0\}$$

gilt. Aufgrund der Inklusion (5.2.3) und Lemma 5.2.4 bekommen wir $v \in D(M; x^*)$ und die Voraussetzung, daß die Hessesche Matrix $\nabla^2 f(x^*)$ auf der Menge $D(M; x^*)$ positiv definit ist, impliziert mit $v \neq 0$

$$v^T \nabla^2 f(x^*) v > 0.$$

Wegen $f_{k_i}(p_{k_i}) - f(x_{k_i}) < 0$ für alle $i \in \mathbb{N}$ ist

$$\frac{1}{2} \|p_{k_i}\| \, {v_i}^T \nabla^2 f(x_{k_i}) v_i < -\nabla f(x_{k_i})^T v_i \quad \text{für alle } i \in \mathbb{N}.$$

Der Übergang zum Limes für $i \to \infty$ sowie $\|p_{k_i}\| \geq \varepsilon$ für alle $i \in \mathbb{N}$ impliziert jetzt

$$0 < \frac{1}{2} \varepsilon v^T \nabla^2 f(x^*) v \leq -\nabla f(x^*)^T v = 0.$$

Dieser Widerspruch vollendet den Beweis. √

■

5.3 Konvergenzgeschwindigkeit

In diesem Abschnitt werden wir zeigen, daß das Newton-Verfahren unter gewissen Voraussetzungen mindestens linear oder superlinear konvergiert. Dieses Resultat zur Konvergenzgeschwindigkeit basiert darauf, daß wir zeigen, daß die Iterationspunkte des Newton-Verfahrens für alle hinreichend großen Iterationsindizes nicht auf dem Rand der Kugel, die durch den Radius $\mu_1\Delta_k$ für $k \in \mathbb{N}$ charakterisiert ist, liegen. Die Zwischenschritte helfen uns dabei, eine weitere Bedingung zu erfüllen, die zum Beweis des Resultates über die Konvergenzgeschwindigkeit wichtig ist. Als Hilfsmittel benötigen wir als erstes zwei Lemmata.

Lemma 5.3.1 *Sei $\varphi : \mathbb{R} \to \mathbb{R}$ eine auf dem Intervall $[0,1]$ zweimal stetig differenzierbare Abbildung und es existiere ein $\varepsilon > 0$, so daß*

$$\varphi''(\alpha) \geq \varepsilon \quad \text{für alle } 0 \leq \alpha \leq 1$$

gilt und ein $0 < \mu < 1$, so daß

$$\varphi(1) \leq \varphi(0) + \mu\varphi'(0)$$

ist. Dann gilt

$$\varphi(0) - \varphi(1) \geq \frac{\mu}{2(1-\mu)}\varepsilon.$$

Beweis. Nach dem Satz von Taylor existiert ein $0 < \hat{\alpha} < 1$, für das

$$\varphi(1) = \varphi(0) + \varphi'(0) + \frac{1}{2}\varphi''(\hat{\alpha})$$

gilt. Daraus ergibt sich

$$\begin{aligned}\frac{1}{2}\varphi''(\hat{\alpha}) =& \varphi(1) - \varphi(0) - \varphi'(0)\\ \leq& \varphi(0) + \mu\varphi'(0) - \varphi(0) - \varphi'(0)\\ =& -(1-\mu)\varphi'(0)\end{aligned}$$

und so mit

$$\begin{aligned}\varphi(0) - \varphi(1) \geq& \mu(-\varphi'(0))\\ \geq& \frac{\mu}{2(1-\mu)}\varphi''(\hat{\alpha})\\ \geq& \frac{\mu}{2(1-\mu)}\varepsilon\end{aligned}$$

das gewünschte Ergebnis. ■

Lemma 5.3.2 *Sei* $\{0\} \neq V \subseteq \mathbb{R}^n$ *ein Vektorraum und* $\{v_1, \ldots, v_s\} \subset \mathbb{R}^n$ *mit* $1 \leq s \leq n$ *eine Orthonormalbasis dieses Vektorraums. Sei weiterhin*

$$A := \begin{pmatrix} v_1 & \ldots & v_s \end{pmatrix} \in \mathbb{R}^{n \times s}$$

die Matrix, deren Spaltenvektoren die Vektoren dieser Orthonormalbasis sind. Dann gelten für alle $x \in \mathbb{R}^n$

(i) $\Pi_V(x) = AA^T x$,

(ii) $\|x\|^2 = \|x - \Pi_V(x)\|^2 + \|\Pi_V(x)\|^2$.

Beweis.

(i) Sei $x \in \mathbb{R}^n$ beliebig gewählt. Nach Lemma 2.1.2 müssen wir lediglich prüfen, ob

$$(AA^T x - x)^T (y - AA^T x) \geq 0 \quad \text{für alle } y \in V$$

gilt. Es ist

$$\begin{aligned} &(AA^T x - x)^T (y - AA^T x) \\ =&(AA^T x)^T y - (AA^T x)^T AA^T x - x^T y + x^T AA^T x \\ =&x^T AA^T y - x^T A \underbrace{A^T A}_{=id_{\mathbb{R}^s}} A^T x - x^T y + x^T AA^T x \\ =&x^T AA^T y - x^T y \quad \text{für alle } y \in V, \end{aligned}$$

da die Matrix A die Vektoren einer Orthonormalbasis von V als Spaltenvektoren besitzt. Sei nun $y \in V$ beliebig gewählt. Dann haben wir die Darstellung $y = Az$ für ein $z \in \mathbb{R}^s$ und setzen wir dies in die obige Gleichung ein, so bekommen wir

$$\begin{aligned} x^T AA^T y - x^T y =&x^T AA^T Az - x^T Az \\ =&x^T Az - x^T Az \\ =&0 \quad \text{für alle } y \in V, \end{aligned}$$

womit wir die erste Aussage des Lemmas gezeigt haben. √

(ii) Sei nun wieder $x \in \mathbb{R}^n$ beliebig gewählt. Es gilt

$$\begin{aligned} \|x\|^2 =&x^T x \\ =&(x - \Pi_V(x) + \Pi_V(x))^T (x - \Pi_V(x) + \Pi_V(x)) \\ =&\|x - \Pi_V(x)\|^2 + 2(x - \Pi_V(x))^T \Pi_V(x) + \|\Pi_V(x)\|^2 \\ =&\|x - \Pi_V(x)\|^2 + 2(x - AA^T x)^T (AA^T x) + \|\Pi_V(x)\|^2 \\ =&\|x - \Pi_V(x)\|^2 + 2(x^T AA^T x - x^T A \underbrace{A^T A}_{=id_{\mathbb{R}^s}} A^T x) + \|\Pi_V(x)\|^2 \\ =&\|x - \Pi_V(x)\|^2 + \|\Pi_V(x)\|^2. \end{aligned}$$

$\surd$

■

Im folgenden Satz sehen wir, daß alle hinreichend großen Iterationsindizes $k \in \mathbb{N}$ erfolgreich sind und die Folge $(\Delta_k)_{k\in\mathbb{N}}$ nach unten durch eine Konstante größer Null beschränkt ist. Da aber die Folge $(x_k)_{k\in\mathbb{N}}$ konvergiert, muß demnach notwendigerweise

$$\|p_k\| < \mu_1 \Delta_k$$

für alle hinreichend großen $k \in \mathbb{N}$ sein und der Rand der Kugel um den Nullpunkt mit dem Radius $\mu_1\Delta_k$ wird nicht aktiv.

Satz 5.3.3 *Sei $f : \mathbb{R}^n \to \mathbb{R}$ eine auf einer offenen Obermenge von M zweimal stetig differenzierbare Abbildung. Das Newton-Verfahren breche nicht vorzeitig ab und sei $(x_k)_{k\in\mathbb{N}}$ die Folge, die durch dieses Verfahren erzeugt wird. Sei weiterhin $\nabla^2 f$ auf der Niveaumenge $L(M; x_0)$ beschränkt und seien die Forderungen (4.2.5) und (4.2.6) an die Zwischenschritte erfüllt. Sei $x^* \in M$ der Grenzwert der Folge $(x_k)_{k\in\mathbb{N}}$ und ein stationärer Punkt der Optimierungsaufgabe (P). Sei weiterhin die Hessesche Matrix $\nabla^2 f(x^*)$ auf der Menge $D(M; x^*)$ positiv definit. Dann sind alle hinreichend großen Iterationsindizes $k \in \mathbb{N}$ erfolgreich und es gilt*

$$\inf_{k\in\mathbb{N}} \Delta_k > 0.$$

Beweis. Der Beweis erfolgt in zwei Schritten. Im ersten Schritt leiten wir eine Abschätzung über die Zwischenschritte her, im zweiten zeigen wir

$$\lim_{k\to\infty} \chi_k = 1,$$

womit wir durch die Definition von erfolgreichen Schritten und der Iterationsvorschrift für Δ_k die Aussage des Satzes gezeigt haben werden.

- Als erstes wollen wir Satz 4.1.12 anwenden und müssen dazu die Voraussetzungen dieses Satzes überprüfen. Wie im Beweis zu Korollar 5.2.1 zeigt man, daß ∇f auf $L(M; x_0)$ gleichmäßig stetig ist und Implikation (3.5.11) gilt. Die Forderung (4.2.5) an die Zwischenschritte stellt sicher, daß Inklusion (4.1.4) erfüllt ist. Damit wissen wir, daß ein Iterationsindex $\hat{k} \in \mathbb{N}$ existiert, so daß

  $$x_k, x_k^C, x_k + p_k \in E(M; -\nabla f(x^*)) \quad \text{für alle } k \geq \hat{k}$$

  sind. Aufgrund von Implikation (4.1.3) und wiederum Forderung (4.2.5) haben wir auch

  $$\{x_k^1 := x_k^C, \ldots, x_k^{r+1}\} \subset E(M; -\nabla f(x^*)) \quad \text{für alle } k \geq \hat{k}.$$

Da x^* ein stationärer Punkt der Optimierungsaufgabe (P) ist, gilt nach Satz 4.1.2

$$x^* \in E(M; -\nabla f(x^*)).$$

Da die Folge $(x_k)_{k\in\mathbb{N}}$ gegen x^* konvergiert und $\nabla^2 f(x^*)$ auf der Menge $D(M; x^*)$ positiv definit ist, existiert ein $\tilde{k} \geq \hat{k}$, so daß auch $\nabla^2 f(x_k)$ für alle $k \geq \tilde{k}$ auf $D(M; x^*)$ positiv definit ist. Sei nun $k \geq \tilde{k}$ ein beliebiger Iterationsindex. Wir definieren der Einfachheit halber den Zwischenschritt

$$x_k^0 := x_k,$$

und wählen jetzt einen beliebigen Index $0 \leq j \leq r$ eines Zwischenschritts. Durch Anwendung der Definition der exponierten Seitenfläche auf x_k^j und x^* bekommt man

$$\begin{aligned} -\nabla f(x^*)^T x_k^j \geq & -\nabla f(x^*)^T x^* \text{ und} \\ -\nabla f(x^*)^T x^* \geq & -\nabla f(x^*)^T x_k^j \end{aligned}$$

und damit

$$-\nabla f(x^*)^T (x^* - x_k^j) = 0$$

Erneute Anwendung dieser Definition auf x_k^{j+1} ergibt

$$-\nabla f(x^*)^T (x_k^{j+1} - x_k^j + x^*) \geq -\nabla f(x^*)^T y \quad \text{für alle } y \in M$$

und so $x_k^{j+1} - x_k^j + x^* \in E(M; -\nabla f(x^*))$, woraus wir

$$x_k^{j+1} - x_k^j \in D(M; x^*)$$

erhalten. Da die Menge

$$(\{x_k \mid k \geq \tilde{k}\} \cup \{x^*\}) \times \{p \in D(M; x^*) \mid \|p\| = 1\}$$

das cartesische Produkt von kompakten Mengen ist und demnach ebenfalls kompakt ist und stetige Funktionen auf kompakten Mengen ihr Minimum annehmen, existiert eine Konstante $\hat{c} > 0$, so daß

(5.3.4)
$$p^T \nabla^2 f(x_k) p \geq \hat{c}\|p\|^2 \quad \text{für alle } k \geq \tilde{k} \quad \text{und alle } p \in D(M; x^*)$$

ist. Daraus schließen wir

$$(x_k^{j+1} - x_k^j)^T \nabla^2 f(x_k)(x_k^{j+1} - x_k^j) \geq \hat{c}\left\|x_k^{j+1} - x_k^j\right\|^2$$

für alle $k \geq \tilde{k}$ und alle dazugehörigen Indizes $0 \leq j \leq r$ von Zwischenschritten. Sei nun wieder $k \geq \tilde{k}$ ein beliebiger Iterationsindex und

$0 \leq j \leq r$ der Index eines beliebigen Zwischenschritts sowie jetzt die Funktion $\varphi : \mathbb{R} \to \mathbb{R}$ durch

$$\varphi(\alpha) := q_k(\alpha x_k^{j+1} + (1-\alpha)x_k^j)$$

definiert. Ausgeschrieben bedeutet das

$$\begin{aligned}
\varphi(\alpha) =& \nabla f(x_k)^T(\alpha x_k^{j+1} + (1-\alpha)x_k^j - x_k) \\
&+ \frac{1}{2}(\alpha x_k^{j+1} + (1-\alpha)x_k^j - x_k)^T \nabla^2 f(x_k)(\alpha x_k^{j+1} + (1-\alpha)x_k^j - x_k) \\
=& \alpha \nabla f(x_k)^T(x_k^{j+1} - x_k^j) + \nabla f(x_k)^T(x_k^j - x_k) \\
&+ \frac{1}{2}\alpha^2(x_k^{j+1} - x_k^j)^T \nabla^2 f(x_k)(x_k^{j+1} - x_k^j) \\
&+ \alpha(x_k^{j+1} - x_k^j)^T \nabla^2 f(x_k)(x_k^j - x_k) \\
&+ \frac{1}{2}(x_k^j - x_k)^T \nabla^2 f(x_k)(x_k^j - x_k) \quad \text{für alle } \alpha \in \mathbb{R}.
\end{aligned}$$

Die Ableitung der Funktion φ lautet dann

$$\begin{aligned}
\varphi'(\alpha) =& \nabla f(x_k)^T(x_k^{j+1} - x_k^j) \\
&+ \alpha(x_k^{j+1} - x_k^j)^T \nabla^2 f(x_k)(x_k^{j+1} - x_k^j) \\
&+ (x_k^{j+1} - x_k^j)^T \nabla^2 f(x_k)(x_k^j - x_k) \quad \text{für alle } \alpha \in \mathbb{R}
\end{aligned}$$

und die zweite Ableitung

$$\varphi''(\alpha) = (x_k^{j+1} - x_k^j)^T \nabla^2 f(x_k)(x_k^{j+1} - x_k^j) \quad \text{für alle } \alpha \in \mathbb{R}.$$

Mit dem oben Gezeigten gilt nun

$$\varphi''(\alpha) \geq \hat{c}\left\|x_k^{j+1} - x_k^j\right\|^2 \quad \text{für alle } \alpha \in \mathbb{R}.$$

Weiterhin impliziert Forderung (4.2.6) an die Zwischenschritte insbesondere

$$\begin{aligned}
\varphi(1) =& q_k(x_k^{j+1}) \\
\leq& q_k(x_k^j) + \mu_0 \nabla q_k(x_k^j)^T(x_k^{j+1} - x_k^j) \\
=& \varphi(0) + \mu_0 \varphi'(0)
\end{aligned}$$

und somit können wir Lemma 5.3.1 anwenden und wir erhalten

$$\begin{aligned}
q_k(x_k^j) - q_k(x_k^{j+1}) \geq& \frac{\mu_0}{2(1-\mu_0)}\hat{c}\left\|x_k^{j+1} - x_k^j\right\|^2 \\
=& \tilde{c}\left\|x_k^{j+1} - x_k^j\right\|^2
\end{aligned} \tag{5.3.5}$$

für alle $k \geq \tilde{k}$ und alle Indizes $0 \leq j \leq r$ von Zwischenschritten mit der Definition

$$\tilde{c} := \frac{\mu_0}{2(1-\mu_0)}\hat{c}.$$

$\surd$

- Jetzt zeigen wir $\lim_{k\to\infty} \chi_k = 1$. Sei dazu $k \geq \tilde{k}$ ein beliebiger Iterationsindex. Aus der eben gewonnenen Abschätzung (5.3.5) erhalten wir

$$\begin{aligned} f(x_k) - f_k(p_k) =& q_k(x_k^0) - q_k(x_k^{r+1}) \\ =& \sum_{j=0}^{r} q_k(x_k^j) - q_k(x_k^{j+1}) \\ \geq& \sum_{j=0}^{r} \tilde{c}\left\|x_k^{j+1} - x_k^j\right\|^2 \\ \geq& \tilde{c} \max_{0\leq j\leq r} \left\|x_k^{j+1} - x_k^j\right\|^2. \end{aligned}$$

Es gilt nach der Dreiecksungleichung natürlich auch

$$\begin{aligned} \|p_k\| =& \left\|\sum_{j=0}^{r}(x_k^{j+1} - x_k^j)\right\| \\ \leq& \sum_{j=0}^{r} \left\|x_k^{j+1} - x_k^j\right\| \\ \leq& (r+1) \max_{0\leq j\leq r} \left\|x_k^{j+1} - x_k^j\right\|, \end{aligned}$$

so daß wir

$$\begin{aligned} f(x_k) - f_k(p_k) \geq& \tilde{c} \max_{0\leq j\leq r} \left\|x_k^{j+1} - x_k^j\right\|^2 \\ \geq& \frac{\tilde{c}}{(r+1)^2}\|p_k\|^2 \end{aligned}$$

bekommen. Weiterhin existiert nach dem Satz von Taylor ein $0 < t_k < 1$, für das

$$f(x_k + p_k) - f_k(p_k) = \frac{1}{2}{p_k}^T(\nabla^2 f(x_k + t_k p_k) - \nabla^2 f(x_k))p_k$$

gilt, womit wir die weitere Ungleichung

$$\begin{aligned} |f(x_k + p_k) - f_k(p_k)| =& \frac{1}{2}\left|{p_k}^T(\nabla^2 f(x_k + t_k p_k) - \nabla^2 f(x_k))p_k\right| \\ \leq& \|p_k\|^2 \left\|\nabla^2 f(x_k + t_k p_k) - \nabla^2 f(x_k)\right\| \\ \leq& \|p_k\|^2 \sup_{0<t<1} \left\|\nabla^2 f(x_k + t p_k) - \nabla^2 f(x_k)\right\| \end{aligned}$$

erhalten. Aus diesen beiden Abschätzungen ergibt sich nun

$$\begin{aligned}
|\chi_k - 1| &= \left| \frac{f(x_k) - f(x_k + p_k) - (f(x_k) - f_k(p_k))}{f(x_k) - f_k(p_k)} \right| \\
&= \left| \frac{f(x_k + p_k) - f_k(p_k)}{f(x_k) - f_k(p_k)} \right| \\
&\leq \frac{\|p_k\|^2 \sup_{0<t<1} \|\nabla^2 f(x_k + tp_k) - \nabla^2 f(x_k)\|}{\frac{\tilde{c}}{(r+1)^2}\|p_k\|^2} \\
&= \frac{(r+1)^2}{\tilde{c}} \sup_{0<t<1} \left\|\nabla^2 f(x_k + tp_k) - \nabla^2 f(x_k)\right\|.
\end{aligned}$$

Es reicht demnach,

$$\lim_{k\to\infty} \sup_{0<t<1} \left\|\nabla^2 f(x_k + tp_k) - \nabla^2 f(x_k)\right\| = 0$$

zu zeigen. Da die Folge $(x_k)_{k\in\mathbb{N}}$ gegen x^* konvergiert, ist diese Aussage erfüllt, falls auch die Folge $(p_k)_{k\in\mathbb{N}}$ konvergiert. Am Anfang des Beweises haben wir bereits gezeigt, daß

$$x_k, x_k^C, x_k + p_k \in E(M; -\nabla f(x^*)) \quad \text{für alle } k \geq \tilde{k}$$

sind und die mit der gleichen Argumentation wie oben schließen wir

$$p_k \in D(M; x^*) \quad \text{für alle } k \geq \tilde{k}.$$

Insbesondere gilt $\Pi_{D(M;x^*)}(p_k) = p_k$ für alle $k \geq \tilde{k}$ und da $D(M; x^*)$ ein Vektorraum ist, ist nach Lemma 5.3.2 die Projektion auf $D(M; x^*)$ eine lineare Abbildung. Aus der Ungleichung

$$f_k(p_k) - f(x_k) \leq 0 \quad \text{für alle } k \geq \tilde{k}$$

erhalten wir

$$\frac{1}{2} {p_k}^T \nabla^2 f(x_k) p_k \leq -\nabla f(x_k)^T p_k \quad \text{für alle } k \geq \tilde{k}$$

und daher ergeben Ungleichung (5.3.4) und $\Pi_{D(M;x^*)}(p_k) = p_k$

$$\begin{aligned}
\hat{c}\|p_k\|^2 \leq {p_k}^T \nabla^2 f(x_k) p_k &= (\Pi_{D(M;x^*)}(p_k))^T \nabla^2 f(x_k) p_k \\
&\leq -2(\Pi_{D(M;x^*)}(\nabla f(x_k)))^T p_k \\
&\leq 2\left\|\Pi_{D(M;x^*)}(\nabla f(x_k))\right\| \|p_k\| \quad \text{für alle } k \geq \tilde{k}
\end{aligned}$$

und somit

$$\|p_k\| \leq \frac{2}{\hat{c}} \left\|\Pi_{D(M;x^*)}(\nabla f(x_k))\right\| \quad \text{für alle } k \geq \tilde{k}.$$

Weiterhin ist

$$\nabla f(x^*)^T v = 0 \quad \text{für alle } v \in D(M; x^*),$$

da

$$-\nabla f(x^*)^T x = -\nabla f(x^*)^T x^* \quad \text{für alle } x \in E(M; -\nabla f(x^*))$$

ist, das heißt, daß $\nabla f(x^*)$ senkrecht auf dem Vektorraum $D(M; x^*)$ und damit insbesondere auf allen Vektoren von dessen Basis steht. Nach Lemma 5.3.2 gilt somit

$$\Pi_{D(M;x^*)}(\nabla f(x^*)) = 0.$$

Zusammen mit der Konvergenz der Folge $(x_k)_{k\in\mathbb{N}}$ gegen x^*, der stetigen Differenzierbarkeit von f auf einer offenen Obermenge von M und der Stetigkeit von Projektionen erhalten wir nun

$$0 \leq \lim_{k\to\infty} \|p_k\| \leq \frac{2}{\hat{c}} \lim_{k\to\infty} \left\|\Pi_{D(M;x^*)}(\nabla f(x_k))\right\| = 0$$

und der Beweis ist beendet. √

■

Der Satz über die Konvergenzgeschwindigkeit des Verfahrens benötigt allerdings weitere Voraussetzungen an die Menge M und an die Zwischenschritte. So fordern wir, daß

$$\text{span}(\{c_0, \ldots, c_m\}) = \mathbb{R}^n \tag{5.3.6}$$

und

$$\{l_i, u_i\} \not\subseteq \{-\infty, \infty\} \quad \text{für alle } 1 \leq i \leq m \tag{5.3.7}$$

ist. Die zweite Forderung bedeutet also, daß entweder l_i oder u_i reellwertig sein muß. Andernfalls könnte man die erste Forderung umgehen, indem man bei dem jeweiligen c_i die Schranken $l_i := -\infty$ und $u_i := \infty$ wählt. Bezüglich der Zwischenschritte stellen wir neben den Forderungen (4.2.5) und (4.2.6) eine weitere Bedingung und zwar an den letzten Zwischenschritt x_k^{r+1}. Wie in Bemerkung 4.2.2 bereits erwähnt wurde definieren wir $p_k := x_k^{r+1} - x_k$, um so die Zwischenschritte in das Newton-Verfahren einzubetten. Wie im Beweis zu Satz 4.2.3 seien jetzt wieder $\{v_1, \ldots, v_s\} \subset \mathbb{R}^n$ mit $1 \leq s \leq n$ eine Orthonormalbasis des Vektorraums

$$V_k^j := \{v \in \mathbb{R}^n \mid {c_i}^T v = 0 \text{ für alle } i \in I(M; x_k^j)\} \tag{5.3.8}$$

und

$$A_k^j := \begin{pmatrix} v_1 & \dots & v_s \end{pmatrix} \in \mathbb{R}^{n \times s} \tag{5.3.9}$$

die Matrix, die die Vektoren dieser Orthonormalbasis als Spalten hat. Es ist dann s die Dimension dieses Vektorraums. Wir fordern nun, daß für eine Folge $(\tau_k)_{k \in \mathbb{N}} \subset \mathbb{R}_+ \cup \{0\}$

$$\left\| A_k^{r\,T} (\nabla f(x_k) + \nabla^2 f(x_k) p_k) \right\| \leq \tau_k \left\| A_k^{r\,T} \nabla f(x_k) \right\| \quad \text{für alle } k \in \mathbb{N} \tag{5.3.10}$$

gilt.

Lemma 5.3.4 *Seien die Voraussetzungen von Satz 5.3.3 erfüllt und die Forderungen (5.3.6) und (5.3.7) sichergestellt. Sei weiterhin eine Folge*

$$(\tau_k)_{k \in \mathbb{N}} \subset \mathbb{R}_+ \cup \{0\}$$

gewählt. Dann läßt sich mit Hilfe einer geeigneten Wahl von Zwischenschritten Forderung (5.3.10) für alle hinreichend großen $k \in \mathbb{N}$ erfüllen.

Beweis. Wir werden ein Verfahren kennenlernen, mit dem man sicherstellt, daß spätestens mit dem letzten Zwischenschritt x_k^{r+1} und der Definition $p_k := x_k^{r+1} - x_k$ die Forderung (5.3.10) erfüllt ist. Ist diese Forderung schon vorher als

$$\left\| A_k^{j-1\,T} (\nabla f(x_k) + \nabla^2 f(x_k) p_k) \right\| \leq \tau_k \left\| A_k^{j-1\,T} \nabla f(x_k) \right\|$$

für ein $2 \leq j \leq r$ und $p_k := x_k^j - x_k$ sichergestellt, so wählt man einfach $x_k^{r+1} := x_k^j$ sowie $x_k^r := \ldots := x_k^j := x_k^{j-1}$ und die Forderungen (4.2.5) sowie (4.2.6) sind erfüllt. Denn da wir $x_k + p_k \in M$ und $\|p_k\| \leq \mu_1 \Delta_k$ als primäre Forderungen an den Schritt p_k stellen, ist nicht ohne weiteres einsehbar, daß eine soche Wahl überhaupt möglich ist. Im Beweis zu Satz 5.3.3 haben wir Ungleichung (5.3.4) gezeigt. Das heißt, daß die Hessesche Matrix $\nabla^2 f(x_k)$ für alle hinreichend großen $k \in \mathbb{N}$ auf $D(M; x^*)$ positiv definit ist. Nach dem gleichen Satz gilt auch

$$\|p_k\| < \mu_1 \Delta_k$$

für alle hinreichend großen $k \in \mathbb{N}$, wählt man den Schritt p_k nach den Iterationsvorschriften des Newton-Verfahrens beliebig aus. Wir betrachten ein solch hinreichend großes $k \in \mathbb{N}$, das diesen beiden Eigenschaften genügt. Seien also bereits die Zwischenschritte

$$\{x_k^1 := x_k^C, \ldots, x_k^j\} \subset M$$

mit $1 \leq j \leq r$ berechnet, von denen keiner mit Hilfe der Definition $p_k := x_k^j - x_k$ Forderung (5.3.10) erfülle.

- Nach Satz 4.2.3 lassen sich die Bedingungen (4.2.5) und (4.2.6) in die dazu hinreichenden Bedingungen (4.2.7) und (4.2.8) umformen. Es reicht also, diese neuen Bedingungen zu zeigen, falls wir einen Zwischenschritt x_k^{j+1} finden, der Forderung (5.3.10) erfüllt. Als erstes suchen wir ein Minimum der Funktion $q_k : \mathbb{R}^n \to \mathbb{R}$, definiert durch

$$q_k(x) := \nabla f(x_k)^T(x - x_k) + \frac{1}{2}(x - x_k)^T \nabla^2 f(x_k)(x - x_k),$$

auf $x_k^j + V_k^j$. Dieses Minimum existiert, da k so hinreichend groß gewählt wurde, daß $\nabla^2 f(x_k)$ auf $D(M; x^*)$ positiv definit ist. Haben wir ein solches $x_k^{j+1} \in x_k^j + V_k^j$ gefunden, so ist (5.3.10) erfüllt. Denn gilt $x_k^{j+1} = x_k^j + A_k^j w$ für ein $w \in \mathbb{R}^s$, dann ist der Gradient der Funktion $q : \mathbb{R}^s \to \mathbb{R}$, definiert durch

$$q(w) := q_k(x_k^j + A_k^j w) - q_k(x_k^j),$$

im Punkt w gleich Null und für diesen gilt

$$\begin{aligned} 0 = \nabla q(w) =& {A_k^j}^T(\nabla f(x_k) + \nabla^2 f(x_k)(x_k^j + A_k^j w - x_k)) \\ =& {A_k^j}^T(\nabla f(x_k) + \nabla^2 f(x_k)(x_k^{j+1} - x_k)) \\ =& {A_k^j}^T(\nabla f(x_k) + \nabla^2 f(x_k) p_k). \end{aligned}$$

Außerdem folgt aus der letzten Gleichung

$${A_k^j}^T(\nabla f(x_k) + \nabla^2 f(x_k)(x_k^j - x_k)) = -{A_k^j}^T \nabla^2 f(x_k) A_k^j w$$

und die Hessesche Matrix

$$\nabla^2 q(w) = {A_k^j}^T \nabla^2 f(x_k) A_k^j$$

ist positiv semidefinit. Daraus erhalten wir nun mit Hilfe des Mittelwertsatzes

$$\begin{aligned} 0 =& \nabla q(w)^T w \\ =& (\nabla q(0) + \nabla^2 q(w) w)^T w \\ =& \nabla q(0)^T w + w^T {A_k^j}^T \nabla^2 f(x_k) A_k^j w, \end{aligned}$$

also

$$\nabla q(0)^T w = -w^T {A_k^j}^T \nabla^2 f(x_k) A_k^j w \leq 0.$$

Somit ergibt sich Forderung (4.2.8) durch

$$\begin{aligned} q(w) =& (A_k^{j\,T}(\nabla f(x_k) + \nabla^2 f(x_k)(x_k^j - x_k)))^T w + \frac{1}{2} w^T A_k^{j\,T} \nabla^2 f(x_k) A_k^j w \\ =& - w^T A_k^{j\,T} \nabla^2 f(x_k) A_k^j w + \frac{1}{2} w^T A_k^{j\,T} \nabla^2 f(x_k) A_k^j w \\ =& - \frac{1}{2} w^T A_k^{j\,T} \nabla^2 f(x_k) A_k^j w \\ <& - \mu_0 w^T A_k^{j\,T} \nabla^2 f(x_k) A_k^j w \\ =& \mu_0 \nabla q(0)^T w \\ =& \mu_0 \min\{\nabla q(0)^T w, 0\}. \end{aligned}$$

Ist nun noch $x_k^{j+1} \in M$, so ist auch (4.2.7) erfüllt. Denn mit Hilfe der Definition $p_k := x_k^{j+1} - x_k$ wird Forderung (5.1.1) genügt und am Anfang des Beweises wurde $k \in \mathbb{N}$ so gewählt, daß auch für alle nachfolgenden $k \in \mathbb{N}$ die Ungleichung $\|p_k\| < \mu_1 \Delta_k$ gilt. $\surd$

- Ist allerdings $x_k^{j+1} \notin M$, so wählen* wir als x_k^{j+1} einen Punkt, der die Forderungen (4.2.5) und (4.2.6) erfüllt, aber eine aktive Restriktion mehr besitzt als x_k^j. Da es insgesamt nur $r+1$ Zwischenschritte gibt, von denen der erste x_k^C ist, sind im Fall der Definition

$$r := m + 1$$

spätestens nach r Zwischenschritten alle m Restriktionen aktiv. Da die Forderungen (5.3.6) und (5.3.7) vorausgesetzt werden, enthält der Vektorraum V_k^r demnach nur die Null und damit ist Forderung (5.3.10) trivialerweise erfüllt. $\surd$

■

Der jetzt folgende Satz stellt das Hauptresultat zur Analyse der Konvergenzgeschwindigkeit des Newton-Verfahrens dar. Er nennt die Voraussetzungen, unter denen das Verfahren mindestens linear oder superlinear konvergiert.

Satz 5.3.5 *Sei $f : \mathbb{R}^n \to \mathbb{R}$ eine auf einer offenen Obermenge von M zweimal stetig differenzierbare Abbildung. Das Newton-Verfahren breche nicht vorzeitig ab und sei $(x_k)_{k\in\mathbb{N}}$ die Folge, die durch dieses Verfahren erzeugt wird. Sei weiterhin $\nabla^2 f$ auf der Niveaumenge $L(M; x_0)$ beschränkt und seien die Forderungen (4.2.5) und (4.2.6) sowie (5.3.10) an die Zwischenschritte erfüllt. Sei $x^* \in M$ der Grenzwert der Folge $(x_k)_{k\in\mathbb{N}}$ und ein stationärer Punkt der*

*Die Existenz eines solchen Punktes konnten wir letztendlich nur ansatzweise sicherstellen.

Optimierungsaufgabe (P). Sei weiterhin die Hessesche Matrix $\nabla^2 f(x^*)$ auf der Menge $D(M;x^*)$ positiv definit und es sei

$$\tau := \limsup_{k\to\infty} \tau_k.$$

Dann konvergiert die Folge $(x_k)_{k\in\mathbb{N}}$ gegen x^* mindestens linear, falls τ hinreichend klein ist und mindestens superlinear, falls $\tau = 0$ ist.

Beweis. Der Beweis erfolgt in mehreren Teilen.

- Als erstes werden wir

  (5.3.11) $V(x) := \{v \in \mathbb{R}^n \mid {c_i}^T v = 0 \text{ für alle } i \in I(M;x)\} \subseteq D(M;x^*)$

  für alle $x \in E(M;-\nabla f(x^*))$ zeigen. Sei dazu $x \in E(M;-\nabla f(x^*))$ beliebig gewählt und wir bemerken zunächst, daß wir am Anfang von Abschnitt 4.1 mit Hilfe der Lagrange-Multiplikatoren $\{\lambda_1,\ldots,\lambda_m\} \subset \mathbb{R}$ für den stationären Punkt x^* die Darstellung der exponierten Seitenfläche

  $$E(M;-\nabla f(x^*)) = \left\{ z \in M \;\middle|\; \begin{array}{l} {c_i}^T z = l_i \text{ falls } \lambda_i > 0 \text{ und} \\ {c_i}^T z = u_i \text{ falls } \lambda_i < 0 \text{ für alle } 1 \le i \le m \end{array} \right\}$$

  hergeleitet haben und so folgern wir nun leicht

  $$\{i \in \{1,\ldots,m\} \mid \lambda_i \neq 0\} \subseteq I(M;x).$$

  Sei jetzt $v \in V(x)$ beliebig gewählt. Es gilt also

  $$\nabla f(x^*)^T v = \sum_{i=1}^{m} \lambda_i ({c_i}^T v) = 0$$

  und da jedes $v \in V(x)$ eine zulässige Richtung an M in x ist, erhalten wir $x + tv \in M$ für alle hinreichend kleinen $t > 0$ und somit

  $$x + tv \in E(M;-\nabla f(x^*))$$

  für alle hinreichend kleinen $t > 0$. Da aber $x \in E(M;-\nabla f(x^*))$ gilt und $D(M;x^*)$ ein Vektorraum ist, haben wir für ein beliebiges hinreichend kleines $t > 0$

  $$v = \frac{1}{t}((x + tv - x^*) - (x - x^*)) \in D(M;x^*)$$

  und somit die gewünschte Inklusion. $\surd$

- Die Voraussetzungen aus Satz 5.3.3 sind auch hier erfüllt und wie am Anfang des Beweises zu jenem Satz gezeigt wurde existiert ein Iterationsindex $\hat{k} \in \mathbb{N}$, so daß

$$\{x_k^1 := x_k^C, \ldots, x_k^{r+1}\} \subset E(M; -\nabla f(x^*)) \quad \text{für alle } k \geq \hat{k}$$

ist. Aus Inklusion (5.3.11) folgt zusammen mit der Definition des Vektorraums V_k^r in (5.3.8)

$$V_k^r = V(x_k^r) \subseteq D(M; x^*) \quad \text{für alle } k \geq \hat{k}.$$

Sei nun ein beliebiger Iterationsindex $k \geq \hat{k}$ gegeben. Nach Lemma 5.3.2 und mit der Definition der Matrix A_k^r in (5.3.9) ist die Projektion auf V_k^r durch

$$\Pi_{V_k^r}(x) = A_k^r {A_k^r}^T x \quad \text{für alle } x \in \mathbb{R}^n.$$

gegeben. Da auch zu $D(M; x^*)$ eine Orthonormalbasis existiert, können wir Lemma 5.3.2 auch auf $D(M; x^*)$ anwenden und wir haben, da V_k^r nach (5.3.11) in $D(M; x^*)$ enthalten ist,

$$\begin{aligned} \left\|\Pi_{D(M;x^*)}(x) - x\right\| &= \inf_{y \in D(M;x^*)} \|y - x\| \\ &\leq \inf_{y \in V_k^r} \|y - x\| \\ &= \left\|\Pi_{V_k^r}(x) - x\right\| \quad \text{für alle } x \in \mathbb{R}^n. \end{aligned}$$

Mit Lemma 5.3.2, Teil (ii) erhalten wir

$$\begin{aligned} \left\|\Pi_{V_k^r}(x)\right\|^2 &= \|x\|^2 - \left\|x - \Pi_{V_k^r}(x)\right\|^2 \\ &\leq \|x\|^2 - \left\|x - \Pi_{D(M;x^*)}(x)\right\|^2 \\ &= \left\|\Pi_{D(M;x^*)}(x)\right\|^2 \quad \text{für alle } x \in \mathbb{R}^n \end{aligned}$$

und somit

$$\left\|\Pi_{V_k^r}(x)\right\| \leq \left\|\Pi_{D(M;x^*)}(x)\right\| \quad \text{für alle } x \in \mathbb{R}^n. \tag{5.3.12}$$

√

- Jetzt zeigen wir

$$\limsup_{k \to \infty} \frac{\left\|\Pi_{V_k^r}(\nabla f(x_{k+1}))\right\|}{\left\|\Pi_{D(M;x^*)}(\nabla f(x_k))\right\|} \leq \limsup_{k \to \infty} \tau_k. \tag{5.3.13}$$

Als erstes definieren wir die Folge $(\varepsilon_k)_{k \in \mathbb{N}}$ durch

$$\varepsilon_k := \frac{\left\|{A_k^r}^T(\nabla f(x_k + p_k) - \nabla f(x_k) - \nabla^2 f(x_k) p_k)\right\|}{\|p_k\|}$$

und es ist

$$\begin{aligned}
\varepsilon_k &= \left\| A_k^{r\,T} \int_0^1 (\nabla^2 f(x_k + tp_k) - \nabla^2 f(x_k)) \left(\frac{p_k}{\|p_k\|} \right) \, dt \right\| \\
&\leq \underbrace{\left\| A_k^{r\,T} \right\|}_{=1} \left\| \int_0^1 (\nabla^2 f(x_k + tp_k) - \nabla^2 f(x_k)) \left(\frac{p_k}{\|p_k\|} \right) \, dt \right\| \\
&\leq \int_0^1 \left\| \nabla^2 f(x_k + tp_k) - \nabla^2 f(x_k) \right\| \left(\frac{\|p_k\|}{\|p_k\|} \right) \, dt \\
&\leq \int_0^1 \left(\left\| \nabla^2 f(x_k + tp_k) - \nabla^2 f(x^*) \right\| + \left\| \nabla^2 f(x^*) - \nabla^2 f(x_k) \right\| \right) \, dt.
\end{aligned}$$

für alle $k \in \mathbb{N}$. Aufgrund der Konvergenz von $(x_k)_{k\in\mathbb{N}}$ gegen x^* und von $(p_k)_{k\in\mathbb{N}}$ gegen Null folgt die Konvergenz von $(\varepsilon_k)_{k\in\mathbb{N}}$ gegen Null. Nach Satz 5.3.3 existiert ein $\bar{k} \in \mathbb{N}$, so daß alle Iterationsindizes $k \geq \bar{k}$ erfolgreich sind. Da die Matrizen A_k^r für alle $k \in \mathbb{N}$ die Vektoren von Orthonormalbasen als Spalten haben, gilt $\|A_k^r x\| = \|x\|$ für alle $x \in \mathbb{R}^n$ und alle $k \in \mathbb{N}$. Mit unserer in diesem Satz neuen Voraussetzung (5.3.10) an die Zwischenschritte erhalten wir jetzt

$$\begin{aligned}
&\left\| \Pi_{V_k^r}(\nabla f(x_{k+1})) \right\| \\
=& \left\| A_k^r A_k^{r\,T} \nabla f(x_k + p_k) \right\| \\
=& \left\| A_k^{r\,T} \nabla f(x_k + p_k) \right\| \\
\leq& \left\| A_k^{r\,T} (\nabla f(x_k + p_k) - \nabla f(x_k) - \nabla^2 f(x_k) p_k) \right\| \\
&+ \left\| A_k^{r\,T} (\nabla f(x_k) + \nabla^2 f(x_k) p_k) \right\| \\
\leq& \varepsilon_k \|p_k\| + \tau_k \left\| A_k^{r\,T} \nabla f(x_k) \right\| \\
=& \varepsilon_k \|p_k\| + \tau_k \left\| A_k^r A_k^{r\,T} \nabla f(x_k) \right\| \\
=& \varepsilon_k \|p_k\| + \tau_k \left\| \Pi_{V_k^r}(\nabla f(x_k)) \right\| \qquad \text{für alle } k \geq \bar{k}.
\end{aligned} \tag{5.3.14}$$

Am Ende des Beweises zu Satz 5.3.3 haben wir gezeigt, daß ein $\tilde{k} \geq \hat{k}$ und eine Konstante $\hat{c} > 0$ existieren, so daß

$$\|p_k\| \leq \frac{2}{\hat{c}} \left\| \Pi_{D(M;x^*)}(\nabla f(x_k)) \right\| \qquad \text{für alle } k \geq \tilde{k}$$

ist. Zusammen ergibt sich nun mit (5.3.12)

$$\begin{aligned}
&\limsup_{k\to\infty} \frac{\left\|\Pi_{V_k^r}(\nabla f(x_{k+1}))\right\|}{\left\|\Pi_{D(M;x^*)}(\nabla f(x_k))\right\|} \\
\leq &\limsup_{k\to\infty} \frac{\varepsilon_k \|p_k\| + \tau_k \left\|\Pi_{V_k^r}(\nabla f(x_k))\right\|}{\left\|\Pi_{D(M;x^*)}(\nabla f(x_k))\right\|} \\
\leq &\limsup_{k\to\infty} \frac{\frac{2}{\hat{c}}\varepsilon_k \left\|\Pi_{D(M;x^*)}(\nabla f(x_k))\right\| + \tau_k \left\|\Pi_{D(M;x^*)}(\nabla f(x_k))\right\|}{\left\|\Pi_{D(M;x^*)}(\nabla f(x_k))\right\|} \\
= &\limsup_{k\to\infty} \left(\frac{2}{\hat{c}}\varepsilon_k + \tau_k\right) \\
\leq &\frac{2}{\hat{c}} \limsup_{k\to\infty} \varepsilon_k + \limsup_{k\to\infty} \tau_k \\
= &\limsup_{k\to\infty} \tau_k .
\end{aligned}$$

$\surd$

- Wir schätzen nun den Zähler aus (5.3.13) ab und zeigen die Existenz eines Iterationsindex $\check{k} \in \mathbb{N}$ und einer Folge $(\delta_k)_{k\in\mathbb{N}} \subset \mathbb{R}_+$ mit Grenzwert Null, für die

  (5.3.15) $\quad \left\|\Pi_{V_k^r}(\nabla f(x_{k+1}))\right\| \geq (\hat{c} - \delta_k) \|x_{k+1} - x^*\| \quad$ für alle $k \geq \check{k}$

  gilt. Dazu benötigen wir einige Vorbemerkungen.

  (i) Als erstes zeigen wir

  $$\Pi_{V_k^r}(\nabla f(x^*)) = 0 \quad \text{für alle } k \in \mathbb{N}.$$

  Am Ende des Beweises von Satz 5.3.3 haben wir bemerkt, daß

  $$\nabla f(x^*)^T v = 0 \quad \text{für alle } v \in D(M; x^*)$$

  ist und damit gilt diese Aussage insbesondere für alle $v \in V_k^r$. Mit Lemma 5.3.2 ergibt sich

  $$\begin{aligned}
  \Pi_{V_k^r}(\nabla f(x^*)) =& A_k^r {A_k^r}^T \nabla f(x^*) \\
  =& A_k^r (\underbrace{\nabla f(x^*)^T A_k^r}_{=0})^T \\
  =& 0 \quad \text{für alle } k \in \mathbb{N}.
  \end{aligned}$$

  $\surd$

(ii) Als nächstes bemerken wir, daß ein Iterationsindex $\check{k} \in \mathbb{N}$ existiert, für den

$$x_{k+1} - x^* \in V_k^r \quad \text{für alle } k \geq \check{k}$$

gilt. Denn es sind alle Iterationsindizes $k \geq \bar{k}$ erfolgreich und daher ist $x_{k+1} = x_k + p_k = x_k^{r+1}$ für alle $k \geq \bar{k}$, womit wir zusammen mit Forderung (4.2.5)

$$I(M; x_k^r) \sqsubset I(M; x_{k+1}) \quad \text{für alle } k \geq \bar{k}$$

bekommen. Außerdem existiert ein Iterationsindex $\check{k} \geq \bar{k}$, so daß

$$I(M; x_k^r) \sqsubset I(M; x^*) \quad \text{für alle } k \geq \check{k}$$

ist, denn da $(x_k)_{k\in\mathbb{N}}$ gegen x^* konvergiert, ist eine nicht aktive Komponente von x^* auch bei x_k für alle hinreichend großen $k \in \mathbb{N}$ nicht aktiv und zusammen mit der ersten Inklusion erhält man diese Inklusion. Somit bekommen wir $x_{k+1} \in V_k^r$ und $x^* \in V_k^r$ für alle $k \geq \check{k}$ und da V_k^r ein Vektorraum ist, gilt auch $x_{k+1} - x^* \in V_k^r$ für alle $k \geq \check{k}$. $\surd$

(iii) Sei jetzt wieder $k \in \mathbb{N}$ beliebig gewählt. Als weitere Aussage zeigen wir

$$\left\|\Pi_{V_k^r}(\nabla^2 f(x^*)v)\right\| \geq \hat{c}\,\|v\| \quad \text{für alle } v \in V_k^r.$$

Da die Matrix $\nabla^2 f(x^*)$ auf der Menge $D(M; x^*)$ positiv definit ist, ist sie dies auch insbesondere auf V_k^r und es existiert eine Konstante $\hat{c} > 0$, für die

$$p^T \nabla^2 f(x^*) p \geq \hat{c}\|p\|^2 \quad \text{für alle } p \in D(M; x^*)$$

gilt. Daraus folgt aufgrund der Identität $\Pi_{V_k^r}(v) = v$ für alle $v \in V_k^r$ aus

$$\begin{aligned}
\hat{c}\|v\|^2 &\leq \left\|v^T \nabla^2 f(x^*) v\right\| \\
&= \left\|\left(A_k^r A_k^{r\,T} v\right)^T \nabla^2 f(x^*) v\right\| \\
&= \left\|v^T A_k^r A_k^{r\,T} \nabla^2 f(x^*) v\right\| \\
&\leq \left\|A_k^r A_k^{r\,T} \nabla^2 f(x^*) v\right\| \|v\| \\
&= \left\|\Pi_{V_k^r}(\nabla^2 f(x^*) v)\right\| \|v\| \quad \text{für alle } v \in V_k^r
\end{aligned}$$

die Ungleichung

$$\left\|\Pi_{V_k^r}(\nabla^2 f(x^*)v)\right\| \geq \hat{c}\,\|v\| \quad \text{für alle } v \in V_k^r.$$

$\surd$

Jetzt haben wir alle Zwischenbemerkungen zum Beweis von (5.3.15) erledigt und wir definieren die Folge $(\delta_k)_{k\in\mathbb{N}}$ durch

$$\delta_k := \frac{\left\|A_k^r A_k^{r\,T}(\nabla f(x_{k+1}) - \nabla f(x^*) - \nabla^2 f(x^*)(x_{k+1}-x^*))\right\|}{\|x_{k+1}-x^*\|}.$$

Es gilt

$$\begin{aligned}
\delta_k &= \left\| A_k^{r\,T} \int_0^1 (\nabla^2 f(x^* + t(x_{k+1}-x^*)) - \nabla^2 f(x^*)) \left(\frac{x_{k+1}-x^*}{\|x_{k+1}-x^*\|}\right) dt \right\| \\
&\le \underbrace{\left\|A_k^{r\,T}\right\|}_{=1} \left\| \int_0^1 (\nabla^2 f(x^* + t(x_{k+1}-x^*)) - \nabla^2 f(x^*)) \left(\frac{x_{k+1}-x^*}{\|x_{k+1}-x^*\|}\right) dt \right\| \\
&\le \int_0^1 \left\|(\nabla^2 f(x^*) + t(x_{k+1}-x^*)) - \nabla^2 f(x^*)\right\| dt \qquad \text{für alle } k \in \mathbb{N},
\end{aligned}$$

so daß zusammen mit der Konvergenz von $(x_k)_{k\in\mathbb{N}}$ gegen x^* die Konvergenz von $(\delta_k)_{k\in\mathbb{N}}$ gegen Null folgt. Daraus erhalten wir mit (i)

$$\begin{aligned}
&\left\|\Pi_{V_k^r}(\nabla^2 f(x^*)(x_{k+1}-x^*))\right\| \\
&= \left\|\Pi_{V_k^r}(\nabla f(x_{k+1})) + \Pi_{V_k^r}(\nabla f(x_{k+1}) - \nabla f(x^*) - \nabla^2 f(x^*)(x_{k+1}-x^*))\right\| \\
&\le \left\|\Pi_{V_k^r}(\nabla f(x_{k+1}))\right\| \\
&\quad + \left\|\Pi_{V_k^r}(\nabla f(x_{k+1}) - \nabla f(x^*) - \nabla^2 f(x^*)(x_{k+1}-x^*))\right\| \\
&= \left\|\Pi_{V_k^r}(\nabla f(x_{k+1}))\right\| \\
&\quad + \left\|A_k^r A_k^{r\,T}(\nabla f(x_{k+1}) - \nabla f(x^*) - \nabla^2 f(x^*)(x_{k+1}-x^*))\right\| \\
&\le \left\|\Pi_{V_k^r}(\nabla f(x_{k+1}))\right\| + \delta_k \|x_{k+1}-x^*\| \qquad \text{für alle } k \in \mathbb{N}.
\end{aligned}$$

Da nach (ii) die Inklusion $x_{k+1} - x^* \in V_k^r$ für alle $k \ge \check{k}$ gilt, ergibt sich mit (iii)

$$\left\|\Pi_{V_k^r}(\nabla^2 f(x^*)(x_{k+1}-x^*))\right\| \ge \hat{c}\,\|x_{k+1}-x^*\| \qquad \text{für alle } k \ge \check{k}.$$

Wir erhalten Ungleichung (5.3.15) mit Hilfe der letzten beiden Ungleichungen durch

$$\begin{aligned}
\hat{c}\,\|x_{k+1}-x^*\| &\le \left\|\Pi_{V_k^r}(\nabla^2 f(x^*)(x_{k+1}-x^*))\right\| \\
&\le \left\|\Pi_{V_k^r}(\nabla f(x_{k+1}))\right\| + \delta_k \|x_{k+1}-x^*\| \qquad \text{für alle } k \ge \check{k}.
\end{aligned}$$

√

- Wir schätzen jetzt den Nenner aus (5.3.13) ab und zeigen die Existenz einer Konstanten $\tilde{c} > 0$ und einer Folge $(\eta_k)_{k\in\mathbb{N}} \subset \mathbb{R}_+$ mit Grenzwert Null, für die

$$\left\|\Pi_{D(M;x^*)}(\nabla f(x_k))\right\| \le (\tilde{c} + \eta_k)\,\|x_k - x^*\| \qquad \text{für alle } k \in \mathbb{N} \tag{5.3.16}$$

gilt. Sei dazu $\{v_1, \ldots, v_s\} \subset \mathbb{R}^n$ mit $1 \leq s \leq n$ eine Orthonormalbasis des Vektorraums $D(M; x^*)$ und

$$A := \begin{pmatrix} v_1 & \ldots & v_s \end{pmatrix} \in \mathbb{R}^{n \times s}$$

die Matrix, die die Vektoren dieser Orthonormalbasis als Spalten hat. Nach Lemma 5.3.2 ist dann die Projektion auf $D(M; x^*)$ gegeben durch

$$\Pi_{D(M;x^*)}(v) = AA^T v \quad \text{für alle } v \in \mathbb{R}^n.$$

und wir erhalten

$$\begin{aligned} \Pi_{D(M;x^*)}(\nabla f(x^*)) =& AA^T \nabla f(x^*) \\ =& A(\underbrace{\nabla f(x^*)^T A}_{=0})^T \\ =& 0. \end{aligned}$$

Wir definieren jetzt die Folge $(\eta_k)_{k \in \mathbb{N}}$ durch

$$\eta_k := \frac{\left\| AA^T(\nabla f(x_k) - \nabla f(x^*) - \nabla^2 f(x^*)(x_k - x^*)) \right\|}{\|x_k - x^*\|}$$

und es gilt

$$\begin{aligned} \eta_k &= \left\| A^T \int_0^1 (\nabla^2 f(x^* + t(x_k - x^*)) - \nabla^2 f(x^*)) \left(\frac{x_k - x^*}{\|x_k - x^*\|} \right) dt \right\| \\ &\leq \underbrace{\left\| A^T \right\|}_{=1} \left\| \int_0^1 (\nabla^2 f(x^* + t(x_k - x^*)) - \nabla^2 f(x^*)) \left(\frac{x_k - x^*}{\|x_k - x^*\|} \right) dt \right\| \\ &\leq \int_0^1 \left\| (\nabla^2 f(x^*) + t(x_k - x^*)) - \nabla^2 f(x^*) \right\| dt \quad \text{für alle } k \in \mathbb{N}, \end{aligned}$$

so daß zusammen mit der Konvergenz von $(x_k)_{k \in \mathbb{N}}$ gegen x^* die Konvergenz von $(\eta_k)_{k \in \mathbb{N}}$ gegen Null folgt. Jetzt erhalten wir mit $x_k - x^* \in D(M; x^*)$ für alle $k \geq \hat{k}$ und damit $AA^T(x_k - x^*) = x_k - x^*$ für alle $k \geq \hat{k}$ die Ungleichungskette

$$\begin{aligned} & \left\| \Pi_{D(M;x^*)}(\nabla f(x_k)) \right\| \\ \leq & \left\| \Pi_{D(M;x^*)}(\nabla^2 f(x^*)(x_k - x^*)) \right\| \\ & + \left\| \Pi_{D(M;x^*)}(\nabla f(x_k) - \nabla f(x^*) - \nabla^2 f(x^*)(x_k - x^*)) \right\| \\ \leq & \left\| AA^T \nabla^2 f(x^*) AA^T (x_k - x^*) \right\| \\ & + \left\| \Pi_{D(M;x^*)}(\nabla f(x_k) - \nabla f(x^*) - \nabla^2 f(x^*)(x_k - x^*)) \right\| \\ \leq & \left\| AA^T \nabla^2 f(x^*) AA^T \right\| \|x_k - x^*\| + \eta_k \|x_k - x^*\| \\ = & \left(\left\| AA^T \nabla^2 f(x^*) AA^T \right\| + \eta_k \right) \|x_k - x^*\| \quad \text{für alle } k \in \mathbb{N} \end{aligned}$$

und daraus (5.3.16) mit der Definition

$$\tilde{c} := \left\| AA^T \nabla^2 f(x^*) AA^T \right\|.$$

√

- Jetzt haben wir alle Abschätzungen, die wir zum Beweis der Aussage des Satzes benötigen. Setzt man (5.3.15) und (5.3.16) in (5.3.13) ein, so ergibt sich

$$\begin{aligned}
\limsup_{k\to\infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} &\leq \limsup_{k\to\infty} \frac{(\tilde{c} + \eta_k) \left\| \Pi_{V_k^r}(\nabla f(x_{k+1})) \right\|}{(\hat{c} - \delta_k) \left\| \Pi_{D(M;x^*)}(\nabla f(x_k)) \right\|} \\
&\leq \left(\limsup_{k\to\infty} \frac{\tilde{c} + \eta_k}{\hat{c} - \delta_k} \right) \left(\limsup_{k\to\infty} \frac{\left\| \Pi_{V_k^r}(\nabla f(x_{k+1})) \right\|}{\left\| \Pi_{D(M;x^*)}(\nabla f(x_k)) \right\|} \right) \\
&\leq \left(\frac{\tilde{c}}{\hat{c}} \right) \limsup_{k\to\infty} \tau_k \\
&= \left(\frac{\tilde{c}}{\hat{c}} \right) \tau.
\end{aligned}$$

Hieraus erkennt man, daß die Folge $(x_k)_{k\in\mathbb{N}}$ gegen x^* mindestens linear konvergiert, falls

$$\tau < \frac{\hat{c}}{\tilde{c}}$$

ist und mindestens superlinear, falls $\tau = 0$ ist. √

■

Es ist auch möglich, quadratische Konvergenz des Newton-Verfahrens zu zeigen. Wir benötigen dabei aber eine weitere Voraussetzung, von der wir nicht wissen, ob sie praktisch erfüllbar ist.

Korollar 5.3.6 *Seien die Voraussetzungen aus Satz 5.3.5 gegeben. Sei weiterhin $\nabla^2 f$ auf der Niveaumenge $L(M; x_0)$ Lipschitz-stetig, das heißt es existiere eine Konstante $\bar{c} > 0$, so daß*

$$\left\| \nabla^2 f(x) - \nabla^2 f(y) \right\| \leq \bar{c} \|x - y\| \qquad \textit{für alle } x, y \in L(M; x_0)$$

ist. Außerdem existiere ein Konstante $\check{c} > 0$, so daß

$$\tau_k \leq \check{c} \left\| \Pi_{V_k^r}(\nabla f(x_k)) \right\| \qquad \textit{für alle } k \in \mathbb{N}$$

gilt. Dann konvergiert die Folge $(x_k)_{k\in\mathbb{N}}$ mindestens quadratisch gegen x^*.

Beweis. Da der Beweis in großen Teilen identisch zum Beweis von Satz 5.3.5 ist, geben wir nur die Stellen an, die sich ändern. Wir ändern Ungleichung (5.3.13) in

$$\limsup_{k\to\infty} \frac{\left\|\Pi_{V_k^r}(\nabla f(x_{k+1}))\right\|}{\left\|\Pi_{D(M;x^*)}(\nabla f(x_k))\right\|^2} \leq \bar{c}\left(\frac{2}{\hat{c}}\right)^2 + \check{c}.$$

Denn für jedes $k \in \mathbb{N}$ existiert nach dem Mittelwertsatz ein $0 < t_k < 1$, so daß

$$\nabla f(x_k + p_k) - \nabla f(x_k) = \nabla^2 f(x_k + t_k p_k) p_k$$

gilt. Mit den zusätzlichen Voraussetzungen modifizieren wir die Ungleichungskette (5.3.14) nun zu

$$\begin{aligned}
\left\|\Pi_{V_k^r}(\nabla f(x_{k+1}))\right\| &= \left\|A_k^r A_k^{r\,T} \nabla f(x_k + p_k)\right\| \\
&= \left\|A_k^{r\,T} \nabla f(x_k + p_k)\right\| \\
&\leq \left\|A_k^{r\,T}(\nabla f(x_k + p_k) - \nabla f(x_k) - \nabla^2 f(x_k) p_k)\right\| \\
&\quad + \left\|A_k^{r\,T}(\nabla f(x_k) + \nabla^2 f(x_k) p_k)\right\| \\
&= \left\|A_k^{r\,T}(\nabla^2 f(x_k + t_k p_k) - \nabla^2 f(x_k)) p_k\right\| \\
&\quad + \left\|A_k^{r\,T}(\nabla f(x_k) + \nabla^2 f(x_k) p_k)\right\| \\
&\leq \underbrace{\left\|A_k^{r\,T}\right\|}_{=1} \bar{c} t_k \|p_k\|^2 + \tau_k \left\|A_k^{r\,T} \nabla f(x_k)\right\| \\
&\leq \bar{c}\|p_k\|^2 + \tau_k \left\|A_k^r A_k^{r\,T} \nabla f(x_k)\right\| \\
&= \bar{c}\|p_k\|^2 + \tau_k \left\|\Pi_{V_k^r}(\nabla f(x_k))\right\| \qquad \text{für alle } k \geq \bar{k}
\end{aligned}$$

und mit zusammen mit (5.3.12) ergibt sich nun

$$\begin{aligned}
&\limsup_{k\to\infty} \frac{\left\|\Pi_{V_k^r}(\nabla f(x_{k+1}))\right\|}{\left\|\Pi_{D(M;x^*)}(\nabla f(x_k))\right\|^2} \\
\leq &\limsup_{k\to\infty} \frac{\bar{c}\|p_k\|^2 + \tau_k \left\|\Pi_{V_k^r}(\nabla f(x_k))\right\|}{\left\|\Pi_{D(M;x^*)}(\nabla f(x_k))\right\|^2} \\
\leq &\limsup_{k\to\infty} \frac{\bar{c}\left(\frac{2}{\hat{c}}\left\|\Pi_{D(M;x^*)}(\nabla f(x_k))\right\|\right)^2 + \check{c}\left\|\Pi_{V_k^r}(\nabla f(x_k))\right\|^2}{\left\|\Pi_{D(M;x^*)}(\nabla f(x_k))\right\|^2} \\
= &\limsup_{k\to\infty} \frac{\bar{c}\left(\frac{2}{\hat{c}}\left\|\Pi_{D(M;x^*)}(\nabla f(x_k))\right\|\right)^2 + \check{c}\left\|\Pi_{D(M;x^*)}(\nabla f(x_k))\right\|^2}{\left\|\Pi_{D(M;x^*)}(\nabla f(x_k))\right\|^2} \\
= &\limsup_{k\to\infty} \left(\bar{c}\left(\frac{2}{\hat{c}}\right)^2 + \check{c}\right) \\
= &\bar{c}\left(\frac{2}{\hat{c}}\right)^2 + \check{c}.
\end{aligned}$$

Setzt man nun (5.3.15) und (5.3.16) in unsere obige Ungleichung ein, so ergibt sich

$$\begin{aligned}\limsup_{k\to\infty} \frac{\|x_{k+1}-x^*\|}{\|x_k-x^*\|^2} &\leq \limsup_{k\to\infty} \frac{(\tilde{c}+\eta_k)\left\|\Pi_{V_k^r}(\nabla f(x_{k+1}))\right\|}{(\hat{c}-\delta_k)\left\|\Pi_{D(M;x^*)}(\nabla f(x_k))\right\|^2} \\ &\leq \left(\limsup_{k\to\infty} \frac{\tilde{c}+\eta_k}{(\hat{c}-\delta_k)^2}\right)\left(\limsup_{k\to\infty} \frac{\left\|\Pi_{V_k^r}(\nabla f(x_{k+1}))\right\|}{\left\|\Pi_{D(M;x^*)}(\nabla f(x_k))\right\|^2}\right) \\ &\leq \left(\frac{\tilde{c}}{\hat{c}^2}\right)\left(\bar{c}\left(\frac{2}{\hat{c}}\right)^2+\check{c}\right).\end{aligned}$$

Hieraus erkennt man, daß die Folge $(x_k)_{k\in\mathbb{N}}$ mindestens quadratisch gegen x^* konvergiert. ■

Kapitel 6

Numerische Tests

In diesem Kapitel werden die numerischen Ergebnisse präsentiert, die mit einer Implementation des Newton-Verfahrens in MATLAB erzielt wurden. Wir beschränken uns bei dieser Betrachtung auf die Optimierungsaufgabe (Q), das heißt den Box-restringierten Fall. In unseren Tests versuchen wir, Satz 5.2.8 und Satz 5.3.5 an drei Beispielfunktionen zu verifizieren. Dabei verändern wir die Startpunkte x_0 sowie den wichtigen Parameter des Verfahrens μ_0. Für die Folge $(\tau_k)_{k\in\mathbb{N}}$ wählen wir der Einfachheit halber verschiedene Konstanten τ und versuchen herauszufinden, welche Konvergenzgeschwindigkeit gemäß Satz 5.3.5 erzielt wird. Das Maß für den Aufwand des Verfahrens ist die Anzahl an Iterationen, die benötigt werden, um eine gewisse Genauigkeit im Ergebnis zu erreichen. Die Quelltexte des Programms sind in Abschnitt 6.3 zu finden.

6.1 Testaufbau

Wir betrachten also die Optimierungsaufgabe (Q). Als Testfunktionen für die Zielfunktion $f : \mathbb{R}^n \to \mathbb{R}$ benutzen wir

- die Beliebig-dimensionale Funktion,
- die Erweiterte Rosenbrock-Funktion,
- die Wood-Funktion.

Die Funktionen stammen aus Kapitel C von Geiger and Kanzow, 1999. Sie haben alle die Gestalt

$$f(x) = \sum_{i=1}^{r} (F_i(x))^2$$

mit $r \in \mathbb{N} \setminus \{0\}$ und den Funktionen $F_i : \mathbb{R}^n \to \mathbb{R}$ für $1 \leq i \leq r$. Alle Funktionen sind Polynome in mehreren Veränderlichen und damit zweimal

stetig differenzierbar. Wir wählen als Menge $[l, u]$ das n-dimensionale Intervall

$$[l, u] := \left[\begin{pmatrix} -\infty \\ \vdots \\ -\infty \end{pmatrix}, \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \right],$$

denn alle drei Funktionen besitzen in $(1, \ldots, 1)$ einen stationären Punkt der Optimierungsaufgabe (Q). Um letztendlich die Aussagen der Sätze 5.2.8 und 5.3.5 zu verifizieren, müssen wir sicherstellen, daß die jeweilige Hessesche Matrix $\nabla^2 f(1, \ldots, 1)$ auf der Menge $D([l, u]; (1, \ldots, 1))$ positiv definit ist. In allen drei Fällen ist $\nabla^2 f(1, \ldots, 1)$ sogar auf dem ganzen $\mathbb{R}^n$ positiv definit.

Testfunktion 6.1.1 (Beliebig-dimensionale Funktion) Es darf $n \in \mathbb{N} \setminus \{0\}$ beliebig gewählt werden und es gilt $r := n+2$. Für die Funktionen $F_i : \mathbb{R}^n \to \mathbb{R}$ mit $1 \leq i \leq r$ gilt

$$F_i(x) := \begin{cases} x_i - 1 & \text{falls } 1 \leq i \leq n, \\ \sum_{j=1}^{n} j(x_j - 1) & \text{falls } i = n+1, \\ \left(\sum_{j=1}^{n} j(x_j - 1) \right)^2 & \text{falls } i = n+2. \end{cases}$$

Testfunktion 6.1.2 (Erweiterte Rosenbrock-Funktion) Bei dieser Funktion muß $n \in \mathbb{N} \setminus \{0\}$ gerade sein und es ist $r := n$. Es gilt für die Funktionen $F_{2i-1} : \mathbb{R}^n \to \mathbb{R}$ mit $1 \leq i \leq \frac{r}{2}$

$$F_{2i-1}(x) := 10(x_{2i} - {x_{2i-1}}^2)$$

und für die Funktionen $F_{2i} : \mathbb{R}^n \to \mathbb{R}$ mit $1 \leq i \leq \frac{r}{2}$

$$F_{2i}(x) := 1 - x_{2i-1}.$$

Testfunktion 6.1.3 (Wood-Funktion) Bei dieser Funktion ist $n := 4$ und $r := 6$. Wir haben also in diesem Fall nicht die Möglichkeit, ein hochdimensionales Beispiel zu konstruieren. Es gilt für die Funktionen $F_i : \mathbb{R}^n \to \mathbb{R}$ mit $1 \leq i \leq 6$

$$\begin{aligned} F_1(x) :=& 10(x_2 - {x_1}^2), \\ F_2(x) :=& 1 - x_1, \\ F_3(x) :=& \sqrt{90}(x_4 - {x_3}^2), \\ F_4(x) :=& 1 - x_3, \\ F_5(x) :=& \sqrt{10}(x_2 + x_4 - 2), \\ F_6(x) :=& \frac{1}{\sqrt{10}}(x_2 - x_4). \end{aligned}$$

Nun kommen wir zum Newton-Verfahren selbst. Wir setzen die Konstanten

$$\mu_1 := 1, \quad \rho_0 := 10^{-3}, \quad \rho_1 := 0.25, \quad \rho_2 := 0.75$$

und

$$\sigma_1 := 0.25, \quad \sigma_2 := 0.5, \quad \sigma_3 := 1.5$$

und wählen für die Konstante μ_0 mehrere Werte, da sie sowohl bei der Berechnung des Cauchy-Schritts als auch bei der Berechnung der Zwischenschrittte wichtig ist und daher im Verfahren eine große Rolle spielt. Wir wählen $n := 4$ bei allen Funktionen und $n := 16$ bei der Beliebig-dimensionalen Funktion und der Erweiterten Rosenbrock-Funktion. Neben der Testreihe haben wir mit den beiden letztgenannten Funktionen noch einzelne Tests mit höheren Werten für n durchgeführt, wegen der hohen Rechendauer der Implementation in MATLAB jedoch auf die Wahl verschiedener Parameter verzichtet und somit nicht in die Tabellen eingefügt. Für die verschiedenen Werte von n suchen wir nun verschiedene Startwerte $x_0 \in [l, u]$ aus. Für $n := 16$ lassen wir wir außer den Startwerten $(0, \ldots, 0)$ und $(-5, \ldots, -5)$ einen Zufallsgenerator weitere Startwerte x_0 erzeugen und mitteln dann über die Anzahl an Iterationen. Wir definieren nun

$$\Delta_0 := \left\| \nabla_{[l,u]} f(x_0) \right\|.$$

Das Verfahren bricht ab, sobald

$$\left\| \nabla_{[l,u]} f(x_k) \right\| < 10^{-2}$$

ist. Wir setzen die Konstanten zur Berechnung des Cauchy-Schritts

$$\gamma_1 := 1, \quad \gamma_2 := 0.8, \quad \gamma_3 := 4, \quad \gamma_4 := 1.25,$$

wobei γ_4 eine neu eingeführte Konstante darstellt. Erfüllt $\alpha_k := \gamma_1$ Forderung (3.3.3) schon von Anfang an, so wird α_k solange mit γ_4 multipliziert, bis es im darauffolgenden Schritt Forderung (3.3.5) erfüllt oder größer als γ_3 ist. Ansonsten greift das Verfahren aus Satz 3.3.2, also die Multiplikation mit γ_2. Der Einfachheit halber wählen wir in unserer Implementation für die Folge $(\tau_k)_{k \in \mathbb{N}}$ die konstante Folge

$$\tau_k := \tau \quad \text{für alle } k \in \mathbb{N}$$

mit einer Konstanten $\tau \in \mathbb{R}_+ \cup \{0\}$, die während der Testreihe verschiedene Werte annimmt. Wir überprüfen nun, ob das Verfahren konvergiert. Sobald die Folge $(x_k)_{k \in \mathbb{N}}$ einen Häufungspunkt besitzt, konvergiert sie nach Satz 5.2.8 auch. Um die Konvergenzordung zu ermitteln, berechnen wir die Quotienten

$$\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} \quad \text{für alle } k \in \mathbb{N}$$

für verschiedene Werte von τ. Nach Satz 5.3.5 sollten wir bei hinreichend kleinen $\tau > 0$ mindestens lineare Konvergenz und im Fall $\tau := 0$ sogar mindestens superlineare Konvergenz der Folge $(x_k)_{k\in\mathbb{N}}$ erhalten. In Lemma 5.3.4 suchen wir ein Minimum der Funktion $q_k : \mathbb{R}^n \to \mathbb{R}$, definiert durch

$$q_k(x) := \nabla f(x_k)^T(x - x_k) + \frac{1}{2}(x - x_k)^T\nabla^2 f(x_k)(x - x_k),$$

auf $x_k^j + V_k^j$. Da es zwar nicht für die Rechenzeit, aber für die Anzahl an Iterationen unerheblich ist, wie man dieses Minimum bestimmt, verwenden wir die Funktion `'fmincon'` aus MATLAB, wobei wir das Ergebnis wieder auf die Menge $[l, u]$ projizieren, da andernfalls das Verfahren nicht weiterarbeiten kann.

6.2 Testergebnisse

Wir stellen die Ergebnisse für jede der Funktionen einzeln dar. In den beiden linken Spalten sind die verschiedenen Werte für n und die verschiedenen Startpunkte x_0 angegeben. Jede darauffolgende Spalte steht für ein Paar der Konstanten (μ_0, τ) und enthält in jedem Eintrag die Anzahl an Iterationen, die das Newton-Verfahren bis zur Erfüllung der Abbruchbedingung benötigt. Bei allen Tests konnten wir die Anzahl an Iterationen angeben, denn die Grundvoraussetzung dafür, nämlich die Konvergenz gegen den stationären Punkt $(1, \ldots, 1)$, war immer erfüllt.

Ergebnis 6.2.1 (Beliebig-dimensionale Funktion) Bei der Testreihe mit der Beliebig-dimensionalen Funktion fällt auf, daß die Anzahl an Iterationen kaum von der Wahl der Konstanten μ_0 und τ abhängt. Die Iterationsschritte waren jedoch verschieden und es gab geringe Abweichungen. Auffällig ist, daß bei den höherdimensionalen Tests die Anzahl an Iterationen im Unterschied zu den anderen Testfunktionen ansteigt, jedoch geringer als n, was durch Tests mit $n := 64$ bestätigt wird. Die Konvergenzgeschwindigkeit konnte nicht genau ermittelt werden, jedoch war die Folge der Quotienten der Form

$$\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} \text{ mit } k \in \mathbb{N}$$

ab einem gewissen $k \in \mathbb{N}$ absteigend und die letzten 5 bis 6 Iterationschritte lagen nahe bei Null, so daß wir superlineare Konvergenz zumindest nicht ausschließen können.

Beliebig-dimensionale Funktion					
$\mu_0 :=$		10^{-2}		10^{-1}	
$\tau :=$		10^{-3}	1	10^{-3}	1
$n := 4$	$(0,0,0,0)$	8	8	8	8
	$(-5,0,0,0)$	8	8	9	9
	$(-5,-5,0,0)$	10	10	10	10
	$(-5,-5,-5,0)$	11	11	11	11
	$(-5,-5,-5,-5)$	12	12	12	12
$n := 16$	$(0,\ldots,0)$	14	14	14	14
	$\varnothing$ von 5 Tests	17	17	17	17
	$(-5,\ldots,-5)$	18	19	19	19

Ergebnis 6.2.2 (Erweiterte Rosenbrock-Funktion) Bei der Testreihe mit der Erweiterten Rosenbrock-Funktion sieht man hingegen, daß die Anzahl an Iterationen zwar kaum von μ_0, jedoch von τ abhängt. Diesmal gibt es auch keine nennenswerte Abhängigkeit von der Dimension des Problems, auch für $n := 64$ benötigt das Newton-Verfahren eine ähnliche Anzahl an Iterationen. Auch bei dieser Testreihe konnte die Konvergenzgeschwindigkeit nicht genau bestimmt werden, jedoch lagen die Quotienten der Form

$$\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} \text{ mit } k \in \mathbb{N}$$

für die letzten 4 bis 5 Iterationschritte nahe bei Null.

Erweiterte Rosenbrock-Funktion					
$\mu_0 :=$		10^{-2}		10^{-1}	
$\tau :=$		10^{-3}	1	10^{-3}	1
$n := 4$	$(0,0,0,0)$	16	17	16	17
	$(-5,0,0,0)$	38	38	38	38
	$(-5,-5,0,0)$	38	35	38	35
	$(-5,-5,-5,0)$	40	39	40	39
	$(-5,-5,-5,-5)$	39	35	39	35
$n := 16$	$(0,\ldots,0)$	16	16	16	17
	$\varnothing$ von 5 Tests	35	35	35	34
	$(-5,\ldots,-5)$	39	35	39	35

Ergebnis 6.2.3 (Wood-Funktion) Leider ließ sich die Wood-Funktion nur für $n := 4$ testen, da sie nur für diese Dimension definiert ist. Auch hier erkennt man eine stärkere Abhängigkeit von τ als von μ_0.

Wood-Funktion					
$\mu_0 :=$		10^{-2}		10^{-1}	
$\tau :=$		10^{-3}	1	10^{-3}	1
$n := 4$	$(0,0,0,0)$	6	6	6	6
	$(-5,0,0,0)$	9	9	9	9
	$(-5,-5,0,0)$	9	8	9	8
	$(-5,-5,-5,0)$	9	10	9	8
	$(-5,-5,-5,-5)$	9	8	9	8

6.3 Implementation

In diesem Abschnitt werden wir einige Anmerkungen zur Implementation des Newton-Verahrens in MATLAB machen und abschließend die Quelltexte angeben. Zunächst ist zu beachten, daß der aktuelle Iterationspunkt mit `'z'` und der gerade berechnete Iterationspunkt mit `'z_neu'` bezeichnet wird. Das hat den Grund, daß MATLAB die Variable `'x'` für seine Unterfunktion `'fmincon'` so verwendet, daß sie im Rest der Funktion `'zwischenschritte'` nicht mehr benutzt werden kann. Außerdem verwenden wir während des gesamten Programms Zeilenvektoren statt Spaltenvektoren als Standardvektoren.

```
%*************************************************************
% Funktion 'newton'
%
% Beschreibung: Implementation des Newton-Verfahrens in Matlab
%
%*************************************************************


function k = newton(zielfunktion , z , mu_0 , tau)

%-------------------------------------------------------------
% 'k' ist die Anzahl an Iterationen, die das Newton-Verfahren
% benoetigt. 'zielfunktion' ist das Handle der Zielfunktion, 'z'
% ist ein Punkt der Menge '[l , u]'.

%-------------------------------------------------------------
% Datei oeffnen
datei = fopen('ergebnis' , 'w');

%-------------------------------------------------------------
% Konstanten
%mu_0 = 1e-2;
mu_1 = 1;
rho_0 = 1e-3;
rho_1 = 0.25;
rho_2 = 0.75;
sigma_1 = 0.25;
sigma_2 = 0.5;
sigma_3 = 1.5;
%tau = 1e-3;
abbruch_konstante = 1e-2;

%-------------------------------------------------------------
% Beliebig-dimensionale Funktion, Nummer 6
%zielfunktion = @beliebig;
%z = [0 0 0 0];

%-------------------------------------------------------------
% Erweiterte Rosenbrock-Funktion, Nummer 14
%zielfunktion = @rosenbrock;
%z = [0 0 0 0];
```

```
%-------------------------------------------------------------
% Wood-Funktion, Nummer 17
%zielfunktion = @wood;
%z = [0 0 0 0];

%-------------------------------------------------------------
% Alle verwendeten Funktionen ausser der Wood-Funktion
%n = 16;
%rand('state' , sum(100 * clock));
%z = -5 * rand(1 , n);

%-------------------------------------------------------------
% Stationaerer Punkt aller verwendeten Funktionen
z_stationaer = ones(1 , length(z));

%-------------------------------------------------------------
% Initialisierung
%.............................................................
k = 0;
[f , g , H] = feval(zielfunktion , z);
delta = norm(g);
abbruch = norm(projizierter_gradient(zielfunktion , z));
%.............................................................
fprintf(datei , 'Funktion ''%s''\n\n' , func2str(zielfunktion));
fprintf(datei , '\t mu_0 = %16.8f\n' , mu_0);
fprintf(datei , '\t tau = %16.8f\n\n' , tau);
fprintf(datei , '%d. Iterationsschritt\n' , k);
fprintf(datei , '\t z = %10.2f %10.2f %10.2f %10.2f\n' , z);
fprintf(datei , '\n');
fprintf(datei , '\t f = %10.2f\n' , f);
fprintf(datei , '\t delta = %10.2f\n' , delta);
fprintf(datei , '\t norm(pg) = %10.2f\n\n\n' , abbruch);

%-------------------------------------------------------------
% Schleife
while (abbruch > abbruch_konstante)
%.............................................................
     k = k + 1;
     p_C = cauchy_schritt(zielfunktion , z , delta , mu_0 ,
          mu_1);
     fprintf(datei , 'Zwischenschritte des %d.
```

```
          Iterationsschritts\n\n' , k);
     z_neu = zwischenschritte(zielfunktion , z , p_C , delta ,
          mu_1 , tau , datei);
%.....................................................................
     [f , g , H] = feval(zielfunktion , z);
     [f_neu , g_neu , H_neu] = feval(zielfunktion , z_neu);
     chi = (f - f_neu)
          / (f - modellfunktion(zielfunktion , z , z_neu - z));
     if (chi > rho_0)
          quotient = norm(z_neu - z_stationaer)
               / norm(z - z_stationaer);
          z = z_neu;
          f = f_neu;
          g = g_neu;
          H = H_neu;
     else
  quotient = 1;
     end
     if (chi <= rho_1)
          delta = sigma_1 * delta;
     elseif (chi >= rho_2)
          delta = sigma_3 * delta;
     else
          delta = sigma_2 * delta;
     end
     abbruch = norm(projizierter_gradient(zielfunktion , z));
%.....................................................................
     fprintf(datei , '%d. Iterationsschritt\n' , k);
     fprintf(datei , '\t z = %10.2f %10.2f %10.2f %10.2f\n' ,
          z);
     fprintf(datei , '\n');
     fprintf(datei , '\t f = %10.2f\n' , f);
     fprintf(datei , '\t delta = %10.2f\n' , delta);
     fprintf(datei , '\t norm(pg) = %10.2f\n' , abbruch);
     fprintf(datei , '\t quotient = %10.2f\n\n\n' , quotient);
end

%---------------------------------------------------------------------
% Datei schliessen
fclose(datei);
```

```
%*************************************************************
```

```
%*****************************************************************
% Funktion 'cauchy_schritt'
%
% Beschreibung: Cauchy-Schritt des Newton-Verfahrens
%
%*****************************************************************


function p_C = cauchy_schritt(zielfunktion , z , delta , mu_0 ,
     mu_1)

%----------------------------------------------------------------
% 'gamma_1' , 'gamma_2' , 'gamma_3' sind die Konstanten zur
% Berechnung des Cauchy-Schritts aus Abschnitt 3.3. Erfuellt
% 'alpha = gamma_1' Forderung (3.3.3) schon von Anfang an, so
% wird 'alpha' solange mit 'gamma_4' multipliziert, bis es im
% darauffolgenden Schritt Forderung (3.3.5) erfuellt oder
% groesser als 'gamma_3' ist. Ansonsten greift das Verfahren aus
% Satz 3.3.2, also die Multiplikation mit 'gamma_2'.
% 'zielfunktion' ist das Handle der Zielfunktion, 'z' ist ein
% Punkt der Menge '[l , u]'.

%----------------------------------------------------------------
% Konstanten
gamma_1 = 1;
gamma_2 = 0.8;
gamma_3 = 4;
gamma_4 = 1.25;

%----------------------------------------------------------------
% Initialisierung
[f , g , H] = feval(zielfunktion , z);
alpha = gamma_1;
pi = projektion(z - alpha * g) - z;
mf = modellfunktion(zielfunktion , z , pi);

%----------------------------------------------------------------
% Schleife
%................................................................
while ((mf - f <= mu_0 * (g * pi')) & (norm(pi) <= mu_1 * delta)
     & (alpha <= gamma_3))
```

```
     alpha = alpha * gamma_4;
     pi = projektion(z - alpha * g) - z;
     mf = modellfunktion(zielfunktion , z , pi);
end
%.....................................................
alpha = alpha / gamma_4;
pi = projektion(z - alpha * g) - z;
mf = modellfunktion(zielfunktion , z , pi);
%.....................................................
while ((mf - f > mu_0 * (g * pi')) | (norm(pi) > mu_1 * delta))
     alpha = alpha * gamma_2;
     pi = projektion(z - alpha * g) - z;
     mf = modellfunktion(zielfunktion , z , pi);
end

%-------------------------------------------------------------
% Ergebnis
p_C = pi;


%*************************************************************
```

```
%****************************************************************
% Funktion 'modellfunktion'
%
% Beschreibung: Modellfunktion der Zielfunktion des
%               Newton-Verfahrens
%
%****************************************************************


function mf = modellfunktion(zielfunktion , z , p)

%---------------------------------------------------------------
% 'zielfunktion' ist das Handle der Zielfunktion, 'z' und 'p'
% sind Punkte des 'R^n'.

%---------------------------------------------------------------
% Initialisierung
[f , g , H] = feval(zielfunktion , z);

%---------------------------------------------------------------
% Ergebnis
mf = f + g * p' + (p * H * p') / 2;


%****************************************************************
```

```
%*****************************************************************
% Funktion 'projektion'
%
% Beschreibung: Projektion auf die Menge '[l , u]'
%
%*****************************************************************


function y = projektion(z)

%----------------------------------------------------------------
% 'z' ist ein Punkt des 'R^n'.

%----------------------------------------------------------------
% Initialisierung
n = length(z);
l = schranke(z , 0);
u = schranke(z , 1);

%----------------------------------------------------------------
% Ergebnis
for (i = 1 : n)
     if (z(i) < l(i))
          y(i) = l(i);
     elseif (z(i) > u(i))
          y(i) = u(i);
     else
          y(i) = z(i);
     end
end


%*****************************************************************
```

```
%****************************************************************
% Funktion 'projizierter_gradient'
%
% Beschreibung: Projizierter Gradient der Zielfunktion des
%               Newton-Verfahrens
%
%****************************************************************


function pg = projizierter_gradient(zielfunktion , z)

%----------------------------------------------------------------
% 'zielfunktion' ist das Handle der Zielfunktion, 'z' ist ein
% Punkt des 'R^n'.

%----------------------------------------------------------------
% Initialisierung
n = length(z);
l = schranke(z , 0);
u = schranke(z , 1);
[f , g , H] = feval(zielfunktion , z);

%----------------------------------------------------------------
% Ergebnis
for (i = 1 : n)
     if (z(i) == l(i))
          pg(i) = max(- g(i) , 0);
     elseif (z(i) == u(i))
          pg(i) = min(- g(i) , 0);
     else
          pg(i) = - g(i);
     end
end


%****************************************************************
```

```
%*****************************************************************
% Funktion 'schranke'
%
% Beschreibung: Schranken der Menge '[l , u]'
%
%*****************************************************************


function s = schranke(z , i)

%----------------------------------------------------------------
% 'i = 0' gibt als Vektor die Schranke 'l' aus, 'i = 1' gibt als
% Vektor die Schranke 'u' aus. 'z' ist ein Punkt des 'R^n'.

%----------------------------------------------------------------
% Initialisierung
n = length(z);

%----------------------------------------------------------------
% Ergebnis
if (i == 0)
     s = -Inf * ones(1 , n);
else
     s = ones(1 , n);
end


%*****************************************************************
```

```
%*****************************************************************
% Funktion 'zwischenschritte'
%
% Beschreibung: Zwischenschritte des Newton-Verfahrens
%
%*****************************************************************


function zwischenschritt = zwischenschritte(zielfunktion , z ,
     p_C , delta , mu_1 , tau , datei)

%----------------------------------------------------------------
% Die Zwischenschritte werden nach der Beschreibung in Abschnitt
% 4.2 und Lemma 5.3.4 berechnet. 'zielfunktion' ist das Handle
% der Zielfunktion, 'z' ist ein Punkt der Menge '[l , u]'.

%----------------------------------------------------------------
% Optionen der Unterfunktion 'fmincon'
options = optimset;
%options = optimset('LargeScale' , 'off');

%----------------------------------------------------------------
% Initialisierung
%................................................................
j = 2;
n = length(z);
l = schranke(z , 0);
u = schranke(z , 1);
z_neu = z + p_C;
[f , g , H] = feval(zielfunktion , z);
%................................................................
A = eye(n);
lb = l;
ub = u;
for (i = 1 : n)
     if ((z_neu(i) == l(i)) | (z_neu(i) == u(i)))
          A(i , i) = 0;
     end
end
i = 1;
while (i <= size(A , 1))
```

```
     if (norm(A(i , :)) == 0)
          A(i , :) = [];
  lb(i) = [];
  ub(i) = [];
  i = i - 1;
     end
     i = i + 1;
end
%.....................................................
x_start = zeros(1 , size(A , 1));
if (~isempty(A))
     x = fmincon(@problemfunktion , x_start , [] , [] , [] ,
          [] , lb , ub , @trustregiongrenze , options , z , g ,
          H , z_neu , A , delta , mu_1);
     z_neu = projektion((z_neu)' + A' * x');
end
%.....................................................
fprintf(datei , '%d. Zwischenschritt\n' , j);
fprintf(datei , '\t z_neu = %10.2f %10.2f %10.2f %10.2f\n' ,
     z_neu);
fprintf(datei , '\n');
for (i = 1 : size(A , 1))
     fprintf(datei , '\t A(%d , :) = %d %d %d %d %d %d %d %d' ,
          i , A(i , :));
     fprintf(datei , '\n');
end
fprintf(datei , '\n');

%-----------------------------------------------------------
% Schleife
while ((~isempty(A)) & (j <= n + 1)
     & (norm(A * (g' + H * (z_neu - z)')) > tau * norm(A * g')))
%.....................................................
     j = j + 1;
     A = eye(n);
     lb = l;
     ub = u;
     for (i = 1 : n)
          if ((z_neu(i) == l(i)) | (z_neu(i) == u(i)))
               A(i , i) = 0;
          end
     end
```

```
     i = 1;
     while (i <= size(A , 1))
          if (norm(A(i , :)) == 0)
               A(i , :) = [];
       lb(i) = [];
       ub(i) = [];
       i = i - 1;
          end
          i = i + 1;
     end
%.....................................................................
     x_start = zeros(1 , size(A , 1));
     if (~isempty(A))
          x = fmincon(@problemfunktion , x_start , [] , [] ,
               [] , [] , lb , ub , @trustregiongrenze ,
               options , z , g , H , z_neu , A , delta , mu_1);
          z_neu = projektion((z_neu)' + A' * x');
     end
%.....................................................................
     fprintf(datei , '%d. Zwischenschritt\n' , j);
     fprintf(datei ,
          '\t z_neu = %10.2f %10.2f %10.2f %10.2f\n' , z_neu);
     fprintf(datei , '\n');
     for (i = 1 : size(A , 1))
          fprintf(datei ,
               '\t A(%d , :) = %d %d %d %d %d %d %d %d' , i ,
               A(i , :));
          fprintf(datei , '\n');
     end
     fprintf(datei , '\n');
end

%---------------------------------------------------------------
% Ergebnis
zwischenschritt = z_neu;


%===============================================================
% Unterfunktion 'problemfunktion'
%
% Beschreibung: Funktion 'q' in Lemma 5.3.4.
%
```

```
%=============================================================

function [f , g , H] = problemfunktion(x , z , g , H , z_neu ,
     A , delta , mu_1)

%-------------------------------------------------------------
% Ergebnis
f = (A * (g' + H * (z_neu - z)'))' * x'
     + 1 / 2 * x * A * H * A' * x';
if (nargout > 1)
     g = (A * (g' + H * (z_neu - z)'))' + A * H * A' * x';
elseif (nargout > 2)
     H = A * H * A';
end


%=============================================================
% Unterfunktion 'trustregiongrenze'
%
% Beschreibung: Forderung 4.2.7.
%
%=============================================================

function [kleinergleich , gleich] = trustregiongrenze(x , z ,
     g , H , z_neu , A , delta , mu_1)

%-------------------------------------------------------------
% Ergebnis
%.............................................................
kleinergleich = norm((z_neu)' + A' * x' - z') - mu_1 * delta;
%.............................................................
gleich = [];


%*************************************************************
```

```
%*****************************************************************
% Funktion 'funktionsauswertung'
%
% Beschreibung: Ausgabe des Funktionswertes der Zielfunktion des
%               Newton-Verfahrens zusammen mit ihrem Gradienten,
%               ihrem projizierten Gradienten und ihrer
%               Hesseschen Matrix, auch in Diagonalform
%
%*****************************************************************


function funktionsauswertung(zielfunktion , z)

%----------------------------------------------------------------
% 'zielfunktion' ist das Handle der Zielfunktion, 'z' ist ein
% Punkt des 'R^n'.

%----------------------------------------------------------------
% Datei oeffnen
datei = fopen('zielfunktion' , 'w');

%----------------------------------------------------------------
% Beliebig-dimensionale Funktion, Nummer 6
%zielfunktion = @beliebig;
%z = [1 1 1 1];

%----------------------------------------------------------------
% Erweiterte Rosenbrock-Funktion, Nummer 14
%zielfunktion = @rosenbrock;
%z = [1 1 1 1];

%----------------------------------------------------------------
% Wood-Funktion, Nummer 17
%zielfunktion = @wood;
%z = [1 1 1 1];

%----------------------------------------------------------------
% Initialisierung
[f , g , H] = feval(zielfunktion , z);
pg = projizierter_gradient(zielfunktion , z);
[U , D] = eig(H);
```

```
%-------------------------------------------------------------
% Ausgabe
fprintf(datei , 'Funktion ''%s''\n\n' , func2str(zielfunktion));
fprintf(datei , 'z = %10.2f %10.2f %10.2f %10.2f\n' , z);
fprintf(datei , '\n\n');
fprintf(datei , 'f = %10.2f' , f);
fprintf(datei , '\n\n');
fprintf(datei , 'g = %10.2f %10.2f %10.2f %10.2f\n' , g);
fprintf(datei , '\n\n');
fprintf(datei , 'pg = %10.2f %10.2f %10.2f %10.2f\n' , pg);
fprintf(datei , '\n\n');
for (i = 1 : size(H , 1))
     fprintf(datei , 'H(%d , :) = %10.2f %10.2f %10.2f %10.2f
          %10.2f %10.2f %10.2f %10.2f' , i , H(i , :));
     fprintf(datei , '\n');
end
fprintf(datei , '\n');
for (i = 1 : size(D , 1))
     fprintf(datei , 'D(%d , :) = %10.2f %10.2f %10.2f %10.2f
          %10.2f %10.2f %10.2f %10.2f' , i , D(i , :));
     fprintf(datei , '\n');
end

%-------------------------------------------------------------
% Datei schliessen
fclose(datei);


%*************************************************************
```

```
%*****************************************************************
% Funktion 'beliebig'
%
% Beschreibung: Beliebig-dimensionale Funktion, Nummer 6
%
%*****************************************************************


function [f , g , H] = beliebig(z)

%----------------------------------------------------------------
% 'z' ist ein Punkt des 'R^n', 'f' der Funktionswert im Punkt
% 'z', 'g' der Gradient im Punkt 'z' und 'H' die Hessesche
% Matrix im Punkt 'z'.

%----------------------------------------------------------------
% Initialisierung
n = length(z);
r = n + 2;
F = zeros(r);
dF = zeros(r , n);
ddF = zeros(r , n , n);

%----------------------------------------------------------------
% Partielle Ableitungen
%................................................................
for (p = 1 : n)
     F(p) = z(p) - 1;
     dF(p , p) = 1;
end
%................................................................
F(n + 1) = 0;
for (p = 1 : n)
     F(n + 1) = F(n + 1) + p * (z(p) - 1);
     dF(n + 1 , p) = p;
end
%................................................................
F(n + 2) = (F(n + 1))^2;
for (p = 1 : n)
     dF(n + 2 , p) = 2 * F(n + 1) * p;
     for (q = 1 : n)
```

```
          ddF(n + 2 , p , q) = 2 * q * p;
     end
end

%-------------------------------------------------------------
% Ergebnis
%.............................................................
f = 0;
for (p = 1 : r)
     f = f + (F(p))^2;
end
%.............................................................
for (i = 1 : n)
     g(i) = 0;
     for (p = 1 : r)
          g(i) = g(i) + 2 * F(p) * dF(p , i);
     end
end
%.............................................................
for (i = 1 : n)
     for (j = 1 : n)
          H(i , j) = 0;
          for (p = 1 : r)
               H(i , j) = H(i , j) + 2 * (dF(p , i) * dF(p , j)
                    + F(p) * ddF(p , i , j));
          end
     end
end


%*************************************************************
```

```
%*************************************************************
% Funktion 'rosenbrock'
%
% Beschreibung: Erweiterte Rosenbrock-Funktion, Nummer 14
%
%*************************************************************


function [f , g , H] = rosenbrock(z)

%-------------------------------------------------------------
% 'z' ist ein Punkt des 'R^n', 'f' der Funktionswert im Punkt
% 'z', 'g' der Gradient im Punkt 'z' und 'H' die Hessesche
% Matrix im Punkt 'z'.

%-------------------------------------------------------------
% Initialisierung
n = length(z);
r = n;
F = zeros(r);
dF = zeros(r , n);
ddF = zeros(r , n , n);

%-------------------------------------------------------------
% Partielle Ableitungen
for (q = 1 : r / 2)
%.............................................................
     F(2 * q - 1) = 10 * (z(2 * q) - (z(2 * q - 1))^2);
     dF(2 * q - 1 , 2 * q - 1) = -20 * z(2 * q - 1);
     dF(2 * q - 1 , 2 * q) = 10;
     ddF(2 * q - 1 , 2 * q - 1 , 2 * q - 1) = -20;
%.............................................................
     F(2 * q) = 1 - z(2 * q - 1);
     dF(2 * q , 2 * q - 1) = -1;
end

%-------------------------------------------------------------
% Ergebnis
%.............................................................
f = 0;
for (p = 1 : r)
```

```
     f = f + (F(p))^2;
end
%.............................................................
for (i = 1 : n)
     g(i) = 0;
     for (p = 1 : r)
          g(i) = g(i) + 2 * F(p) * dF(p , i);
     end
end
%.............................................................
for (i = 1 : n)
     for (j = 1 : n)
          H(i , j) = 0;
          for (p = 1 : r)
               H(i , j) = H(i , j) + 2 * (dF(p , i) * dF(p , j)
                    + F(p) * ddF(p , i , j));
          end
     end
end


%*************************************************************
```

```
%*****************************************************************
% Funktion 'wood'
%
% Beschreibung: Wood-Funktion, Nummer 17
%
%*****************************************************************


function [f , g , H] = wood(z)

%-----------------------------------------------------------------
% 'z' ist ein Punkt des 'R^n', 'f' der Funktionswert im Punkt
% 'z', 'g' der Gradient im Punkt 'z' und 'H' die Hessesche
% Matrix im Punkt 'z'.

%-----------------------------------------------------------------
% Initialisierung
n = 4;
r = 6;
F = zeros(r);
dF = zeros(r , n);
ddF = zeros(r , n , n);

%-----------------------------------------------------------------
% Partielle Ableitungen
F(1) = 10 * (z(2) - (z(1))^2);
dF(1 , 1) = -20 * z(1);
dF(1 , 2) = 10;
ddF(1 , 1 , 1) = -20;
%.................................................................
F(2) = 1 - z(1);
dF(2 , 1) = -1;
%.................................................................
F(3) = sqrt(90) * (z(4) - (z(3))^2);
dF(3 , 3) = -2 * sqrt(90) * z(3);
dF(3 , 4) = sqrt(90);
ddF(3 , 3 , 3) = -2 * sqrt(90);
%.................................................................
F(4) = 1 - z(3);
dF(4 , 3) = -1;
%.................................................................
```

```
F(5) = sqrt(10) * (z(2) + z(4) - 2);
dF(5 , 2) = sqrt(10);
dF(5 , 4) = sqrt(10);
%.....................................................
F(6) = 1 / sqrt(10) * (z(2) - z(4));
dF(6 , 2) = 1 / sqrt(10);
dF(6 , 4) = -1 / sqrt(10);

%-------------------------------------------------------------
% Ergebnis
%.....................................................
f = 0;
for (p = 1 : r)
     f = f + (F(p))^2;
end
%.....................................................
for (i = 1 : n)
     g(i) = 0;
     for (p = 1 : r)
          g(i) = g(i) + 2 * F(p) * dF(p , i);
     end
end
%.....................................................
for (i = 1 : n)
     for (j = 1 : n)
          H(i , j) = 0;
          for (p = 1 : r)
               H(i , j) = H(i , j) + 2 * (dF(p , i) * dF(p , j)
                    + F(p) * ddF(p , i , j));
          end
     end
end


%*************************************************************
```

Literaturverzeichnis

[Burke and Moré, 1994] Burke, J. V. and Moré, J. J. (1994). Exposing constraints. *SIAM Journal on Optimization*, 4(3):573–595.

[Burke et al., 1990] Burke, J. V., Moré, J. J., and Toraldo, G. (1990). Convergence properties of trust region methods for linear and convex constraints. *Mathematical Programming*, 47:305–336.

[Calamai and Moré, 1987] Calamai, P. H. and Moré, J. J. (1987). Projected gradient methods for linearly constrained problems. *Mathematical Programming*, 39:93–116.

[Dunn, 1987] Dunn, J. C. (1987). On the convergence of projected gradient processes to singular critical points. *Journal of Optimization Theory and Applications*, 55:203–216.

[Geiger and Kanzow, 1999] Geiger, C. and Kanzow, C. (1999). *Numerische Verfahren zur Lösung unrestringierter Optimierungsaufgaben*. Springer-Verlag Berlin Heidelberg New York.

[Lin and Moré, 1999] Lin, C.-J. and Moré, J. J. (1999). Newton's method for large bound-constrained optimization problems. *SIAM Journal on Optimization*, 9(4):1100–1127.

[Moré, 1988] Moré, J. J. (1988). Trust regions and projected gradients. *Lecture Notes in Control and Information Sciences*, 113:1–13.

[Toint, 1988] Toint, P. L. (1988). Global convergence of a class of trust region methods for nonconvex minimization in hilbert space. *IMA Journal of Numerical Analysis*, 8:231–252.

[Zarantonello, 1971] Zarantonello, E. H. (1971). Projections on convex sets in hilbert space and spectral theory, part i. In Zarantonello, E. H., editor, *Contributions to Nonlinear Functional Analysis*, number 27 in Contributions to Nonlinear Functional Analysis, pages 237–341, 111 Fifth Avenue, New York, New York 10003. Mathematics Research Center, The University of Wisconsin, Academic Press Inc.

Ich widme diese Arbeit meinen Eltern, die mir dieses Studium und damit diese Arbeit erst ermöglicht haben. Ich möchte mich bei Herrn Professor Dr. Jochen Werner für die interessante Themenstellung sowie die stets freundliche Betreuung während der Erstellung dieser Arbeit bedanken. Mein Dank gilt auch Andrea Haske, die mich währenddessen psychologisch, kulinarisch und geduldig unterstützt und letztendlich Korrektur gelesen hat.