

# **Nichtglatte Optimierung**

Michael Ulbrich

Technische Universität München

Juni 2016



# Inhaltsverzeichnis

<b>1</b>	<b>Einführung</b>	<b>3</b>
1.1	Beispiele für nichtglatte Optimierungsprobleme . . . . .	4
1.1.1	Minimax . . . . .	4
1.1.2	Compressive Sensing / Sparse Optimization . . . . .	5
1.2	Beispiele für nichtglatte Gleichungssysteme . . . . .	6
1.2.1	Das quadratische Penalty-Verfahren . . . . .	6
1.2.2	Reformulierung von Komplementaritätsbedingungen . . . . .	7
<b>2</b>	<b>Nichtglatte Optimierungsprobleme</b>	<b>9</b>
2.1	Einführung . . . . .	9
2.2	Richtungsableitung und eine Optimalitätsbedingung . . . . .	9
2.3	Verfahren des steilsten Abstiegs . . . . .	11
2.4	Konvexe Funktionen und ihr Subdifferential . . . . .	16
2.5	Das Subgradienten-Verfahren . . . . .	24
2.5.1	Subgradienten-Verfahren bei bekanntem Optimalwert . . . . .	25
2.5.2	Subgradienten-Verfahren bei unbekanntem Optimalwert . . . . .	28
2.6	Schnittebenen-Verfahren . . . . .	31
2.7	Das $\varepsilon$ -Subdifferential . . . . .	34
2.8	Bundle Methoden . . . . .	38
2.8.1	Das Bundle-Verfahren aus Sicht der Schnittebenenmethode . . . . .	38
2.8.2	Eine duale Sichtweise des Bundle-Verfahrens . . . . .	41
2.8.3	Globale Konvergenz . . . . .	45
<b>3</b>	<b>Verfahren für Nichtglatte Gleichungssysteme</b>	<b>51</b>
3.1	Ein allgemeines Newton-artiges Verfahren . . . . .	51
3.1.1	Spezialfall: Das gewöhnliche Newton-Verfahren . . . . .	54

3.2	Verallgemeinerte Differentiale . . . . .	54
3.3	Semiglattheit . . . . .	58
3.4	Semiglatte Newton-Verfahren . . . . .	61
<b>4</b>	<b>Konzepte und Methoden für erweitert-reellwertige Funktionen</b>	<b>63</b>
4.1	Konvexe Analysis für erweitert-reellwertige Funktionen . . . . .	63
4.1.1	Wichtige Begriffe . . . . .	63
4.1.2	Die Proximalabbildung . . . . .	64
4.2	Das proximale Gradientenverfahren . . . . .	65
4.2.1	Konvergenz des Verfahrens . . . . .	67
	<b>Literaturverzeichnis</b>	<b>71</b>

# Kapitel 1

## Einführung

Diese Vorlesung gibt eine Einführung in Theorie und Verfahren für nichtlineare Optimierungsprobleme und nichtlineare Gleichungssysteme, die nichtdifferenzierbare Funktionen enthalten. Die beschriebenen Verfahren benutzen Techniken der nichtglatten Analysis, die wir nebenbei entwickeln.

Wir betrachten zwei Klassen nichtglatter Probleme:

- Nichtglatte Optimierungsprobleme:

$$(1.1) \quad \min_{x \in X} f(x),$$

mit lokal Lipschitz-stetiger, aber i.a. nicht überall differenzierbarer Zielfunktion  $f : X \rightarrow \mathbb{R}$  und abgeschlossenem zulässigem Bereich  $X \subset \mathbb{R}^n$ . Man kann auch noch einen Schritt weiter gehen und  $f : \mathbb{R}^n \rightarrow \bar{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$  betrachten. Eine Nebenbedingung  $x \in X$  kann man dann durch  $f(x) = \infty$  für alle  $x \in X$  berücksichtigen.

Um die Darstellung nicht zu technisch werden zu lassen, betrachten wir vornehmlich (aber nicht ausschließlich) konvexe Probleme (d.h.  $f$  und  $X$  konvex) und beschränken uns häufig auf den unrestringierten Fall  $X = \mathbb{R}^n$ .

- Nichtglatte Gleichungssysteme:

$$(1.2) \quad F(x) = 0,$$

mit lokal Lipschitz-stetiger, aber i.a. nicht überall differenzierbarer Funktion  $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ .

Es ist wichtig, zu bemerken, dass die Nichtdifferenzierbarkeit der Funktion  $f$  in (1.1) noch schwerer wiegt als die Nichtdifferenzierbarkeit von  $F$  in (1.2). Denn ist im Fall  $X = \mathbb{R}^n$  die Zielfunktion  $f$  in (1.1) sogar stetig differenzierbar, aber nicht zweimal differenzierbar, so ist  $\nabla f$  nichtglatt und die Berechnung eines stationären Punktes von (1.1) reduziert sich auf das nichtglatte Gleichungssystem

$$\nabla f(x) = 0.$$

Wir haben es somit (falls  $\nabla f$  lokal L-stetig ist) mit einem Problem vom Typ (1.2) zu tun.

Es gibt aber eine Situation, wo man eine enge Brücke zwischen nichtglatten Optimierungsproblemen und nichtglatten Gleichungssystemen schlagen kann: Hat  $f$  die Struktur  $f(x) = g(x) + h(x)$  mit  $g : \mathbb{R}^n \rightarrow \mathbb{R}$  hinreichend glatt und  $h : \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$  konvex, unterhalbstetig (und zur praktischen

Durchführung hinreichend gut strukturiert), dann kann man die (später noch einzuführende) *Proximalabbildung*  $\text{prox}_h : \mathbb{R}^n \rightarrow \mathbb{R}^n$  von  $h$  nutzen, um Optimalitätsbedingungen 1. Ordnung in der Form des Gleichungssystems

$$x - \text{prox}_h(x - \nabla g(x)) = 0$$

zu schreiben. Die Abbildung  $\text{prox}_h$  ist nichtexpansiv, d.h. Lipschitz-stetig mit Konstante 1. Wir erhalten also ein nichtglattes Gleichungssystem, das in wichtigen Spezialfällen semiglatt ist.

Für  $X \subset \mathbb{R}^n$  nichtleer, abgeschlossen und konvex ist die *Indikatorfunktion* von  $X$  definiert durch  $\iota_X(x) = 0$ , falls  $x \in X$ ,  $\iota_X(x) = +\infty$  falls  $x \notin X$ . Es gilt dann  $\text{prox}_{\iota_X}(x) = P_X(x) =$  metrische Projektion auf  $X$ .

Unsere Anforderungen an Verfahren für (1.1) bzw. (1.2) sind daher unterschiedlich:

- a) Für das Problem (1.1) konzentrieren wir uns auf global konvergente Verfahren, d.h. z.B. solche, deren Häufungspunkte lokale Minima von  $f$  sind (oder stationäre Punkte; dieses Konzept müssen wir für nichtglattes  $f$  aber erst noch einführen).
- b) Im Fall der Problemklasse (1.2) zielen wir auf lokal schnell konvergente, Newton-artige Verfahren ab. Diese sind dann insbesondere auch geeignet, um für das Minimierungsproblem (1.1) lokal schnell konvergente Verfahren zu entwickeln, falls  $f \in C^1$ , aber nicht  $C^2$  ist (prominentes Beispiel: Quadratische Penalty-Funktion).

## 1.1 Beispiele für nichtglatte Optimierungsprobleme

### 1.1.1 Minimax

Viele Probleme der Praxis lassen sich in Form einer Minimaxaufgabe schreiben:

$$\min_{x \in X} \max_{y \in Y} \psi(x, y)$$

mit abgeschlossenen Mengen  $X \subset \mathbb{R}^n$ ,  $Y \subset \mathbb{R}^m$  und stetiger Funktion  $\psi : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ .

Insbesondere tritt diese Situation in der *Lagrange-Dualität* auf: Das Problem

$$\min_x f(x) \quad \text{u.d.N.} \quad g(x) \leq 0$$

kann äquivalent in der Form

$$\min_x \max_{\lambda \geq 0} L(x, \lambda)$$

geschrieben werden mit der Lagrangefunktion  $L(x, \lambda) = f(x) + \lambda^T g(x)$ . Das *Lagrange-Dualproblem* lautet

$$\max_{\lambda \geq 0} \min_x L(x, \lambda)$$

und  $d(\lambda) := \min_x L(x, \lambda)$  heißt *duale Zielfunktion*. Diese ist konkav, da  $L$  bzgl.  $\lambda$  konkav ist, im Allgemeinen nichtglatt und  $d$  kann den Wert  $-\infty$  annehmen.

Wir kehren nun zurück zum Minimaxproblem für  $\psi$ . Setzen wir  $f(x) := \max_{y \in Y} \psi(x, y)$ , so ist das Minimaxproblem äquivalent zum (i.a. nichtglatten) Optimierungsproblem

$$\min_{x \in X} f(x).$$

Betrachte z.B.

$$\psi(x, y) = xy, \quad X = \mathbb{R}, \quad Y = [0, 1].$$

Dann gilt

$$f(x) = \max_{y \in [0, 1]} xy = \max\{x, 0\}.$$

Diese Funktion hat einen Knick bei  $x = 0$ .

Man kann zeigen, dass  $f$  konvex ist, falls  $\psi(\cdot, y)$  konvex ist für alle  $y \in Y$ . Weiter gibt es Aussagen über das Subdifferential von  $f$  und darüber, wann  $f$  differenzierbar ist (Satz von Danskin).

### 1.1.2 Compressive Sensing / Sparse Optimization

Aktuell werden in der internationalen Forschung die vielen Möglichkeiten des Compressive Sensing systematisch untersucht. Es geht hierbei darum, approximative Lösungen von stark unterbestimmten (häufig linearen) Gleichungssystemen zu finden, die möglichst viele Nullkomponenten enthalten:

Bei gegebenem Fehlerniveau  $\varepsilon > 0$  sucht man also eine Lösung von

$$(1.3) \quad \min_{x \in \mathbb{R}^n} \|x\|_0 \quad \text{u.d.N.} \quad \|F(x)\|_2^2 \leq \varepsilon^2$$

mit der „Nichtnullen-zählenden Norm“ (aber dies ist natürlich keine Norm)

$$\|x\|_0 := |\{i; x_i \neq 0\}|.$$

Hierbei ist  $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$  stetig differenzierbar mit  $m \ll n$ . Beim Compressive Sensing ist  $F$  linear, d.h.  $F(x) = Ax - b$ .

Die Nebenbedingung ist äquivalent zu  $\|F(x)\|_2 \leq \varepsilon$ , wir haben quadriert, um nicht unnötig zusätzliche Nichtglattheit zu erzeugen.

Das angegebene Problem ist ein kombinatorisches Optimierungsproblem und man kann zeigen, dass es selbst im linearen Fall NP-vollständig ist, außer die Matrix  $A$  hat geeignete Struktur.

Ist eine solche Struktur, die erfreulich häufig vorliegt, gegeben, dann kann man zeigen, dass das Problem (1.3) mit  $F(x) = Ax - b$  äquivalent ist zu dem  $\ell_1$ -Optimierungsproblem

$$\min_{x \in \mathbb{R}^n} \|x\|_1 \quad \text{u.d.N.} \quad \|F(x)\|_2^2 \leq \varepsilon^2$$

Hierbei ist die (konvexe, aber nichtglatte)  $\ell_1$ -Norm gegeben durch:

$$\|x\|_1 = \sum_{i=1}^n |x_i|.$$

Wir erhalten also ein nichtglattes Optimierungsproblem mit Lipschitz-stetiger, konvexer Zielfunktion und stetig differenzierbarer Nebenbedingung. Im Fall, dass  $F$  linear ist, ist die Nebenbedingung zudem konvex und quadratisch.

#### Anwendungen:

- Bildrekonstruktion
- Segmentierung bewegter Bilder

- Verbesserte medizinische Bildgebung mit signifikant reduzierter Anzahl benötigter Aufnahmen (insbes. Reduktion der Strahlenbelastung)
- Neues Paradigma zur komprimierten Signalübertragung von Signalen  $y \in \mathbb{R}^n$ :

Voraussetzung: Es gibt eine (bekannte) Basis  $B$  (oder ein Erzeugendensystem), in der sich alle relevanten Signale  $y$  mit vergleichsweise wenigen Nichtnullen sehr gut approximieren lassen:

$$y \approx Bx, \quad \|x\|_0 \leq \gamma m$$

mit  $m \ll n$  und einem „Sicherheitsfaktor“  $\gamma \in (0, 1)$ .

1. Kompression: Bestimme  $m$  Projektionen  $b_i := (\phi_i, y)_H$ ,  $\|\phi_i\|_H = 1$ , des Signals  $y$  mit geeigneten  $\phi_i$  (häufig können die  $\phi_i$  zufällig gewählt werden). Daraus leitet sich der Begriff Compressive Sensing (komprimierende Abtastung) ab.
2. Übertrage  $b$  statt  $y$ .
3. Rekonstruktion: Bestimme eine dünn besetzte Lösung  $x^*$  von

$$Ax = b, \quad \text{wobei} \quad a_{ij} = (\phi_i, B e_j)_H$$

mit  $e_j = (0, \dots, 0, 1, 0, \dots, 0)^T$  der  $j$ -te Einheitsvektor ist. Das rekonstruierte Signal ist dann  $y^* = Bx^*$ .

Dem Empfänger muss bekannt sein, welche  $\phi_i$  verwendet wurden (z.B. identische Initialisierung eines Zufallsgenerators).

*Ziel* des ersten Teils der Vorlesung ist die Untersuchung nichtglatter Optimierungsprobleme und die Entwicklung geeigneter numerischer Lösungsverfahren.

## 1.2 Beispiele für nichtglatte Gleichungssysteme

Wir geben nun einige Beispiele an, die auf nichtglatte Gleichungssysteme führen.

### 1.2.1 Das quadratische Penalty-Verfahren

Beim quadratische Penalty-Verfahren für das nichtlineare Optimierungsproblem

$$(1.4) \quad \min f(x) \quad \text{u.d.N.} \quad g(x) \leq 0, \quad h(x) = 0$$

mit  $C^2$ -Funktionen  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ ,  $h : \mathbb{R}^n \rightarrow \mathbb{R}^p$  wird das quadratische Penalty-Problem

$$(1.5) \quad \min_{x \in \mathbb{R}^n} P_\alpha(x), \quad P_\alpha(x) = f(x) + \frac{\alpha}{2} \sum_{i=1}^m (g_i(x))_+^2 + \frac{\alpha}{2} \sum_{i=1}^p h_i(x)^2$$

für eine Folge  $(\alpha_k) \rightarrow \infty$  von  $\alpha$ -Werten gelöst. Hierbei ist  $(t)_+ = \max(t, 0)$  und  $(t)_+^2 = ((t)_+)^2$ . Zur Lösung des Problems (1.5) wird ein unrestringiertes Optimierungsverfahren herangezogen, bevorzugt natürlich ein Newton-artiges Verfahren aufgrund seiner Effizienz. Nun gilt aber

$$\nabla P_\alpha(x) = \nabla f(x) + \alpha \sum_{i=1}^m (g_i(x))_+ \nabla g_i(x) + \alpha \sum_{i=1}^p h_i(x) \nabla h_i(x),$$



und diese Funktion ist i.a. nicht stetig differenzierbar, weil  $(g_i(x))_+$  nicht  $C^1$  ist (betrachte z.B.  $g_i(x) = x$ , dann gilt  $(g_i(x))_+ \nabla g_i(x) = (x)_+$  und diese Funktion hat bei  $x = 0$  einen Knick). Das Newton-Verfahren ist also auf (1.5) nicht anwendbar, weil  $P_\alpha(x)$  zwar  $C^1$ , aber nicht  $C^2$  ist. Wir werden aber sehen, dass die nichtglatte Funktion  $\nabla P_\alpha$  allen Voraussetzungen genügt, um die in dieser Vorlesung entwickelten semiglaten Newtonverfahren anwenden zu können.

### 1.2.2 Reformulierung von Komplementaritätsbedingungen

Die Bedeutung nichtglatter Gleichungssysteme beruht wesentlich auf der Tatsache, dass viele Ungleichungssysteme in äquivalente nichtglatte Gleichungssysteme überführt werden können. Ein Beispiel hierfür bildet die folgende wichtige Problemklasse:

#### Nichtlineares Komplementaritätsproblem $\text{NCP}(F)$ :

$$(1.6) \quad x_i \geq 0, \quad F_i(x) \geq 0, \quad x_i F_i(x) = 0, \quad i = 1, \dots, n,$$

wobei  $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$  stetig differenzierbar sei. Die Abkürzung NCP steht hierbei für Nonlinear Complementarity Problem (Nichtlineares Komplementaritätsproblem).

Komplementaritätsprobleme, und allgemeiner Variationsungleichungen, zahlreiche Anwendungen (Hindernisprobleme, Kontaktprobleme, Ökonomie, Optimale Steuerung, Sattelpunktprobleme, Spieltheorie, ...). Insbesondere lassen sich die Kuhn–Tucker-Bedingungen des differenzierbaren Optimierungsproblems

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{u.d.N.} \quad x_i \geq 0, \quad i = 1, \dots, n,$$

in der Form  $\text{NCP}(\nabla f)$  schreiben.

Die *grundlegende Idee* besteht nun darin,  $\text{NCP}(F)$  in folgender Form zu schreiben:

$$(1.7) \quad \Phi(x) \stackrel{\text{def}}{=} \begin{pmatrix} \phi(x_1, F_1(x)) \\ \vdots \\ \phi(x_n, F_n(x)) \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix},$$

wobei  $\phi$  eine *NCP-Funktion* ist, d.h.

$$\phi : \mathbb{R}^2 \rightarrow \mathbb{R}, \quad \phi(a, b) = 0 \iff a \geq 0, \quad b \geq 0, \quad ab = 0.$$

#### Beispiele für NCP-Funktionen:

$$(1.8) \quad \phi_{FB}(a, b) = a + b - \sqrt{a^2 + b^2} \quad \text{Fischer–Burmeister Funktion}$$

$$(1.9) \quad \phi_M(a, b) = \min\{a, b\}.$$

Die Äquivalenz von  $\text{NCP}(F)$  und (1.7) ist offensichtlich. Sowohl  $\phi_{FB}$  als auch  $\phi_M$  sind Lipschitz-stetig, aber nicht überall differenzierbar. Bei (1.7) handelt es sich somit um ein nichtglattes Gleichungssystem der Form (1.2).

Wir betrachten nun das allgemeine nichtlineare Optimierungsproblem (1.4). Die KKT-Bedingungen für (1.4) bilden ein System aus nichtlinearen Gleichungen und Komplementaritätsbedingungen:

$$\begin{aligned} \nabla f(x) + \nabla g(x)\lambda + \nabla h(x)\mu &= 0 \\ h(x) &= 0 \\ g_i(x) \leq 0, \quad \lambda_i \geq 0, \quad \lambda_i g_i(x) &= 0 \quad (i = 1, \dots, m) \end{aligned}$$

Die Komplementaritätsbedingungen können wieder mithilfe einer NCP-Funktion  $\phi$  reformuliert werden:

$$\phi(-g_i(x), \lambda_i) = 0 \quad (i = 1, \dots, m)$$

Damit ist das KKT-System äquivalent zu einem nichtglatten Gleichungssystem

$$F(x, \lambda, \mu) := \begin{pmatrix} \nabla f(x) + \nabla g(x)\lambda + \nabla h(x)\mu \\ h(x) \\ (\phi(-g_i(x), \lambda_i))_{1 \leq i \leq m} \end{pmatrix} = 0.$$

*Ziel* des zweiten Teils der Vorlesung ist die Entwicklung und Analyse einer Klasse q-superlinear bzw. q-quadratisch konvergenter Newton-Verfahren zur Lösung von (1.2) für eine möglichst große Klasse von Funktionen  $F$ , zu der insbesondere  $\nabla P_\alpha$ ,  $\Phi_{FB}$  und  $\Phi_M$  gehören.

## Kapitel 2

# Nichtglatte Optimierungsprobleme

### 2.1 Einführung

Im Mittelpunkt dieses Kapitels steht das folgende Optimierungsproblem:

$$(2.1) \quad \min_{x \in X} f(x),$$

wobei  $f$  lokal Lipschitz-stetig, aber nicht überall differenzierbar ist. Der zulässige Bereich  $X \subset \mathbb{R}^n$  sei abgeschlossen und nichtleer. Wie bereits erwähnt, werden wir hauptsächlich den Fall betrachten, dass  $f$  konvex ist und  $X = \mathbb{R}^n$  gilt.

Wir wiederholen zunächst die Definition der lokalen Lipschitz-Stetigkeit:

**Definition 2.1.1.** Die Funktion  $F : X \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$  heißt *lokal Lipschitz-stetig* bei  $x \in X$  oder *Lipschitz-stetig nahe*  $x \in X$ , falls es Zahlen  $L = L(x) > 0$  und  $\delta = \delta(x) > 0$  gibt mit

$$\|F(y) - F(z)\| \leq L\|y - z\| \quad \forall y, z \in B_\delta(x) \cap X.$$

Hier bezeichnet  $B_\delta(x) = \{y \in \mathbb{R}^n; \|y - x\| < \delta\}$  die offene  $\delta$ -Kugel um  $x$  und  $\|v\| = \sqrt{v^T v}$  ist die Euklidische Norm.

Die Funktion  $F$  heißt *lokal Lipschitz-stetig auf*  $S \subset X$ , falls  $F$  nahe jedem Punkt  $x \in S$  Lipschitz-stetig ist.

**Bemerkung.** Jede auf einer offenen Menge  $U \subset \mathbb{R}^n$  definierte  $C^1$ -Funktion  $F : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$  ist lokal Lipschitz-stetig auf  $U$ . Warum?

### 2.2 Richtungsableitung und eine Optimalitätsbedingung

Wir betrachten das Problem

$$\min_{x \in X} f(x),$$

und nehmen an, dass die Zielfunktion  $f : U \rightarrow \mathbb{R}$  auf einer offenen Umgebung  $U \supset X$  richtungsdifferenzierbar und lokal Lipschitz-stetig ist. Die Menge  $X \subset \mathbb{R}^n$  sei nichtleer. Wie wir sehen werden, erfüllen konvexe Funktionen diese Voraussetzung.

**Definition 2.2.1** (Richtungsableitung). Die stetige Funktion  $F : U \rightarrow \mathbb{R}^m$ ,  $U \subset \mathbb{R}^n$  offen, heißt *richtungsdifferenzierbar* im Punkt  $x \in U$ , falls für alle  $d \in \mathbb{R}^n$  die Richtungsableitung

$$F'(x, d) := \lim_{t \rightarrow 0^+} \frac{F(x + td) - F(x)}{t} \in \mathbb{R}^m$$

existiert. Die Funktion  $F$  heißt *richtungsdifferenzierbar* (auf  $U$ ), falls  $F$  in allen Punkten  $x \in U$  richtungsdifferenzierbar ist.

**Bemerkung.** Nicht jede stetige Funktion (nicht einmal jede Lipschitz-stetige Funktion) ist in jedem Punkt richtungsdifferenzierbar. Ein Beispiel ist  $f : \mathbb{R} \rightarrow \mathbb{R}$ ,  $f(0) = 0$ ,  $f(x) = x \sin(\ln |x|)$ . Wie wir aber sehen werden, sind konvexe Funktionen richtungsdifferenzierbar.

Die Richtungsableitung hat folgende Eigenschaften:

**Lemma 2.2.2.** Die Funktion  $F : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ ,  $U \subset \mathbb{R}^n$  offen, sei richtungsdifferenzierbar im Punkt  $x$ . Dann gilt

$$F'(x, \lambda d) = \lambda F'(x, d) \quad \forall d \in \mathbb{R}^n, \forall \lambda \geq 0 \quad (\text{positive Homogenität})$$

Ist  $F$  zudem Lipschitz-stetig nahe  $x \in U$  mit Konstante  $L$ , so ist die Abbildung  $s \in \mathbb{R}^n \mapsto F'(x, s)$  Lipschitz-stetig mit Konstante  $L$ .

**Beweis:** Natürlich gilt die Aussage für  $\lambda = 0$ , da  $F'(x, 0) = 0$ . Für  $\lambda > 0$  haben wir

$$\begin{aligned} F'(x, \lambda d) &= \lim_{t \rightarrow 0^+} \frac{F(x + t\lambda d) - F(x)}{t} = \lambda \lim_{t \rightarrow 0^+} \frac{F(x + t\lambda d) - F(x)}{t\lambda} \\ &= \lambda \lim_{\tau \rightarrow 0^+} \frac{F(x + \tau d) - F(x)}{\tau} = \lambda F'(x, d). \end{aligned}$$

Sei nun  $F$  Lipschitz-stetig nahe  $x$ . Dann gilt für  $d_1, d_2 \in \mathbb{R}^n$

$$\begin{aligned} \|F'(x, d_1) - F'(x, d_2)\| &= \lim_{t \rightarrow 0^+} \left\| \frac{F(x + td_1) - F(x)}{t} - \frac{F(x + td_2) - F(x)}{t} \right\|_2 \\ &= \lim_{t \rightarrow 0^+} \left\| \frac{F(x + td_1) - F(x + td_2)}{t} \right\|_2 \leq L \|d_1 - d_2\|. \end{aligned}$$

□

Zur Formulierung einer notwendigen Optimalitätsbedingung erster Ordnung benötigen wir den aus der nichtlinearen Optimierung bekannten Tangentialkegel:

**Definition 2.2.3.** Sei  $X \subset \mathbb{R}^n$  eine nichtleere Menge. Als *Tangentialkegel* an  $X$  im Punkt  $x \in X$  bezeichnen wir die Menge

$$T(X, x) = \left\{ d \in \mathbb{R}^n ; \exists \eta_k > 0, x^k \in X : \lim_{k \rightarrow \infty} x^k = x, \lim_{k \rightarrow \infty} \eta_k (x^k - x) = d \right\}.$$

Wir können nun eine notwendige Optimalitätsbedingung formulieren:

**Satz 2.2.4.** Sei  $\bar{x}$  ein lokales Minimum von  $f : U \rightarrow \mathbb{R}$  auf  $X \subset U$ , und  $U \subset \mathbb{R}^n$  offen und  $f$  lokal Lipschitz-stetig und richtungsdifferenzierbar in  $\bar{x}$ . Dann gilt

a)  $\bar{x} \in X$ ,

b)  $f'(\bar{x}, d) \geq 0$  für alle  $d \in T(X, \bar{x})$ .

**Beweis:**

Natürlich gilt a). Zum Nachweis von b) gelte

$$X \ni x^k \rightarrow \bar{x}, \quad \eta_k > 0, \quad d^k := \eta_k(x^k - \bar{x}) \rightarrow d.$$

Im Fall  $d = 0$  ist  $f'(\bar{x}, d) = 0$  und daher nichts zu zeigen. Sei also  $d \neq 0$ . Wegen  $x^k \rightarrow \bar{x}$  und  $d^k \rightarrow d \neq 0$  folgt dann  $t_k := 1/\eta_k \rightarrow 0^+$  für  $k \rightarrow \infty$ . Weiter gibt es  $l > 0$ ,  $L > 0$  und  $\varepsilon > 0$ , so dass alle  $x^k$  und  $\bar{x} + t_k d$ ,  $k \geq l$ , in einer  $\varepsilon$ -Umgebung von  $\bar{x}$  liegen, auf der  $f$  Lipschitz-stetig mit Konstante  $L$  ist. Da  $\bar{x}$  ein lokales Minimum von  $f$  auf  $X$  ist, erhalten wir  $f(x^k) - f(\bar{x}) \geq 0$  für hinreichend große  $k \geq l$  und daher

$$\begin{aligned} 0 &\leq \eta_k(f(x^k) - f(\bar{x})) = \frac{f(\bar{x} + t_k d^k) - f(\bar{x})}{t_k} \\ &= \frac{f(\bar{x} + t_k d^k) - f(\bar{x} + t_k d)}{t_k} + \frac{f(\bar{x} + t_k d) - f(\bar{x})}{t_k} \\ &\leq L\|d^k - d\| + \frac{f(\bar{x} + t_k d) - f(\bar{x})}{t_k} \rightarrow f'(\bar{x}, d). \end{aligned}$$

□

## 2.3 Verfahren des steilsten Abstiegs

Der vielleicht naheliegendste algorithmische Ansatz zur Lösung eines nichtglaten Optimierungsproblems besteht in dem Versuch, das Verfahren des steilsten Abstiegs zu verallgemeinern.

Wir betrachten das Problem

$$\min_{x \in \mathbb{R}^n} f(x),$$

und nehmen an, dass die Zielfunktion  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  richtungsdifferenzierbar und lokal Lipschitz-stetig ist. Wie wir gesehen haben, ist dann  $f'(\bar{x}, d) \geq 0$  für alle  $d \in \mathbb{R}^n$  eine notwendige Bedingung dafür, dass  $\bar{x}$  eine lokale Lösung dieses Problems ist.

Im Fall der reellwertigen Zielfunktion  $f$  gibt die Richtungsableitung  $f'(x, s)$ ,  $\|s\| = 1$ , die Steigung von  $f$  im Punkt  $x$  in Richtung  $s$  an. Daher ist folgende Definition sinnvoll:

**Definition 2.3.1.**

a) Die Richtung  $s \in \mathbb{R}^n \setminus \{0\}$  heißt Abstiegsrichtung für  $f$  in  $x$ , falls gilt:

$$f'(x, s) < 0.$$

b) Die Abstiegsrichtung  $s$  heißt Richtung des steilsten Abstiegs von  $f$  in  $x$ , falls gilt

$$f'\left(x, \frac{s}{\|s\|}\right) = \min_{\|d\|=1} f'(x, d).$$

**Bemerkung.** a) Wegen der Lipschitz-Stetigkeit von  $f'(x, \cdot)$ , siehe Lemma 2.2.2, existieren Richtungen steilsten Abstiegs, falls es überhaupt Abstiegsrichtungen gibt.

- b) Ist  $f$  differenzierbar in  $x$ , dann gilt  $f'(x, s) = \nabla f(x)^T s$  für alle  $s$  und somit sind im Falle  $\nabla f(x) \neq 0$  die Richtungen steilsten Abstiegs gegeben durch  $s = -\lambda \nabla f(x)$ ,  $\lambda > 0$ .

Wir betrachten nun das folgende Verfahren:

**Algorithmus 2.3.2** (Verfahren des steilsten Abstiegs).

Wähle  $x^0 \in \mathbb{R}^n$ .

Für  $k = 0, 1, 2, \dots$ :

1. Bestimme eine Richtung  $s^k$  steilsten Abstiegs von  $f$  in  $x^k$ . Ist dies nicht möglich, dann STOP.
2. Ermittle die optimale Schrittweite  $\sigma_k \geq 0$  entlang  $s^k$ :

$$f(x^k + \sigma_k s^k) = \min_{\sigma \geq 0} f(x^k + \sigma s^k).$$

3. Setze  $x^{k+1} = x^k + \sigma_k s^k$ .

Wir decken nun anhand von Beispielen einige Schwachpunkte dieses Verfahrens auf:

1. Der schwerwiegendste Nachteil ist die Gefahr der Konvergenz gegen einen Punkt, der nicht stationär ist, d.h. in dem Abstiegsrichtungen existieren. Wir konstruieren nun eine konvexe Funktion  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ , für die dies eintritt:

Wir betrachten zunächst die Funktion

$$f_1(x) = x_1^2 + 2x_2^2.$$

Dann gilt

$$\nabla f_1(x) = \begin{pmatrix} 2x_1 \\ 4x_2 \end{pmatrix}.$$

Man kann sich überlegen, dass das Gradientenverfahren mit optimaler Schrittweite ausgehend vom Startpunkt  $x^0 = (2, 1)^T$  die Folge

$$x^{2k} = 9^{-k} \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \quad x^{2k+1} = 9^{-k} \frac{1}{3} \begin{pmatrix} 2 \\ -1 \end{pmatrix}$$

erzeugt. Insbesondere gilt also  $x^k \rightarrow 0$ . Weiter ergibt sich

$$\nabla f_1(x^{2k}) = 9^{-k} \begin{pmatrix} 4 \\ 4 \end{pmatrix} \rightarrow 0, \quad \nabla f_1(x^{2k+1}) = \frac{1}{3} 9^{-k} \begin{pmatrix} 4 \\ -4 \end{pmatrix} \rightarrow 0$$

Als nächstes betrachten wir

$$f_2(x) = \sqrt{f_1(x)}.$$

Für  $x \neq 0$  ist  $f_1(x) \neq 0$  und daher  $f_2$  stetig differenzierbar bei  $x$  mit

$$\nabla f_2(x) = \frac{1}{2\sqrt{f_1(x)}} \nabla f_1(x),$$

d.h. die Richtungen des steilsten Abstiegs von  $f_2$  in  $x$  stimmen mit denen von  $f_1$  überein. Das Verfahren des steilsten Abstiegs, gestartet in  $x^0 = (2, 1)^T$ , erzeugt daher für  $f_2$  dieselbe Punktfolge wie für  $f_1$  und wir haben wiederum

$$x^k \rightarrow 0.$$

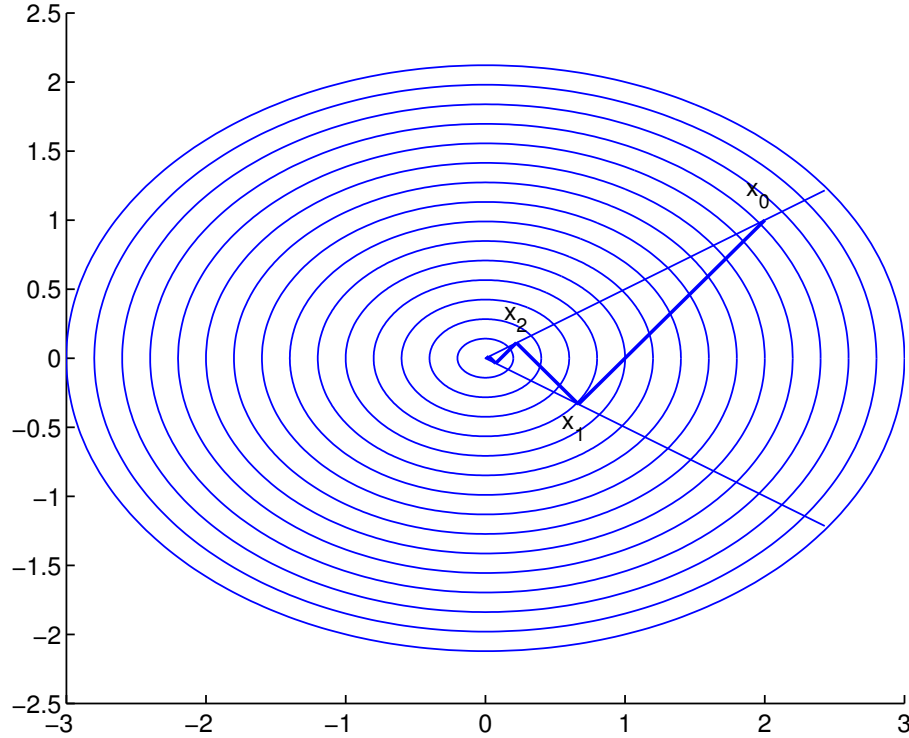


Abbildung 2.1: Verfahren des steilsten Abstiegs für  $f_1(x) = x_1^2 + 2x_2^2$  bzw.  $f_2(x) = \sqrt{f_1(x)}$ .

Die Funktion  $f_2$  ist in 0 nicht differenzierbar. Dies hat zur Folge, dass die Größe des Gradienten keinen Anhaltspunkt über die Nähe zur Lösung 0 liefert:

$$\nabla f_2(x^{2k}) = \frac{\sqrt{6}}{3} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \nabla f_2(x^{2k+1}) = \frac{\sqrt{6}}{3} \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Die Folge  $(x^k)$  verläuft im Sektor  $S = \{(x_1, x_2); |x_2| \leq x_1/2, x_1 > 0\}$ , d.h. wir können die Funktion  $f_2$  außerhalb einer offenen Umgebung des Sektors beliebig abändern, ohne das Verhalten der Iteration zu verändern (sofern dies nicht zu neuen Schrittweiten führt, was aber nicht der Fall sein wird). Wir tun dies durch tangential lineares Fortsetzen der Niveaulinien von  $f_2$  außerhalb des Sektors  $S = \{x; |x_2| \leq 2x_1\}$ . Die resultierende Funktion lautet:

$$f(x) = \begin{cases} f_2(x) = \sqrt{x_1^2 + 2x_2^2} & \text{für } |x_2| \leq 2x_1, \\ \frac{1}{3}(x_1 + 4|x_2|) & \text{für } |x_2| > 2x_1. \end{cases}$$

Die durch das Verfahren des steilsten Abstiegs erzeugte Folge  $(x^k)$  konvergiert dann immer noch gegen 0, aber 0 ist kein stationärer Punkt von  $f$ , denn es gilt:

$$f'(0, (-1, 0)) = -\frac{1}{3}.$$

2. Es ist schwierig, wenn nicht unmöglich, ein Kriterium für endlichen Abbruch des Verfahrens zu finden, das sicherstellt, dass man sich bei Terminierung in der Nähe einer lokalen Lösung befindet (vgl. in diesem Zusammenhang auch  $\nabla f_2(x^k) \not\rightarrow 0$  in 1.). In einer lokalen Lösung  $x$  gilt zwar

$$f'(x, s) = \lim_{t \rightarrow 0+} \frac{f(x + ts) - f(x)}{t} \geq 0 \quad \forall s \in \mathbb{R}^n,$$

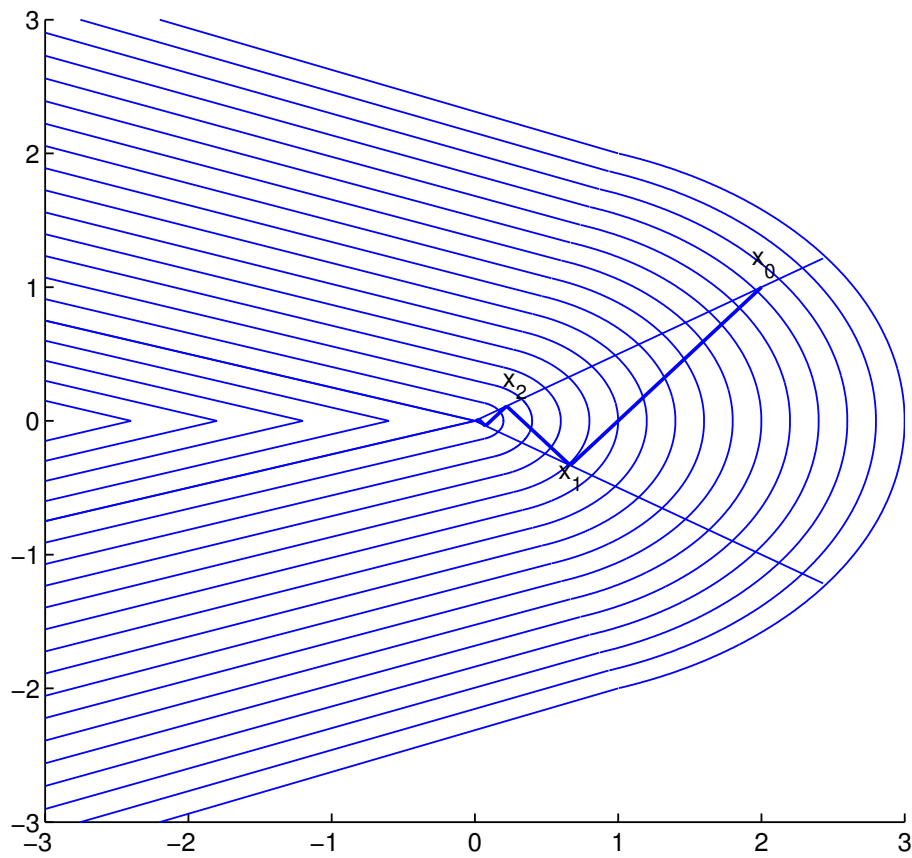


Abbildung 2.2: Verfahren des steilsten Abstiegs für  $f$ : Konvergenz gegen nichtstationären Punkt.



d.h. es existiert keine Richtung steilsten Abstiegs. In Punkten beliebig nahe  $x$  ist dies aber nicht erkennbar. Für das einfache Beispiel  $f(x) = |x|$  z.B. gilt  $f'(0, s) = |s| > 0$  für alle  $s \neq 0$ , d.h. es gibt keine Abstiegsrichtung in  $x = 0$ . Aber:

Für  $x < 0$  ist  $s = 1$  die normierte Richtung steilsten Abstiegs und es gilt  $f'(x, 1) = -1$ , d.h.  $s = 1$  liefert lokal guten Abstieg, auch beliebig nah an der Lösung.

Für  $x > 0$  ist  $s = -1$  die normierte Richtung steilsten Abstiegs und  $f'(x, -1) = -1$ .

Für  $0 < |x| \leq \varepsilon$  ist also nicht zu erkennen, dass  $x$   $\varepsilon$ -optimal ist.

3. Ist  $f$  in  $x$  nicht differenzierbar, so ist die Berechnung einer Richtung steilsten Abstiegs unter Umständen sehr aufwendig. Sei zum Beispiel  $m$  sehr groß und die Vektoren

$$v_k \in \mathbb{R}^n, \quad k = 1, \dots, m,$$

gegeben. Weiter sei

$$f : \mathbb{R}^n \rightarrow \mathbb{R}, \quad f(x) = \max_{1 \leq k \leq m} v_k^T x.$$

Für  $s \in \mathbb{R}^n$  und  $t > 0$  gilt dann

$$f(ts) = t \max_{1 \leq k \leq m} v_k^T s, \text{ d.h. } f'(0, s) = \max_{1 \leq k \leq m} v_k^T s.$$

die Bestimmung der Richtung des steilsten Abstiegs ist daher nicht so einfach. Im Falle ihrer Existenz ist sie gegeben als Lösung von

$$\min_{\|s\| \leq 1} \max_{1 \leq k \leq m} v_k^T s$$

und dieses Problem kann wiederum reformuliert werden in der Form

$$\min \left\{ \alpha ; \alpha \geq v_k^T s, \quad k = 1, \dots, m, \quad \|s\|^2 \leq 1 \right\}.$$

Liefert dieses Problem das Ergebnis  $\alpha = 0$ ,  $s = 0$ , so gibt es keine Abstiegsrichtung. Man kann sich überlegen, dass  $s$  äquivalent gegeben ist durch  $s = -\frac{d}{\|d\|}$ , wobei  $d$  das folgende Problem löst:

$$\min \{ \|d\| ; d \in \text{conv}\{v_1, \dots, v_m\} \}.$$

## 2.4 Konvexe Funktionen und ihr Subdifferential

Wir werden nun die grundlegenden Konzepte der verallgemeinerten Differentialrechnung konvexer Funktionen vorstellen. Zunächst wenden wir uns einigen wichtigen Eigenschaften konvexer Funktionen zu.

**Satz 2.4.1.** *Sei  $f : X \rightarrow \mathbb{R}$  konvex auf der offenen konvexen Menge  $X$ . Dann gilt:*

- a)  $f$  ist lokal Lipschitz-stetig auf  $X$ .
- b)  $f$  ist richtungsdifferenzierbar auf  $X$  und für alle  $x \in X$  gilt

$$f'(x, s) = \inf_{t>0} \Delta(t) \quad \text{mit} \quad \Delta(t) = \frac{f(x + ts) - f(x)}{t}.$$

- c) Für alle  $x, y \in X$  gilt:

$$f(y) - f(x) \geq f'(x, y - x).$$

**Beweis:** zu a):

Sei  $\bar{x} \in X$  beliebig. Da  $X$  offen ist, gibt es  $\delta > 0$ , so dass der achsenparallele Hyperwürfel  $W$  mit Seitenlänge  $6\delta$  und Mittelpunkt  $\bar{x}$  in  $X$  enthalten ist. Seine Ecken seien bezeichnet mit

$$v_k \in \{\bar{x}_1 - 3\delta, \bar{x}_1 + 3\delta\} \times \cdots \times \{\bar{x}_n - 3\delta, \bar{x}_n + 3\delta\}, \quad k = 1, \dots, 2^n.$$

Jeder Punkt  $x \in W$  ist dann Konvexkombination der Ecken, d.h. es gibt  $\lambda_k \geq 0$ ,  $\sum \lambda_k = 1$  mit

$$x = \sum_k \lambda_k v_k.$$

Daher ergibt sich aufgrund der Konvexität von  $f$ :

$$(2.2) \quad f(x) = f\left(\sum_k \lambda_k v_k\right) \leq \sum_k \lambda_k f(v_k) \leq \max_k f(v_k) =: C_W \quad \forall x \in W.$$

Seien nun  $x \in W$  beliebig,  $d = x - \bar{x}$  und  $x' = \bar{x} - d$ . Dann gilt  $[x, x'] \subset W$  und wegen der Konvexität von  $f$  folgt

$$\frac{f(x) + f(x')}{2} \geq f(\bar{x}),$$

also mit (2.2)

$$f(x) \geq 2f(\bar{x}) - f(x') \geq 2f(\bar{x}) - C_W.$$

Damit erhalten wir

$$(2.3) \quad |f(y) - f(x)| \leq C_W - (2f(\bar{x}) - C_W) = 2(C_W - f(\bar{x})) \quad \forall x, y \in W.$$

Wir setzen  $L := (C_W - f(\bar{x}))/\delta$ .

Seien nun  $x, y \in B_\delta(\bar{x})$ ,  $x \neq y$ , beliebig und o.E.  $f(y) \geq f(x)$  (sonst vertausche  $x$  und  $y$ ). Weiter sei  $s = y - x$ ,  $t = \|s\|/(2\delta)$  und  $w = s/t$ . Dann ist  $x + w \in B_{2\delta}(x) \subset B_{3\delta}(\bar{x}) \subset W$ . Wegen

$$\|s\| \leq \|y - \bar{x}\| + \|x - \bar{x}\| < 2\delta$$

gilt außerdem  $0 < t < 1$ . Nun folgt wegen der Konvexität von  $f$  und (2.3):

$$\begin{aligned} |f(y) - f(x)| &= f(y) - f(x) = f(x + tw) - f(x) \leq (1 - t)f(x) + tf(x + w) - f(x) \\ &= t(f(x + w) - f(x)) \leq 2t(C_W - f(\bar{x})) = 2t\delta L = L\|s\| = L\|y - x\|. \end{aligned}$$

Damit ist  $f$  Lipschitz-stetig auf  $B_\delta(\bar{x})$  mit Konstante  $L$ .

zu b):

$\Delta(t)$  ist monoton wachsend, denn für  $0 < \sigma < 1$  und kleine  $t > 0$  gilt  $x, x + ts \in X$  und

$$\Delta(\sigma t) = \frac{f(x + \sigma ts) - f(x)}{\sigma t} \leq \frac{(1 - \sigma)f(x) + \sigma f(x + ts) - f(x)}{\sigma t} = \Delta(t).$$

Wegen der lokalen Lipschitz-Stetigkeit, siehe a), gilt

$$|\Delta(t)| = \frac{|f(x + ts) - f(x)|}{t} \leq \frac{L\|ts\|}{t} = L\|s\|$$

für kleine  $t > 0$ . Daher ergibt sich

$$f'(x, s) = \lim_{t \rightarrow 0^+} \Delta(t) = \inf_{t > 0} \Delta(t) \geq -L\|s\|$$

und das Infimum existiert.

zu c):

Für  $x, y \in X$  und  $s = y - x$  gilt

$$f(x + s) - f(x) = \Delta(1) \geq \inf_{t > 0} \Delta(t) = f'(x, s),$$

wobei Teile des Beweises von b) benutzt wurde. □

**Bemerkung.** Für den Nachweis der Lipschitz-Stetigkeit auf  $X$  benötigen wir, dass  $f$  in einer Umgebung von  $x$  konvex ist. Daher ist die Offenheit von  $X$  hier wichtig. Die Funktion  $f : [0, 1] \rightarrow \mathbb{R}$ ,  $f(0) = 1$ ,  $f(x) = 0$ ,  $x > 0$ , ist beispielsweise konvex, aber nicht stetig.

In der konvexen Analysis übernimmt das Subdifferential jene zentrale Rolle, die im differenzierbaren Fall dem Gradienten zukommt:

**Definition 2.4.2** (Subgradient, Subdifferential). Sei  $f : X \rightarrow \mathbb{R}$  konvex auf der konvexen, offenen Menge  $X \subset \mathbb{R}^n$ . Der Vektor  $g \in \mathbb{R}^n$  heißt *Subgradient* von  $f$  im Punkt  $x \in X$ , wenn gilt:

$$(2.4) \quad f(y) - f(x) \geq g^T(y - x) \quad \forall y \in X.$$

Die Menge  $\partial f(x) \subset \mathbb{R}^n$ ,

$$\partial f(x) = \{g \in \mathbb{R}^n; g \text{ ist Subgradient von } f \text{ in } x\}$$

heißt *Subdifferential* von  $f$  im Punkt  $x \in X$ . Das Subdifferential ist somit eine mengenwertige Abbildung von  $X$  in die Potenzmenge des  $\mathbb{R}^n$ . Wir schreiben hierfür

$$\partial f : X \rightrightarrows \mathbb{R}^n.$$

**Anschauliche Interpretation:**

Der Vektor  $g$  ist Subgradient von  $f$  in  $\bar{x}$ , falls die durch den Punkt  $(\bar{x}, f(\bar{x}))$  verlaufende lineare Funktion  $l$  mit Gradient  $g$ , also

$$l(x) = f(\bar{x}) + g^T(x - \bar{x})$$

auf oder unterhalb des Graphen von  $f$  verläuft. Wir sagen auch: Die Hyperebene  $\text{graph}(l)$  stützt  $\text{graph}(f)$  in  $\bar{x}$  von unten.

**Bemerkung.** Da wir später auch die verallgemeinerte Jacobi-Matrix einführen werden und diese im konvexen Fall mit dem Subdifferential übereinstimmt, müssten wir eigentlich das Subdifferential als Menge von  $1 \times n$ -Matrizen (also Zeilenvektoren) definieren, siehe [Cla83, ]. Da dies aber nicht üblich ist, folgen wir der Konvention und definieren  $\partial f(x)$  als Menge von Spaltenvektoren.

Wir weisen nun einige wichtige Eigenschaften des Subdifferentials nach.

**Satz 2.4.3.** Sei  $f : X \rightarrow \mathbb{R}$  konvex auf der nichtleeren offenen konvexen Menge  $X \subset \mathbb{R}^n$ . Weiter sei  $x \in X$ . Dann gilt:

- a)  $\partial f(x) = \{g \in \mathbb{R}^n; g^T s \leq f'(x, s) \ \forall s \in \mathbb{R}^n\}$ .
- b)  $\partial f(x)$  ist nichtleer, konvex und kompakt. Bezeichnet ferner  $L > 0$  die Lipschitz-Konstante von  $f$  nahe  $x$ , so gilt  $\partial f(x) \subset \bar{B}_L(0)$ .
- c)  $f'(x, s) = \max_{g \in \partial f(x)} g^T s \ \forall s \in \mathbb{R}^n$ .
- d)  $f$  ist nicht nur richtungsdifferenzierbar, sondern sogar Bouligand-differenzierbar:

$$|f(x + s) - f(x) - f'(x, s)| = o(\|s\|) \quad \text{für } \|s\| \rightarrow 0.$$

- e)  $f$  ist genau dann differenzierbar in  $x$  mit  $\nabla f(x) = g$ , wenn  $\partial f(x) = \{g\}$  gilt.

**Beweis:**

zu a):

Sei  $K = \{g \in \mathbb{R}^n; g^T s \leq f'(x, s) \ \forall s \in \mathbb{R}^n\}$ .

•  $\partial f(x) \subset K$ :

Sei  $g \in \partial f(x)$ . Für  $s \in \mathbb{R}^n$  und kleine  $t > 0$  gilt dann  $x + ts \in X$  und wegen (2.4)

$$f(x + ts) \geq f(x) + g^T(ts),$$

folglich

$$g^T s \leq \frac{f(x + ts) - f(x)}{t} \xrightarrow{t \rightarrow 0^+} f'(x, s).$$

Daraus ergibt Grenzübergang  $t \rightarrow 0^+$ :

$$g^T s \leq f'(x, s).$$

•  $K \subset \partial f(x)$ :

Sei  $g \in K$ . Nach Satz 2.4.1 c) haben wir für alle  $y \in X$ :

$$f(y) \geq f(x) + f'(x, y - x) \geq f(x) + g^T(y - x),$$

also  $g \in \partial f(x)$ .

zu b):

•  $\partial f(x)$  abgeschlossen und konvex:

Für alle  $y \in X$  ist die Menge  $M_y = \{g; f(y) \geq f(x) + g^T(y - x)\}$  offensichtlich abgeschlossen und konvex und dies vererbt sich auf

$$\partial f(x) = \bigcap_{y \in X} M_y.$$

•  $\partial f(x) \subset \bar{B}_L(0)$ :

Sei  $s \in \mathbb{R}^n$  beliebig. Ist nun  $L$  die Lipschitz-Konstante von  $f$  nahe  $x$ , so ist  $f'(x, \cdot)$  nach Lemma 2.2.2 Lipschitz-stetig mit Konstante  $L$  und daher

$$|f'(x, s)| = |f'(x, s) - f'(x, 0)| \leq L\|s\|.$$

Sei nun  $g \in \partial f(x)$  beliebig. Für  $s = g$  erhalten wir nun aus a)

$$\|g\|^2 = g^T s \leq f'(x, s) \leq L\|s\| = L\|g\|.$$

also  $g \in \bar{B}_L(0)$ . Insbesondere ist damit auch die Kompaktheit von  $\partial f(x)$  nachgewiesen.

•  $\partial f(x) \neq \emptyset$  und c):

Sei  $s \in \mathbb{R}^n$  beliebig fest und

$$M_1 = \left\{ \begin{pmatrix} y \\ z \end{pmatrix}; y \in X, z \in \mathbb{R}, f(y) < z \right\},$$

$$M_2 = \left\{ \begin{pmatrix} y \\ z \end{pmatrix}; y \in X, y = x + ts, z = f(x) + tf'(x, s), t \geq 0 \right\}$$

Beide Mengen sind nichtleer und konvex. Sie sind auch disjunkt, da für  $(y^T, z)^T \in M_2$  gilt:

$$z = f(x) + tf'(x, s) = f(x) + f'(x, ts) \leq f(x + ts) = f(y).$$

Daher lassen sich  $M_1$  und  $M_2$  durch eine Hyperebene trennen, d.h. es gibt  $(w^T, \alpha)^T \in \mathbb{R}^{n+1} \setminus \{0\}$  mit

$$w^T y_2 + \alpha z_2 \leq w^T y_1 + \alpha z_1 \quad \forall \begin{pmatrix} y_1 \\ z_1 \end{pmatrix} \in M_1, \begin{pmatrix} y_2 \\ z_2 \end{pmatrix} \in M_2.$$

Aus Stetigkeitsgründen gilt dann auch

$$(2.5) \quad w^T y_2 + \alpha z_2 \leq w^T y_1 + \alpha z_1 \quad \forall \begin{pmatrix} y_1 \\ z_1 \end{pmatrix} \in \text{epi}(f), \begin{pmatrix} y_2 \\ z_2 \end{pmatrix} \in M_2,$$

wobei

$$\text{epi}(f) = \left\{ \begin{pmatrix} y \\ z \end{pmatrix}; y \in X, z \in \mathbb{R}, f(y) \leq z \right\}$$

der Epigraph von  $f$  ist. Wegen  $(x^T, f(x) + 1)^T \in \text{epi}(f)$  und  $(x^T, f(x))^T \in M_2$  folgt  $\alpha \geq 0$ .

Angenommen,  $\alpha = 0$ . Dann gilt  $w \neq 0$  und da  $X$  offen ist, gibt es  $\tau > 0$  mit  $y = x - \tau w \in X$ . Nun ist  $(y^T, f(y))^T \in \text{epi}(f)$  sowie  $(x^T, f(x))^T \in M_2$  und daher (wir benutzen  $\alpha = 0$ )

$$w^T x \leq w^T y = w^T x - \tau \|w\|^2 < w^T x.$$

Dies ist ein Widerspruch. Somit ist  $\alpha > 0$  nachgewiesen.

Durchmultiplizieren von (2.5) mit  $1/\alpha$  ergibt für  $g = -w/\alpha$ :

$$-g^T y_2 + z_2 \leq -g^T y_1 + z_1 \quad \forall \begin{pmatrix} y_1 \\ z_1 \end{pmatrix} \in \text{epi}(f), \begin{pmatrix} y_2 \\ z_2 \end{pmatrix} \in M_2,$$

Für  $y \in X$  ergibt nun die spezielle Wahl  $(y_1, z_1) = (y, f(y))$ ,  $(y_2, z_2) = (x, f(x))$ :

$$-g^T x + f(x) \leq -g^T y + f(y) \quad \forall y \in X,$$

und somit ist  $g$  ein Subgradient.

Zum Nachweis von c) nutzen wir, dass für kleine  $t > 0$  gilt ( $X$  ist offen):  $x + ts \in X$ , also  $((x + ts)^T, f(x) + tf'(x, s))^T \in M_2$  und somit

$$-g^T(x + ts) + f(x) + tf'(x, s) \leq -g^T x + f(x).$$

Dies zeigt  $f'(x, s) \leq g^T s$  und zusammen mit a) folgt c).

zu d):

Angenommen, die Aussage gilt nicht. Dann gibt es  $\gamma > 0$ ,  $t_k \rightarrow 0^+$  und  $v_k$ ,  $\|v_k\| = 1$  mit

$$|f(x + t_k v_k) - f(x) - f'(x, t_k v_k)| \geq \gamma t_k \quad \forall k.$$

Durch Übergang auf eine Teilfolge erzielen wir  $v_k \rightarrow v$ . Sei nun  $L$  die Lipschitz-Konstante von  $f$  nahe  $x$ . Wir benutzen die lokale Lipschitz-Stetigkeit von  $f$ , die positive Homogenität und die Lipschitz-Stetigkeit von  $f'(x, \cdot)$ , um zu zeigen, dass für große  $k$  gilt:

$$\begin{aligned} \gamma &\leq \frac{|f(x + t_k v_k) - f(x) - f'(x, t_k v_k)|}{t_k} \\ &\leq \frac{|f(x + t_k v_k) - f(x + t_k v)|}{t_k} + \frac{|f(x + t_k v) - f(x) - f'(x, t_k v_k)|}{t_k} \\ &\leq L\|v_k - v\| + \left| \frac{f(x + t_k v) - f(x)}{t_k} - f'(x, v_k) \right| \\ &\xrightarrow{k \rightarrow \infty} L\|v - v\| + |f'(x, v) - f'(x, v)| = 0. \end{aligned}$$

Dies ist ein Widerspruch und somit ist d) nachgewiesen.

zu e):

Sei  $f$  differenzierbar in  $x$  mit  $g = \nabla f(x)$ . Dann gilt für alle  $s \in \mathbb{R}^n$ :

$$f'(x, s) = \lim_{t \rightarrow 0^+} \frac{f(x + ts) - f(x)}{t} = \nabla f(x)^T s = g^T s.$$

Ist nun  $v \in \partial f(x)$ , so folgt aus a) mit  $s = v - g$ :

$$\|s\|^2 = v^T s - g^T s = v^T s - f'(x, s) \stackrel{\text{a)}}{\leq} 0.$$

Dies zeigt  $v = g$  und somit  $\partial f(x) = \{g\}$ .

Gelte nun umgekehrt  $\partial f(x) = \{g\}$ . Dann gilt wegen c):

$$f'(x, s) = g^T s \quad \forall s \in \mathbb{R}^n$$

und mit d) folgt daraus:

$$|f(x + s) - f(x) - g^T s| = o(\|s\|) \quad \text{für } \|s\| \rightarrow 0.$$

Daher ist  $f$  in  $x$  differenzierbar mit  $\nabla f(x) = g$ . □

Weiter gilt die folgende notwendige und hinreichende Optimalitätsbedingung:

**Satz 2.4.4.** *Sei  $X \subset \mathbb{R}^n$  offen und konvex und  $f : X \rightarrow \mathbb{R}$  sei eine konvexe Funktion. Weiter sei  $x^* \in \mathbb{R}^n$ . Dann sind die folgenden Aussagen äquivalent:*

- a) *Der Punkt  $x^*$  ist (globales = lokales) Minimum von  $f$  auf  $X$ .*
- b) *Es gilt  $f'(x^*, s) \geq 0 \quad \forall s \in \mathbb{R}^n$ .*
- c) *Es gilt  $0 \in \partial f(x^*)$ .*

**Beweis:**

a)  $\implies$  b):

Sei  $x^*$  ein lokales Minimum von  $f$  auf  $X$ . Ist nun  $s \in \mathbb{R}^n$  beliebig, so gilt  $f(x^* + ts) \geq f(x^*)$  für kleine  $t > 0$  und somit

$$f'(x^*, s) = \lim_{t \rightarrow 0^+} \frac{f(x^* + ts) - f(x^*)}{t} \geq 0.$$

b)  $\implies$  c):

Wegen

$$0^T s = 0 \leq f'(x^*, s) \quad \forall s \in \mathbb{R}^n$$

folgt  $0 \in \partial f(x^*)$  gemäß Satz 2.4.3 a).

c)  $\implies$  a):

Da 0 ein Subgradient ist, folgt nach Definition

$$f(x) - f(x^*) \geq 0^T(x - x^*) = 0 \quad \forall x \in X.$$

Somit ist  $x^*$  globales Minimum von  $f$  auf  $X$ . □

Wir können Satz 2.4.3 c) und Satz 2.4.4 zu einer Charakterisierung der Richtungen des steilsten Abstiegs nutzen:

**Satz 2.4.5** (Richtung des steilsten Abstiegs). *Sei  $f : X \rightarrow \mathbb{R}$  konvex auf der offenen und konvexen Menge  $X \subset \mathbb{R}^n$ . Im Punkt  $x \in X$  gelte  $0 \notin \partial f(x)$ . Weiter sei  $s = -P_{\partial f(x)}(0)$ , d.h.  $d = -s$  sei Lösung des Problems*

$$\min_{d \in \mathbb{R}^n} \|d\| \quad \text{u.d.N.} \quad d \in \partial f(x).$$

*Dann ist  $s$  eine Richtung des steilsten Abstiegs von  $f$  in  $x$ .*

**Beweis:** Sei  $d$  die Lösung von

$$\min_{d \in \partial f(x)} \|d\|.$$

Wegen der Voraussetzung  $0 \notin \partial f(x)$  ist  $d \neq 0$ . Sei nun  $s = -d$ .

Da  $d$  die konvexe Funktion  $h(v) = \frac{1}{2}\|v\|^2$  auf der konvexen Menge  $\partial f(x)$  minimiert, gilt:

$$\forall g \in \partial f(x) : 0 \leq \nabla h(d)^T (g - d) = d^T g - d^T d.$$

Daraus folgt  $\min_{g \in \partial f(x)} g^T d = d^T d = \|d\|^2$  und somit

$$f' \left( x, \frac{s}{\|s\|} \right) = \max_{g \in \partial f(x)} g^T \frac{s}{\|s\|} = \max_{g \in \partial f(x)} \frac{-g^T d}{\|d\|} = - \min_{g \in \partial f(x)} \frac{g^T d}{\|d\|} = - \frac{\|d\|^2}{\|d\|} = -\|d\|.$$

Andererseits haben wir

$$\min_{\|v\|=1} f'(x, v) = \min_{\|v\|=1} \max_{g \in \partial f(x)} g^T v \geq \min_{\|v\|=1} d^T v = -\|d\| = f' \left( x, \frac{s}{\|s\|} \right),$$

d.h.  $s = -d$  ist eine Richtung steilsten Abstiegs. □

Dieser Satz führt in Verbindung mit dem folgenden Satz 2.4.6 auf die Formel am Ende von Abschnitt 2.3. Weiter sehen wir, dass zur Berechnung der Richtung des steilsten Abstiegs die Kenntnis des gesamten Subdifferentials erforderlich ist.

Eine wichtige Klasse nichtglatter konvexer Funktionen entsteht durch Maximumsbildung:

**Satz 2.4.6.** Seien die Funktionen  $f_i : X \rightarrow \mathbb{R}$  konvex auf der offenen konvexen Menge  $X \subset \mathbb{R}^n$ . Dann ist die Funktion

$$f : X \rightarrow \mathbb{R}, \quad f(x) = \max_{1 \leq i \leq m} f_i(x)$$

konvex. Im Punkt  $x \in X$  gilt weiter

$$(2.6) \quad f'(x, s) = \max_{i \in I(x)} f'_i(x, s),$$

$$(2.7) \quad \partial f(x) = \text{conv} \left( \bigcup_{i \in I(x)} \partial f_i(x) \right).$$

mit  $I(x) = \{i; f_i(x) = f(x)\}$ .

**Beweis:**  $f$  ist als punktweises Maximum konvexer Funktionen konvex.

Da die  $f_i$  stetig sind, gibt es  $\varepsilon > 0$  mit  $I(y) \subset I(x)$  für alle  $y \in B_\varepsilon(x)$ . Dies zeigt für  $s \in \mathbb{R}^n$  und kleine  $t > 0$

$$\begin{aligned} \frac{f(x+ts) - f(x)}{t} &= \frac{\max_{i \in I(x)} f_i(x+ts) - f(x)}{t} = \frac{\max_{i \in I(x)} (f_i(x+ts) - f_i(x))}{t} \\ &= \max_{i \in I(x)} \frac{f_i(x+ts) - f_i(x)}{t} \xrightarrow{t \rightarrow 0^+} \max_{i \in I(x)} f'_i(x, s). \end{aligned}$$

Sei die rechts stehende Menge mit  $K$  bezeichnet.

•  $K \subset \partial f(x)$ :

Für  $i \in I(x)$  und  $g \in \partial f_i(x)$  gilt wegen Satz 2.4.3 a)

$$g^T s \leq f'_i(x, s) \leq f'(x, s) \quad \forall s \in \mathbb{R}^n$$



und daher  $g \in \partial f(x)$ . Dies zeigt  $\bigcup_{i \in I(x)} \partial f_i(x) \subset \partial f(x)$  und daher

$$K = \text{conv} \left( \bigcup_{i \in I(x)} \partial f_i(x) \right) \subset \text{conv} \partial f(x) = \partial f(x).$$

•  $\partial f(x) \subset K$ :

Die Menge  $K$  ist die konvexe Hülle eines Kompaktums und daher kompakt (als Übung).

Angenommen nun, es gibt  $g \in \partial f(x)$  mit  $g \notin K$ . Wir wenden den strikten Trennungssatz auf die Mengen  $\{g\}$  und  $K$  an. Es gibt also  $v \in \mathbb{R}^n$  und  $\alpha \in \mathbb{R}$  mit

$$v^T y < \alpha < v^T g \quad \forall y \in K.$$

Insbesondere gilt für alle  $i \in I(x)$ :

$$f'_i(x, v) = \max_{y \in \partial f_i(x)} y^T v \leq \max_{y \in K} y^T v \leq \alpha < v^T g.$$

Daraus wiederum folgt

$$f'(x, v) = \max_{i \in I(x)} f'_i(x, v) \leq \alpha < g^T v,$$

aber dies ist ein Widerspruch zu  $g \in \partial f(x)$ . □

Als nächstes geben wir ohne Beweis ein paar Rechenregeln an:

**Satz 2.4.7.** a) Sei  $f : U \rightarrow \mathbb{R}$  konvex,  $U \subset \mathbb{R}^m$  offen und konvex,  $A \in \mathbb{R}^{m \times n}$  und  $b \in \mathbb{R}^m$ . Weiter sei  $X \subset \mathbb{R}^n$  offen und konvex mit  $AX + b \subset U$ . Dann ist die Funktion  $g : X \ni x \rightarrow f(Ax + b)$  konvex und es gilt:

$$\partial g(x) = A^T \partial f(Ax + b).$$

b) Sei  $X \subset \mathbb{R}^n$  offen, nichtleer und konvex und die Funktionen  $f_1, \dots, f_m : X \rightarrow \mathbb{R}$  seien konvex. Weiter sei  $g : \mathbb{R}^m \rightarrow \mathbb{R}$  konvex und es gelte

$$\forall y, z \in \mathbb{R}^m \text{ mit } y \leq z : \quad g(y) \leq g(z).$$

Dann ist  $h : x \mapsto g(f_1(x), \dots, f_m(x))$  konvex und es gilt:

$$\partial h(x) = \left\{ \sum_{i=1}^m v_i w_i ; v \in \partial g(f_1(x), \dots, f_m(x)), w_i \in \partial f_i(x) \right\}.$$

Insbesondere folgt daraus:

$$\partial(f_1 + \dots + f_m)(x) = \partial f_1(x) + \dots + \partial f_m(x).$$

**Beweis:** Siehe [HU96, Thm. VI.4.2.1 und VI.4.3.1].

Aus Satz 2.4.4 erhalten wir folgenden Mittelwertsatz:

**Satz 2.4.8.** Sei  $f : U \rightarrow \mathbb{R}$  konvex auf der offenen konvexen Menge  $U \subset \mathbb{R}^n$  und  $x, y \in U$ . Dann gibt es  $z \in (x, y)$  mit

$$f(y) - f(x) \in \partial f(z)^T (y - x).$$

**Beweis:** Sei  $s = y - x$ . Wir definieren die Funktion

$$\phi(t) = f(x + ts) - tf(y) - (1 - t)f(x), \quad t \in (-\delta, 1 + \delta),$$

$\delta > 0$  hinreichend klein. Sie ist konvex und  $\leq 0$ . Weiter gilt nach den oben angegebenen Rechenregeln:

$$\partial\phi(t) = \partial f(x + ts)^T s - f(y) + f(x).$$

Wegen  $\phi(t) \leq 0 = \phi(0) = \phi(1)$  nimmt  $\phi$  ein lokales Minimum in einem  $t^* \in (0, 1)$  an. Dort gilt  $0 \in \partial\phi(t^*)$  nach Satz 2.4.4 und wir können daher  $z = x + t^*s$  wählen.  $\square$

## 2.5 Das Subgradienten-Verfahren

Wie wir gesehen haben, ist das Verfahren des steilsten Abstiegs einerseits nicht immer konvergent und andererseits erfordert die Bestimmung der Richtung des steilsten Abstiegs die Kenntnis des gesamten Subdifferentials. Wir betrachten nun eine Methode, die in jedem Schritt nur den Funktionswert und einen Subgradienten benötigt.

Zunächst stellen wir fest, dass nicht jeder Subgradient  $g$  eine Abstiegsrichtung  $s = -g$  liefert:

**Beispiel 2.5.1.** Sei  $f(x) = \frac{x_1^2}{2} + 2|x_2|$ . Dann gilt nach der Summenformel

$$\partial f(x_1, 0) = \{(x_1, 0)^T\} + \{(0, \mu)^T; |\mu| \leq 2\} = \{(x_1, \mu)^T; |\mu| \leq 2\}.$$

Betrachte nun im Punkt  $x = (1, 0)^T$  den Subgradienten  $g = (1, 2)^T$ . Dann gilt für  $t > 0$ :

$$f(x - tg) = f(1 - t, -2t) = \frac{(1 - t)^2}{2} + 4t = \frac{1 + t^2}{2} + 3t = f(x) + \frac{t^2}{2} + 3t > f(x).$$

Somit ist  $s = -g$  keine Abstiegsrichtung.

Aber Subgradienten haben eine andere wertvolle Eigenschaft:

Die Richtung  $s = -g$  ist eine Abstiegsrichtung für  $\|x - x^*\|$ , falls  $f(x) > f(x^*)$  gilt.

Da dies ohne nennenswerten Mehraufwand möglich ist, betrachten wir im folgenden das restringierte Problem

$$(2.8) \quad \min_x f(x) \quad \text{u.d.N.} \quad x \in X.$$

Hierbei sei  $X \subset \mathbb{R}^n$  konvex, abgeschlossen und nichtleer (z.B.  $X = \mathbb{R}^n$ ) und  $f : U \rightarrow \mathbb{R}$  sei konvex auf der offenen konvexen Menge  $U \supset X$ .

Mit  $P_X(x) = \operatorname{argmin}_{y \in X} \|y - x\|$  bezeichnen wir die Projektion auf  $X$ .

**Satz 2.5.2.** Sei  $x^*$  eine Lösung von (2.8). Weiter sei  $x \in X$  und  $g \in \partial f(x)$ . Dann gilt:

$$(2.9) \quad \|P_X(x - \sigma g) - x^*\| \leq \|x - \sigma g - x^*\| < \|x - x^*\| \quad \forall 0 < \sigma < \frac{2(f(x) - f(x^*))}{\|g\|^2}.$$

Genauer haben wir:

$$(2.10) \quad \|P_X(x - \sigma g) - x^*\|^2 \leq \|x - \sigma g - x^*\|^2 \leq \|x - x^*\|^2 - \rho(\sigma) \\ \text{mit} \quad \rho(\sigma) = 2\sigma(f(x) - f(x^*)) - \sigma^2\|g\|^2.$$

**Beweis:** Wir verwenden die Ungleichung für Subgradienten (Satz 2.4.3 a)):

$$\begin{aligned}\|x - \sigma g - x^*\|^2 &= \|x - x^*\|^2 - 2\sigma g^T(x - x^*) + \sigma^2 \|g\|^2 \\ &\leq \|x - x^*\|^2 + 2\sigma(f(x^*) - f(x)) + \sigma^2 \|g\|^2 = \|x - x^*\|^2 - \rho(\sigma).\end{aligned}$$

Der Term  $\rho(\sigma)$  ist positiv genau für die in (2.9) angegebenen Werte von  $\sigma$ .

Die Aussage für den projizierten Pfad folgt daraus, dass  $x, x^* \in X$  gilt und  $P_X$  nicht-expansiv ist:

$$\|P_X(x - \sigma g) - x^*\| = \|P_X(x - \sigma g) - P_X(x^*)\| \leq \|x - \sigma g - x^*\|.$$

□

Wir betrachten nun folgende allgemeine Klasse von Verfahren zur Lösung des Problems (2.8):

**Algorithmus 2.5.3** (Projiziertes Subgradienten-Verfahren).

Wähle  $x^0 \in X$ .

Für  $k = 0, 1, 2, \dots$ :

1. Bestimme einen Subgradienten  $g^k \in \partial f(x^k)$ .
2. Falls ein geeignetes Abbruchkriterium erfüllt ist, STOP mit Ergebnis  $x^k$ .
3. Ermittle eine geeignete Schrittweite  $\sigma_k > 0$  und setze

$$x^{k+1} = P_X(x^k - \sigma_k g^k).$$

Aus Beispiel 2.5.1 wissen wir, dass es keinen Sinn macht, zu versuchen, stets Abstieg, d.h.  $f(x^{k+1}) < f(x^k)$ , zu erzielen. Vielmehr zeigt Satz 2.5.2, dass wir stattdessen versuchen sollten,  $\sigma_k > 0$  so zu wählen, dass gilt:

$$\|x^{k+1} - x^*\| < \|x^k - x^*\| \quad \forall k.$$

### 2.5.1 Subgradienten-Verfahren bei bekanntem Optimalwert

Wir betrachten nun zunächst den Fall, dass der optimale Zielfunktionswert  $f(x^*)$  bekannt ist.

**Beispiel 2.5.4.** Gesucht sei ein zulässiger Punkt bezüglich der Menge

$$M = \{x; g_i(x) \leq 0, i = 1, \dots, m\}, \quad g_i : \mathbb{R}^n \rightarrow \mathbb{R} \text{ konvex.}$$

Hierzu können wir die Funktion

$$f(x) = \max\{0, g_1(x), \dots, g_m(x)\}$$

minimieren. Wissen wir a priori (z.B. aufgrund der Problemstellung), dass  $M$  nichtleer ist, dann ist der Optimalwert gleich 0.

**Beispiel 2.5.5.** Angenommen, unser Auftraggeber ist bereits zufrieden, wenn wir den ihm bisher bekannten besten Funktionswert  $f(\bar{x}) > 0$ ,  $\bar{x} \in X$ , um 10% senken können. Dann können wir statt (2.8) das konvexe Problem

$$\min_x \max(0.9f(\bar{x}), f(x)) \quad \text{u.d.N.} \quad x \in X.$$

lösen. Gehen wir davon aus, dass diese Verbesserung erzielbar ist, so ist  $0.9f(\bar{x})$  der optimale Funktionswert dieses Problems.

Bei bekanntem Optimalwert  $f(x^*)$  ist es natürlich sinnvoll,  $\sigma_k$  so zu wählen, dass

$$\rho_k(\sigma_k) = 2\sigma_k(f(x^k) - f(x^*)) - \sigma_k^2 \|g^k\|^2$$

maximal ist. Wegen

$$\rho'_k(\sigma_k) = 2(f(x^k) - f(x^*)) - 2\sigma_k \|g^k\|^2 = 0 \iff \sigma_k = \sigma_k^* := \frac{f(x^k) - f(x^*)}{\|g^k\|^2}$$

betrachten wir Algorithmus 2.5.3 daher im folgenden für die Schrittweitenwahl:

$$(2.11) \quad \sigma_k = \sigma_k^* = \frac{f(x^k) - f(x^*)}{\|g^k\|^2}.$$

Es gilt

$$(2.12) \quad \rho_k(\sigma_k^*) = \frac{(f(x^k) - f(x^*))^2}{\|g^k\|^2}.$$

Wir erhalten das folgende Konvergenzresultat:

**Satz 2.5.6.** *Die Funktion  $f : U \rightarrow \mathbb{R}$  sei konvex auf der offenen konvexen Menge  $U \subset \mathbb{R}^n$ . Die Menge  $X \subset U$  sei abgeschlossen und konvex und  $x^* \in X$  sei eine Lösung von (2.8). Die Folge  $(x^k) \subset X$  werde erzeugt durch Algorithmus 2.5.3, wobei die Schrittweiten  $(\sigma_k)$  gemäß (2.11) gewählt werden, und wir abbrechen, sobald  $f(x^k) = f(x^*)$  gilt. Der Algorithmus erzeuge unendlich viele Iterierte. Dann gilt:*

- a) *Die Folge  $(\|g^k\|)$  ist beschränkt.*
- b)  $\lim_{k \rightarrow \infty} f(x^k) = f(x^*)$ .
- c) *Die Folge  $(x^k)$  konvergiert gegen eine Lösung von (2.8).*

**Beweis:** Sei  $R = \|x^0 - x^*\|$ . Induktiv ergibt sich aus (2.9), (2.10) und (2.12) für alle  $k \geq 0$ :

$$(2.13) \quad \begin{aligned} \|x^{k+1} - x^*\|^2 &\leq \|x^k - x^*\|^2 - \rho_k(\sigma_k) \leq \dots \leq \|x^0 - x^*\|^2 - \sum_{l=0}^k \rho_l(\sigma_l) \\ &= R^2 - \sum_{l=0}^k \rho_l(\sigma_l) < R^2. \end{aligned}$$

Somit gilt  $x^k \in K := \bar{B}_R(x^*) \cap X$ .

zu a): Wir benötigen nun folgendes Hilfsresultat:

**Lemma 2.5.7.** *Sei  $F : M \rightarrow \mathbb{R}^m$  lokal Lipschitz-stetig auf der Menge  $M \subset \mathbb{R}^n$ . Dann ist  $F$  Lipschitz-stetig auf jedem Kompaktum  $K \subset M$ .*

Die Funktion  $f$  ist also auf  $K$  Lipschitz-stetig (Konstante  $L$ ) und nach Satz 2.4.3 b) gilt

$$\|g^k\| \leq L \quad \forall k.$$

Somit existiert  $C = \sup_{k \geq 0} \|g^k\| \leq L$ .

zu b): Wir erhalten mit (2.13)

$$0 \leq \|x^{k+1} - x^*\|^2 \leq R^2 - \sum_{l=0}^k \rho_l(\sigma_l),$$

also folgt wegen  $\rho_k(\sigma_k) > 0$ , dass  $(\rho_k(\sigma_k))$  eine Nullfolge ist. Nun ergibt sich aus (2.11) und (2.12):

$$0 < f(x^k) - f(x^*) = \sqrt{\rho_k(\sigma_k)} \|g^k\| \leq C \sqrt{\rho_k(\sigma_k)} \rightarrow 0.$$

zu c):

Da die Folge  $(x^k)$  im Kompaktum  $K$  liegt, gibt es eine konvergente Teilfolge  $x^{k'} \rightarrow \bar{x} \in K$ . Nun gilt, da  $f$  stetig ist:

$$f(\bar{x}) = \lim_{k' \rightarrow \infty} f(x^{k'}) = f(x^*).$$

Daher ist  $\bar{x}$ , und somit jeder Häufungspunkt von  $(x^k)$ , optimale Lösung.

Wir können also für  $x^*$  speziell einen Häufungspunkt von  $(x^k)$  wählen. Sei  $\bar{x}$  ein weiterer Häufungspunkt.

Nach Wahl von  $\sigma_k$  ist die Folge  $(\|x^k - x^*\|)$  monoton fallend und konvergiert daher gegen eine Zahl  $\alpha \geq 0$ . Da  $x^*$  und  $\bar{x}$  Häufungspunkte von  $(x^k)$  sind, folgt

$$\|x^* - x^*\| = \alpha = \|\bar{x} - x^*\|.$$

Dies zeigt  $\alpha = 0$  und  $\bar{x} = x^*$ . Somit konvergiert  $(x^k)$  gegen eine (globale) Lösung von (2.8).  $\square$

**Beweis von Lemma 2.5.7:**  $F$  ist stetig auf dem Kompaktum  $K$  und daher gilt  $\|F(x)\| \leq C$  für alle  $x \in K$ . Zu jedem  $x \in K$  gibt es  $\delta_x > 0$  und  $L_x > 0$ , so dass  $F$  auf  $B_{\delta_x}(x) \cap K$  Lipschitz-stetig mit Konstante  $L_x$  ist. Die Überdeckung  $B_{\delta_x}(x)$ ,  $x \in K$ , von  $K$  enthält eine endliche Teilüberdeckung  $B_{\delta_{x_j}}(x_j)$ ,  $1 \leq j \leq N$ . Setze  $\delta := \min_j \delta_{x_j}$  und  $L := \max_j L_{x_j}$ .

Sind nun  $x, y \in K$  beliebig, so gilt entweder  $\|x - y\| \geq \delta$  und somit

$$\|F(x) - F(y)\| \leq 2C \leq \frac{2C}{\delta} \|x - y\|$$

oder es gilt  $\|x - y\| < \delta$  und es gibt dann  $i$  mit  $y \in B_{\delta_{x_i}}(x_i)$ , also  $x \in B_{\delta_{x_i} + \delta}(x_i) \subset B_{2\delta_{x_i}}(x_i)$ , woraus

$$\|F(x) - F(y)\| \leq L_i \|x - y\| \leq L \|x - y\|$$

folgt.

Ist  $M$  kompakt und konvex, so können wir anders verfahren: Mit obiger endlicher Überdeckung (wobei es nun sogar genügt, dass  $L_{x_j}$  die Lipschitz-Konstante auf  $B_{\delta_{x_j}}(x_j)$  ist) betrachte beliebige  $x, y \in M$ . Die Strecke  $[x, y]$  liegt in  $M$  und wird von den  $N$  Kugeln überdeckt. Der Schnitt von  $[x, y]$  mit jeder Kugel ist leer oder ein Intervall. Wir können nun  $[x, y]$  in  $m \leq N$  Teilintervalle  $[z_i, z_{i+1}]$  zerlegen mit  $z_i = x + t_i(y - x)$ ,  $0 = t_0 < t_1 < \dots < t_m = 1$ , wobei jedes dieser Intervalle vollständig in einer der  $N$  Kugeln enthalten ist. Es gilt dann

$$\|F(y) - F(x)\| \leq \sum_{i=0}^{m-1} \|F(z_{i+1}) - F(z_i)\| \leq \sum_{i=0}^{m-1} L \|z_{i+1} - z_i\| = L \|y - x\|.$$

$\square$

Wir wenden uns nun der Konvergenzgeschwindigkeit zu:

**Satz 2.5.8.** *Unter den Voraussetzungen von Satz 2.5.6 gilt:*

- a)  $\liminf_{k \rightarrow \infty} \sqrt{k}(f(x^k) - f(x^*)) = 0.$   
 b) *Gibt es ein  $\gamma > 0$  mit*

$$f(x^k) - f(x^*) \geq \gamma \|x^k - x^*\| \quad \forall k,$$

*dann konvergiert  $x^k$   $q$ -linear gegen  $x^*$ :*

$$\|x^{k+1} - x^*\| \leq \alpha \|x^k - x^*\| \quad \text{mit} \quad \alpha = \sqrt{1 - \frac{\gamma^2}{C^2}}, \quad C = \sup_k \|g^k\|.$$

**Bemerkung.** Natürlich muss man in Teil b) für  $x^*$  den Grenzwert von  $(x^k)$  wählen. Dieser existiert nach Satz 2.5.6 c) und ist ein globales Minimum von  $f$ .

**Beweis:** Wir setzen auf dem Beweis von Satz 2.5.6 auf.

zu a): Angenommen, a) gilt nicht. Dann gibt es  $\varepsilon > 0$  und  $K > 0$  mit

$$\sqrt{k}(f(x^k) - f(x^*)) \geq \varepsilon \quad \forall k \geq K.$$

Nun ergibt sich aus (2.12) und (2.13)

$$R^2 \geq \sum_{k=0}^{\infty} \rho_k(\sigma_k) = \sum_{k=0}^{\infty} \frac{(f(x^k) - f(x^*))^2}{\|g^k\|^2} \geq \frac{1}{C^2} \sum_{k=K}^{\infty} (f(x^k) - f(x^*))^2 \geq \frac{1}{C^2} \sum_{k=K}^{\infty} \frac{\varepsilon^2}{k}.$$

Dies ist ein Widerspruch zur Divergenz der harmonischen Reihe. Also stimmt a) doch.

zu b):

$$\begin{aligned} \|x^{k+1} - x^*\|^2 &\leq \|x^k - x^*\|^2 - \rho_k(\sigma_k) = \|x^k - x^*\|^2 - \frac{(f(x^k) - f(x^*))^2}{\|g^k\|^2} \\ &\leq \|x^k - x^*\|^2 - \frac{\gamma^2 \|x^k - x^*\|^2}{\|g^k\|^2} \leq \left(1 - \frac{\gamma^2}{C^2}\right) \|x^k - x^*\|^2. \end{aligned}$$

□

## 2.5.2 Subgradienten-Verfahren bei unbekanntem Optimalwert

Wie wir gesehen haben, besitzt das Subgradienten-Verfahren bei Kenntnis des Optimalwerts relativ gute Konvergenzeigenschaften. Ist jedoch  $f(x^*)$  unbekannt, dann wissen wir nicht, wie wir  $\sigma_k$  wählen müssen. Es gibt hier viele, oft heuristische Verfahrensvarianten, diesem Problem beizukommen.

Wir behandeln hier stellvertretend ein sehr universelles Verfahren zur Lösung von (2.8):

### Algorithmus 2.5.9.

Wähle  $x^0 \in X$ .

Für  $k = 0, 1, 2, \dots$ :

1. Berechne einen Subgradienten  $g^k \in \partial f(x^k)$ . Falls ein geeignetes Abbruchkriterium erfüllt ist, STOP mit Ergebnis  $x^k$ .
2. Setze  $s^k = -\frac{g^k}{\|g^k\|}$ .
3. Ermittle eine Schrittweite  $t_k > 0$  und setze  $x^{k+1} = P_X(x^k + t_k s^k)$ .

Wir betrachten sehr allgemeine Schrittweitenwahlen und setzen lediglich voraus, dass die Schrittweiten folgendes erfüllen:

$$(2.14) \quad t_k \rightarrow 0, \quad \sum_{k=0}^{\infty} t_k = \infty.$$

Die zweite Bedingung ist deshalb sinnvoll, weil wir a priori nicht wissen, wie weit die Lösung  $x^*$  von  $x^0$  entfernt ist (außer wir wüssten etwas über die maximale Ausdehnung von  $X$ , aber solches Spezialwissen wollen wir hier nicht einbeziehen), und weil gilt:

$$\|x^k - x^0\| \leq \sum_{l=0}^{k-1} \|x^{l+1} - x^l\| \leq \sum_{l=0}^{k-1} t_l.$$

Wir weisen nun die Konvergenz des Verfahrens nach:

**Satz 2.5.10.** Die Funktion  $f : U \rightarrow \mathbb{R}$  sei konvex auf der offenen konvexen Menge  $U \subset \mathbb{R}^n$ . Die Menge  $X \subset U$  sei abgeschlossen und konvex und  $x^* \in X$  sei eine Lösung von (2.8). Algorithmus 2.5.9 verwende Schrittweiten, die (2.14) genügen. Dann terminiert das Verfahren entweder aufgrund des Abbruchkriteriums in Schritt 1 oder es erzeugt eine Folge, für die gilt:

a)

$$(2.15) \quad \liminf_{k \rightarrow \infty} f(x^k) = f(x^*).$$

b) Gilt

$$\lim_{\substack{x \in X \\ \|x\| \rightarrow \infty}} f(x) = \infty,$$

so gibt es eine Teilfolge  $(x^{k'})$ , die gegen ein globales Minimum von  $f$  konvergiert.

**Bemerkung.** Die Bedingung in b) ist äquivalent zur Kompaktheit der Menge globaler Minima. Beweis zur Übung.

**Beweis:** Zu betrachten ist lediglich der Fall einer unendlichen Folge von Iterierten  $(x^k)$ .

zu a): Angenommen, (2.15) gilt nicht. Dann gibt es  $\varepsilon > 0$  und  $K > 0$  mit

$$f(x^k) > f(x^*) + \varepsilon \quad \forall k \geq K.$$

Die Menge  $M = \{x \in U; f(x) < f(x^*) + \varepsilon\}$  ist dann offen und enthält kein einziges  $x^k$ ,  $k \geq K$ .

Sei nun  $\delta > 0$  so, dass gilt  $\bar{B}_\delta(x^*) \subset M$ . Weiter sei  $y^k = x^* - \delta s^k$ . Dann gilt  $y^k \in \bar{B}_\delta(x^*)$ .

Nun ergibt sich mit der Subgradienten-Ungleichung für alle  $k \geq K$ :

$$\begin{aligned}
\|x^{k+1} - x^*\|^2 - \|x^k - x^*\|^2 &= \|P_X(x^k + t_k s^k) - P_X(x^*)\|^2 - \|x^k - x^*\|^2 \\
&\leq \|x^k + t_k s^k - x^*\|^2 - \|x^k - x^*\|^2 \\
&= \|(t_k s^k) + (x^k - x^*)\|^2 - \|x^k - x^*\|^2 \\
&= \|t_k s^k\|^2 + 2(t_k s^k)^T (x^k - x^*) \\
&= t_k^2 + 2t_k s^{kT} (y^k - x^*) + 2t_k s^{kT} (x^k - y^k) \\
&= t_k^2 - 2\delta t_k + 2 \frac{t_k}{\|g^k\|} g^{kT} (y^k - x^k) \\
&\leq t_k^2 - 2\delta t_k + 2 \frac{t_k}{\|g^k\|} \underbrace{(f(y^k) - f(x^k))}_{<0} \leq t_k^2 - 2\delta t_k.
\end{aligned}$$

Wegen  $t_k \rightarrow 0$  gilt nach eventuellem Vergrößern von  $K$ :

$$\|x^{k+1} - x^*\|^2 - \|x^k - x^*\|^2 \leq t_k^2 - 2\delta t_k \leq -\delta t_k.$$

Dies ergibt durch Aufsummieren

$$-\|x^K - x^*\|^2 \leq \|x^{K+l} - x^*\|^2 - \|x^K - x^*\|^2 \leq -\delta \sum_{k=K}^{K+l-1} t_k \rightarrow -\infty \quad (l \rightarrow \infty),$$

also einen Widerspruch.

zu b): Nach a) gibt es eine Teilfolge mit  $f(x^{k'}) \rightarrow f(x^*)$ . Nach Voraussetzung ist die Folge  $(x^{k'})$  somit beschränkt und besitzt daher eine konvergente Teilfolge  $x^{k''} \rightarrow \bar{x}$ . Wegen der Stetigkeit von  $f$  gilt

$$f(\bar{x}) = \lim_{k'' \rightarrow \infty} f(x^{k''}) = f(x^*).$$

□

Leider ist die Konvergenzgeschwindigkeit recht unbefriedigend:

**Satz 2.5.11.** *Wir betrachten den unrestringierten Fall  $X = \mathbb{R}^n$  (d.h. ohne Projektion) unter den Voraussetzungen von Satz 2.5.10. Die Folge  $(x^k)$  konvergiere gegen den Punkt  $x^*$ . Dann ist die Konvergenzgeschwindigkeit schlechter als  $r$ -linear:*

$$\limsup_{k \rightarrow \infty} \frac{\|x^k - x^*\|}{\gamma^k} = \infty \quad \forall \gamma \in (0, 1).$$

**Beweis:** Angenommen, die Aussage ist falsch, dann gibt es  $K > 0$ ,  $C > 0$  und  $\gamma \in (0, 1)$  mit

$$\|x^k - x^*\| \leq C\gamma^k \quad \forall k \geq K.$$

Dies zeigt für alle  $k \geq K$

$$t_k = \|x^{k+1} - x^k\| \leq \|x^{k+1} - x^*\| + \|x^k - x^*\| \leq C\gamma^{k+1} + C\gamma^k \leq 2C\gamma^k.$$

Damit ergibt sich aus (2.14):

$$\infty = \sum_{k=K}^{\infty} t_k \leq \sum_{k=K}^{\infty} 2C\gamma^k = \frac{2C\gamma^K}{1-\gamma} < \infty.$$

Dies ist ein Widerspruch.

□



## 2.6 Schnittebenen-Verfahren

Das Schnittebenen-Verfahren nutzt die Tatsache aus, dass für jedes Punkt-Subgradient-Paar  $(x^k, g^k)$  mit  $x^k \in X$  und  $g^k \in \partial f(x^k)$  die lineare Funktion

$$l_k(x) = f(x^k) + g^{kT}(x - x^k)$$

eine Minorante von  $f$  ist, d.h. auf  $U$  gilt  $f(x) \geq l_k(x)$ .

In der  $k$ -ten Iteration des Schnittebenen-Verfahrens stehen die bisherigen Iterierten  $x^j$ , die zugehörigen Funktionswerte  $f^j = f(x^j)$  und jeweils ein Subgradient  $g^j \in \partial f(x^j)$  zur Verfügung. Wir verwenden nun als Modell von  $f$  die Minorante

$$f_k^{\text{se}}(x) = \max_{0 \leq j \leq k} l_j(x) = \max_{0 \leq j \leq k} (f^j + g^{jT}(x - x^j)).$$

Aus  $f \geq l_k$  folgt

$$f(x) \geq f_k^{\text{se}}(x) \quad \forall x \in U.$$

Da der Graph von  $l_k$  jenen von  $f$  im Punkt  $(x^k, f(x^k))$  stützt, ist klar, dass die Hinzunahme jeder neuen Minorante  $l_k$  das Schnittebenenmodell von  $f$  verbessert, insbesondere in der Nähe von  $x^k$ .

Die einfache Idee besteht nun darin, die neue Iterierte  $x^{k+1}$  als Lösung des folgenden Problems zu berechnen:

$$(2.16) \quad \min_x f_k^{\text{se}}(x) \quad \text{u.d.N.} \quad x \in X.$$

Natürlich ist dies nur sinnvoll, wenn das Problem (2.16) (es ist selbst ein nichtglattes Optimierungsproblem) wesentlich einfacher lösbar ist als das Ausgangsproblem (2.8). Ob dies so ist, hängt von der Menge  $X$  ab.

In jedem Fall können wir (2.16) als glattes Problem schreiben (siehe Lemma 2.6.1):

$$(2.17) \quad \min_{x \in \mathbb{R}^n, \alpha \in \mathbb{R}} \alpha \quad \text{u.d.N.} \quad x \in X, \quad f^j + g^{jT}(x - x^j) - \alpha \leq 0, \quad 0 \leq j \leq k.$$

Im Fall  $X = \mathbb{R}^n$  oder  $X = \{x; Ax = b, x \geq 0\}$  oder  $X = \{x; Ax \leq b\}$  ist (2.17) ein lineares Optimierungsproblem und daher relativ leicht zu lösen.

**Lemma 2.6.1.** *Gegeben seien Funktionen  $f_i : X \rightarrow \mathbb{R}$ ,  $i = 1, \dots, m$ , mit  $X \subset \mathbb{R}^n$ . Wir betrachten das (nichtglatte) Problem*

$$(2.18) \quad \min_{x \in X} f(x) \quad \text{mit} \quad f(x) = \max_{1 \leq i \leq m} f_i(x)$$

*und parallel dazu das Problem*

$$(2.19) \quad \min_{x \in \mathbb{R}^n, \alpha \in \mathbb{R}} \alpha \quad \text{u.d.N.} \quad x \in X, \quad f_i(x) \leq \alpha, \quad i = 1, \dots, m.$$

*Dann gilt:*

- a) *Ist  $x^*$  eine globale Lösung von (2.18), so ist  $(x^*, f(x^*))$  eine globale Lösung von (2.19).*
- b) *Ist  $(x^*, \alpha^*)$  eine globale Lösung von (2.19), so ist  $x^*$  eine globale Lösung von (2.18) und es gilt  $f(x^*) = \alpha^*$ .*

**Beweis:** Sei  $x^*$  Lösung von (2.18) und  $\alpha^* = f(x^*)$ . Dann ist natürlich  $(x^*, \alpha^*)$  zulässig für (2.19). Ist nun  $(x, \alpha)$  zulässig für (2.19), so gilt  $x \in X$  und daher

$$\alpha^* = f(x^*) \leq f(x) = \max_{1 \leq i \leq m} f_i(x) \leq \alpha.$$

Damit ist  $(x^*, \alpha^*)$  optimale Lösung von (2.19).

Sei nun  $(x^*, \alpha^*)$  optimale Lösung von (2.19). Dann gilt  $\alpha^* = f(x^*)$ , denn sonst wäre  $(x^*, f(x^*))$  ein zulässiger Punkt mit kleinerem Zielfunktionswert.

Ist nun  $x \in X$  beliebig, so ist  $(x, f(x))$  zulässig für (2.19) und daher folgt

$$f(x^*) = \alpha^* \leq f(x),$$

so dass  $x^*$  optimale Lösung von (2.18) ist. □

Der Name Schnittebenen-Verfahren leitet sich von den Teilproblemen (2.17) ab:

Jede der linearen Ungleichungen

$$f^j + g^{jT}(x - x^j) - \alpha \leq 0$$

beschreibt einen affinen Halbraum im  $\mathbb{R}^{n+1}$  und schneidet somit in der Regel von  $X \times \mathbb{R}$  ein Stück weg ( $\rightarrow$  Schnittebene).

Der zulässige Bereich von (2.17) wird also in jeder Iteration kleiner, weil eine neue Schnittebene hinzukommt. Wie wir in Satz 2.6.4 sehen werden, ist der Punkt  $x^k$  optimale Lösung von (2.8), falls die Hinzunahme der  $k$ -ten Schnittebene die Lösungsmenge von (2.17) nicht ändert. Der zulässige Bereich von (2.17) stimmt überein mit dem Epigraphen  $\text{epi}(f_k^{\text{se}})$  der Funktion  $f_k^{\text{se}}$  und es gilt

$$\text{epi}(f_k^{\text{se}}) \supset \text{epi}(f_{k+1}^{\text{se}}) \supset \text{epi}(f) = \{(x^T, \alpha)^T \in \mathbb{R}^{n+1}; f(x) \leq \alpha\} \quad \forall k.$$

Wir betrachten nun das folgende Verfahren:

**Algorithmus 2.6.2** (Schnittebenen-Verfahren).

0. Wähle  $x^0 \in X$ .

Für  $k = 0, 1, 2, \dots$ :

1. Berechne einen Subgradienten  $g^k \in \partial f(x^k)$ .
2. STOP, falls ein geeignetes Abbruchkriterium erfüllt ist.
3. Berechne  $x^{k+1}$  durch Lösen des Problems (2.16) (oder, äquivalent dazu, von (2.17)).

**Satz 2.6.3.** Algorithmus 2.6.2 erzeuge eine unendliche Folge  $(x^k)$ . Dann ist jeder Häufungspunkt von  $(x^k)$  eine optimale Lösung von (2.8).

**Beweis:** Sei  $(x^k)_K$  eine Teilfolge mit  $(x^k)_K \rightarrow x^*$ . Da  $X$  abgeschlossen ist, gilt  $x^* \in X$ . Weiter ergibt sich für alle  $0 \leq j < k$  und alle  $x \in X$ :

$$\begin{aligned} f(x) &\geq f_{k-1}^{\text{se}}(x) \geq f_{k-1}^{\text{se}}(x^k) \geq f^j + g^{jT}(x^k - x^j) \\ &\xrightarrow{K \ni k \rightarrow \infty} f^j + g^{jT}(x^* - x^j) \xrightarrow{K \ni j \rightarrow \infty} f(x^*). \end{aligned}$$

Beim zweiten Grenzübergang wurde die Stetigkeit von  $f$  benutzt und zusätzlich, dass wegen  $(x^j)_K \rightarrow x^*$  die Abschätzung  $\|g^j\| \leq L$  für große  $j$  gilt, wobei  $L$  die Lipschitz-Konstante von  $f$  nahe  $x^*$  ist.

Aus  $f(x) \geq f(x^*)$  für alle  $x \in X$  folgt, dass  $x^*$  globale Lösung von (2.8) ist.  $\square$

Der folgende Satz zeigt, dass es beim Schnittebenen-Verfahren möglich ist, ein zuverlässiges Abbruchkriterium anzugeben.

**Satz 2.6.4.** *Algorithmus 2.6.2 erzeuge die Iterierten  $x^k$ . Ist dann  $x^*$  optimale Lösung von (2.8) und gilt*

$$f(x^{k+1}) - f_k^{\text{se}}(x^{k+1}) \leq \varepsilon$$

*für ein  $\varepsilon \geq 0$ , so gilt auch*

$$f(x^{k+1}) - f(x^*) \leq \varepsilon$$

**Beweis:** Es gilt

$$f(x^*) \geq f_k^{\text{se}}(x^*) \geq f_k^{\text{se}}(x^{k+1}) \geq f(x^{k+1}) - \varepsilon.$$

$\square$

Wir können also in Schritt 2 das folgende Abbruchkriterium verwenden:

2. Falls  $g^k = 0$  oder  $k \geq 1$  und  $f(x^k) - f_{k-1}^{\text{se}}(x^k) \leq \varepsilon$ , STOP mit Ergebnis  $x^k$ .

**Bemerkung.** Definitionsgemäß ist  $x^{k+1}$  optimale Lösung von (2.16). Tritt der Fall ein, dass die alte Iterierte  $x^k$  bereits Lösung von (2.16) ist, dann kann  $x^{k+1} = x^k$  gewählt werden und wir erhalten mit Satz 2.6.4 (für  $\varepsilon = 0$ )

$$f_k^{\text{se}}(x^{k+1}) = f_k^{\text{se}}(x^k) = f(x^k) = f(x^{k+1}).$$

Damit ist Satz 2.6.4 für  $\varepsilon = 0$  anwendbar und liefert

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

d.h.  $x^{k+1} = x^k$  optimale Lösung von (2.8).

Dies zeigt auch, dass jede neue Schnittebene den zulässigen Bereich von (2.17) echt verkleinert, solange  $x^k$  nicht bereits eine optimale Lösung ist.

Eine grundsätzliche Problematik des Schnittebenen-Verfahrens besteht darin, dass das Problem (2.17) unter Umständen keine Lösung besitzt, wenn  $X$  nicht kompakt ist. Besonders eklatant ist dies im Fall  $X = \mathbb{R}^n$ , denn für  $k = 0$  und  $g^0 \neq 0$  ist die Funktion  $f^{\text{se}} = l_0$  affin-linear, also nicht nach unten beschränkt.

Einer von mehreren möglichen Auswegen besteht darin, eine Regularisierung hinzuzufügen, die dafür sorgt, dass die Zielfunktion gleichmäßig konvex ist:

$$(2.20) \quad \min_x f_k^{\text{se}}(x) + \frac{1}{2\gamma_k} \|x - x^k\|^2 \quad \text{u.d.N.} \quad x \in X.$$

Der Parameter  $\gamma_k > 0$  muss natürlich geeignet gewählt werden. Die Teilprobleme (2.20) bilden die Basis der Bundle-Verfahren, auf die wir nun hinarbeiten.

## 2.7 Das $\varepsilon$ -Subdifferential

Ein großer Nachteil der Richtungsableitung und des Subdifferentials besteht darin, dass man anhand von  $f'(x, \cdot)$  bzw.  $\partial f(x)$  nicht erkennen kann, ob sich  $x$  in der Nähe eines Minimums von  $f$  befindet (betrachte etwa  $f(x) = |x|$ ). Nur, wenn  $x$  bereits das Minimum ist, erkennen wir dies anhand von  $f'(x, \cdot)$  bzw.  $\partial f(x)$ .

Das Problem besteht darin, dass  $f'(x, \cdot)$  und  $\partial f(x)$  keine Umgebungsinformation enthalten.

Wir führen daher nun ein Subdifferential ein, das diese Schwierigkeiten überwindet.

**Definition 2.7.1** ( $\varepsilon$ -Subgradient,  $\varepsilon$ -Subdifferential). Sei  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  konvex und  $\varepsilon \geq 0$ . Der Vektor  $g \in \mathbb{R}^n$  heißt  $\varepsilon$ -Subgradient von  $f$  im Punkt  $x \in \mathbb{R}^n$ , wenn gilt:

$$(2.21) \quad f(y) - f(x) \geq g^T(y - x) - \varepsilon \quad \forall y \in \mathbb{R}^n.$$

Die Menge  $\partial_\varepsilon f(x) \subset \mathbb{R}^n$ ,

$$\partial_\varepsilon f(x) = \{g \in \mathbb{R}^n; g \text{ ist } \varepsilon\text{-Subgradient von } f \text{ in } x\}$$

heißt  $\varepsilon$ -Subdifferential von  $f$  im Punkt  $x \in \mathbb{R}^n$ . Das  $\varepsilon$ -Subdifferential induziert eine mengenwertige Abbildung  $\partial_\varepsilon f : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ .

### Anschauliche Interpretation:

Der Vektor  $g$  ist  $\varepsilon$ -Subgradient von  $f$  in  $\bar{x}$ , falls die durch den Punkt  $(\bar{x}, f(\bar{x}) - \varepsilon)$  verlaufende lineare Funktion  $l$  mit Gradient  $g$ , also

$$l(x) = f(\bar{x}) + g^T(x - \bar{x}) - \varepsilon$$

auf oder unterhalb des Graphen von  $f$  verläuft.

Aufgrund der Definition ist folgendes klar:

$$\partial f(x) = \partial_0 f(x) \subset \partial_\varepsilon f(x) \quad \forall \varepsilon \geq 0, x \in \mathbb{R}^n.$$

Wir können auch eine entsprechende Richtungsableitung definieren:

**Definition 2.7.2** ( $\varepsilon$ -Richtungsableitung). Sei  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  konvex und  $\varepsilon \geq 0$ . Die  $\varepsilon$ -Richtungsableitung von  $f$  im Punkt  $x$  in Richtung  $s \in \mathbb{R}^n$  ist definiert gemäß

$$f'_\varepsilon(x, s) = \inf_{t>0} \frac{f(x + ts) - f(x) + \varepsilon}{t}.$$

### Anschauliche Interpretation:

Betrachten wir  $f$  ausgehend vom Punkt  $x$  entlang der Richtung  $s$ , so ergibt sich die Funktion  $\phi : \mathbb{R}_+ \rightarrow \mathbb{R}$ ,  $\phi(t) = f(x + ts)$ ,  $t \geq 0$ . Der Wert  $a = f'_\varepsilon(x, s)$  ist nun die Steigung jener Geraden  $g : \mathbb{R}_+ \rightarrow \mathbb{R}$ ,  $g(t) = f(x) + at - \varepsilon$ , die durch den Punkt  $(0, f(x) - \varepsilon)$  verläuft und den Graphen der Funktion  $\phi$  von unten berührt, wobei diese Interpretation nur zutrifft, wenn das Infimum in der Definition von  $f'_\varepsilon(x, s)$  in einem  $t = t^*$  angenommen wird. Der Berührungspunkt ist dann  $((x + t^*s)^T, f(x + t^*s))^T$ .

Die folgende Darstellung ist kanonisch und folgt im wesentlichen [GK02]. Wir stellen zunächst einige Zusammenhänge zwischen  $f'$  und  $f'_\varepsilon$  her:

**Satz 2.7.3.** Sei  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  konvex. Dann gilt für alle  $x, s \in \mathbb{R}^n$ :

- a)  $f'(x, s) = f'_0(x, s)$ .
- b)  $f'(x, s) \leq f'_\varepsilon(x, s)$ .

**Beweis:**

zu a): Folgt sofort aus Satz 2.4.1 b).

zu b): Für  $t > 0$  gilt

$$\frac{f(x + ts) - f(x)}{t} \leq \frac{f(x + ts) - f(x) + \varepsilon}{t}$$

Bilden von  $\inf_{t>0}$  liefert nun die Behauptung. □

Für  $\partial_\varepsilon f$  und  $f'_\varepsilon(\cdot, \cdot)$  gelten ganz ähnliche Aussagen wie für  $\partial f$  und  $f'(\cdot, \cdot)$ .

Wir formulieren nun eine  $\varepsilon$ -Entsprechung zu Satz 2.4.3:

**Satz 2.7.4.** Sei  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  konvex und  $x \in \mathbb{R}^n$ . Dann gilt:

- a)  $\partial_\varepsilon f(x) = \{g \in \mathbb{R}^n; g^T s \leq f'_\varepsilon(x, s) \ \forall s \in \mathbb{R}^n\}$ .
- b)  $\partial_\varepsilon f(x)$  ist nichtleer, konvex und kompakt.
- c)  $f'_\varepsilon(x, s) = \max_{g \in \partial_\varepsilon f(x)} g^T s \ \forall s \in \mathbb{R}^n$ .

**Beweis:** Erfolgt in ganz ähnlicher Weise wie der Nachweis der entsprechenden Aussagen in Satz 2.4.3.

**Satz 2.7.5.** Sei  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  eine konvexe Funktion,  $x^* \in \mathbb{R}^n$  und  $\varepsilon \geq 0$ . Dann sind die folgenden Aussagen äquivalent:

- a) Der Punkt  $x^*$  ist  $\varepsilon$ -optimal, d.h.

$$f(x^*) \leq f(x) + \varepsilon \quad \forall x \in \mathbb{R}^n.$$

- b) Es gilt  $f'_\varepsilon(x^*, s) \geq 0 \quad \forall s \in \mathbb{R}^n$ .
- c) Es gilt  $0 \in \partial_\varepsilon f(x^*)$ .

**Beweis:**

a)  $\implies$  b):

Für beliebiges  $s \in \mathbb{R}^n$  folgt  $f(x^* + ts) + \varepsilon \geq f(x^*)$  für  $t > 0$  und somit

$$f'_\varepsilon(x^*, s) = \inf_{t>0} \frac{f(x^* + ts) - f(x^*) + \varepsilon}{t} \geq 0.$$

b)  $\implies$  c):

Wegen

$$0^T s = 0 \leq f'_\varepsilon(x^*, s) \quad \forall s \in \mathbb{R}^n$$

folgt  $0 \in \partial_\varepsilon f(x^*)$  gemäß Satz 2.7.4 a).

c)  $\implies$  a):

Da 0 ein  $\varepsilon$ -Subgradient ist, folgt nach Definition

$$f(x) - f(x^*) \geq 0^T(x - x^*) - \varepsilon = -\varepsilon \quad \forall x \in \mathbb{R}^n.$$

Somit ist  $x^*$   $\varepsilon$ -optimal. □

Zu Beginn hatten wir versprochen, dass für  $\varepsilon > 0$  das  $\varepsilon$ -Subdifferential  $\partial_\varepsilon f(x)$  Informationen aus der Umgebung von  $x$  enthält. Dies wird nun präzisiert:

**Satz 2.7.6.** *Sei  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  konvex,  $x \in \mathbb{R}^n$ . Dann gilt:*

a) *Zu  $\varepsilon > 0$  gibt es  $\delta > 0$ , so dass gilt:*

$$\bigcup_{y \in B_\delta(x)} \partial f(y) \subset \partial_\varepsilon f(x).$$

b) *Zu  $\delta > 0$  gibt es  $\varepsilon > 0$ , so dass gilt:*

$$\bigcup_{y \in B_\delta(x)} \partial f(y) \subset \partial_\varepsilon f(x).$$

**Beweis:**

Fall a): Die Funktion  $f$  ist nach Satz 2.4.1 a) auf einer offenen Umgebung  $U$  von  $x$  Lipschitz-stetig mit Konstante  $L > 0$ . Wähle nun  $\delta > 0$  mit  $B_\delta(x) \subset U$  und  $2L\delta \leq \varepsilon$ . Der Rest des Beweises wird gemeinsam mit b) geführt.

Fall b): Die Funktion  $f$  ist nach Satz 2.4.1 a) und Lemma 2.5.7 Lipschitz-stetig auf  $\bar{B}_\delta(x)$  mit Konstante  $L > 0$ . Wähle nun  $\varepsilon \geq 2L\delta$ .

Fall a) und b) gemeinsam:

Für alle  $y \in B_\delta(x)$  und alle  $g \in \partial f(y)$  gilt  $\|g\| \leq L$  nach Satz 2.4.3 b) und daher ergibt sich für alle  $z \in \mathbb{R}^n$ :

$$\begin{aligned} g^T(z - x) &= g^T(z - y) + g^T(y - x) \leq f(z) - f(y) + \|g\| \|y - x\| \\ &\leq f(z) - f(x) + |f(x) - f(y)| + \|g\| \|y - x\| \\ &\leq f(z) - f(x) + L\|x - y\| + L\|y - x\| \\ &\leq f(z) - f(x) + 2L\delta \leq f(z) - f(x) + \varepsilon. \end{aligned}$$

Dies zeigt  $g \in \partial_\varepsilon f(x)$ . □

Im Gegensatz zu  $f'$  können wir mit  $f'_\varepsilon$  robuste Abstiegsrichtungen berechnen. Genauer gilt:

**Satz 2.7.7.** *Sei  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  konvex,  $\varepsilon \geq 0$  und  $x \in \mathbb{R}^n$  mit  $0 \notin \partial_\varepsilon f(x)$ . Weiter sei  $s = -g$  mit  $g = P_{\partial_\varepsilon f(x)}(0)$ . Dann ist  $\frac{s}{\|s\|}$  Lösung des Problems*

$$(2.22) \quad \min_{\|d\|=1} f'_\varepsilon(x, d)$$

und es gilt

$$(2.23) \quad f'_\varepsilon(x, s) = -\|s\|^2 = -\|g\|^2 < 0$$

Ist weiter  $d$  eine Richtung mit  $f'_\varepsilon(x, d) < 0$  (z.B. obige Richtung  $s$ ), dann gilt

$$(2.24) \quad \inf_{t>0} f(x + ts) < f(x) - \varepsilon.$$

**Beweis:** Für  $d \in \mathbb{R}^n$  mit  $\|d\| = 1$  gilt  $v^T d \geq -\|v\|$  für alle  $v \in \mathbb{R}^n$  und somit

$$\begin{aligned} f'_\varepsilon(x, d) &= \max_{v \in \partial_\varepsilon f(x)} v^T d \geq \max_{v \in \partial_\varepsilon f(x)} -\|v\| = -\min_{v \in \partial_\varepsilon f(x)} \|v\| \\ &= -\|P_{\partial_\varepsilon f(x)}(0)\| = -\|g\| = -\|s\|. \end{aligned}$$

Weiter ergibt sich für alle  $v \in \partial_\varepsilon f(x)$  wegen der Eigenschaft der Projektion

$$g^T(v - g) = (P_{\partial_\varepsilon f(x)}(0) - 0)^T(v - P_{\partial_\varepsilon f(x)}(0)) \geq 0 \quad \forall v \in \partial_\varepsilon f(x).$$

Daraus folgt

$$\min_{v \in \partial_\varepsilon f(x)} v^T g = \|g\|^2.$$

Dies zeigt

$$\begin{aligned} f'_\varepsilon\left(x, \frac{s}{\|s\|}\right) &= \frac{1}{\|g\|} f'_\varepsilon(x, -g) = \frac{1}{\|g\|} \max_{v \in \partial_\varepsilon f(x)} v^T(-g) = -\frac{1}{\|g\|} \min_{v \in \partial_\varepsilon f(x)} v^T g \\ &= -\frac{\|g\|^2}{\|g\|} = -\|g\| = -\|s\|. \end{aligned}$$

Damit ist  $s/\|s\|$  Lösung von (2.22) und es gilt

$$f'_\varepsilon(x, s) = \|s\| f'_\varepsilon\left(x, \frac{s}{\|s\|}\right) = -\|s\|^2 < 0.$$

Ist nun  $d$  eine Richtung mit  $f'_\varepsilon(x, d) < 0$ , dann folgt

$$\inf_{t>0} \frac{f(x + td) - f(x) + \varepsilon}{t} = f'_\varepsilon(x, d) < 0.$$

Daher gibt es  $t^* > 0$  mit

$$\frac{f(x + t^*d) - f(x) + \varepsilon}{t^*} < 0.$$

Daraus erhalten wir

$$\inf_{t>0} f(x + td) \leq f(x + t^*d) < f(x) - \varepsilon.$$

□

Wir können nun in Analogie zur Methode des steilsten Abstiegs das folgende Verfahren betrachten:

**Algorithmus 2.7.8** (Modellalgorithmus).

0. Wähle  $x^0 \in \mathbb{R}^n$  und  $\varepsilon > 0$ .

Für  $k = 0, 1, 2, \dots$ :

1. Bestimme  $g^k = P_{\partial_\varepsilon f(x^k)}(0)$ .
2. Falls  $g^k = 0$ , STOP.
3. Setze  $s^k = -g^k$  und ermittle die optimale Schrittweite  $\sigma_k \geq 0$  entlang  $s^k$ :

$$f(x^k + \sigma_k s^k) = \min_{\sigma \geq 0} f(x^k + \sigma s^k).$$

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

**Bemerkung.** Es würde genügen, wenn die Schrittweite  $\sigma_k$  so gewählt wird, dass sie einen Teil des maximal möglichen Abstiegs realisiert. Zudem könnte man auch mit anderen Richtungen, die  $f'_\varepsilon(x^k, s_k) < 0$  erfüllen, arbeiten.

Dieses Verfahren hat sehr schöne Konvergenzeigenschaften. Es gilt nämlich:

**Satz 2.7.9.** *Sei  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  konvex und nach unten beschränkt. Weiter sei  $\varepsilon > 0$ . Dann terminiert Algorithmus 2.7.8 nach endlich vielen Iterationen mit einem  $\varepsilon$ -optimalen Punkt von  $f$ :*

$$f(x^k) \leq \inf_{x \in \mathbb{R}^n} f(x) + \varepsilon.$$

**Beweis:** Solange  $g^k \neq 0$  ist, gilt  $0 \notin \partial_\varepsilon f(x^k)$  und somit ergibt sich nach Satz 2.7.7:

$$f(x^{k+1}) < f(x^k) - \varepsilon.$$

Irgendwann gilt dann  $f(x^k) \leq \inf_x f(x) + \varepsilon$  und dann muss spätestens  $g^k = 0$  gelten, da eine  $f$ -Abnahme um mehr als  $\varepsilon$  nicht mehr möglich ist. Umgekehrt folgt aus  $g^k = 0$ , dass  $x^k$   $\varepsilon$ -optimal ist, siehe Satz 2.7.5.  $\square$

Da  $\partial_\varepsilon f(x)$  in der Praxis schwer oder gar nicht zu berechnen ist, hat Algorithmus 2.7.8 nur theoretischen Wert. Er legt aber nahe,  $\partial_\varepsilon f(x^k)$  geeignet durch eine Menge  $G_\varepsilon^k$  zu approximieren und dann  $s^k = -P_{G_\varepsilon^k}(0)$  als Suchrichtung zu verwenden. Genau dies macht das Bundle-Verfahren, dem wir uns aber zunächst aus dem Blickwinkel der Schnittebenen-Verfahren nähern wollen.

## 2.8 Bundle Methoden

Bundle (=Bündel) Methoden bilden eine der effizientesten Verfahrensklasse der nichtglatten Optimierung. Die grundlegende Idee besteht darin, wie beim Schnittebenen-Verfahren aus den bisher berechneten Funktionswerten und Subgradienten ein „Bündel“ aus Informationen zu schnüren und dieses zu verwenden, um Suchrichtungen zu berechnen.

Es gibt nun zwei Sichtweisen, die dual zueinander sind:

Die eine interpretiert das Bundle-Verfahren als ein regularisiertes Schnittebenen-Verfahren und verwendet somit das Bündel, um  $f$  durch ein stückweise lineares Schnittebenenmodell zu approximieren. Die Schrittberechnung erfolgt dann durch Minimierung dieses Modells, versehen mit einer geeigneten Penalisierung, welche zu langen Schritten entgegenwirkt und das Teilproblem eindeutig lösbar macht.

In der dualen Sichtweise wird das Bündel verwendet, um durch geeignete Konvexkombinationen der Subgradienten eine innere Approximation  $G_\varepsilon^k$  des  $\varepsilon$ -Subdifferentials  $\partial_{\varepsilon_k} f(x^k)$  zu erzeugen und dieses mittels  $s^k = -P_{G_\varepsilon^k}(0)$  im Sinne des Modellalgorithmus 2.7.8 zur Berechnung von Suchrichtungen zu verwenden.

### 2.8.1 Das Bundle-Verfahren aus Sicht der Schnittebenenmethode

In der  $k$ -ten Iteration stehen für ein Schnittebenenmodell aus früheren Iterationen in dem Bündel die folgenden Informationen zur Verfügung:

$$y^j, \quad f_j = f(y^j), \quad g^j \in \partial f(y^j), \quad j \in J_k \subset \{0, \dots, k\}.$$



Der Punkt  $x^k$  ist stets im Bündel vertreten, d.h. es gibt  $j \in J_k$  mit  $y^j = x^k$ .

Das vom Bündel induzierte Schnittebenenmodell lautet nun

$$f_k^{\text{se}}(x) = \max_{j \in J_k} l_j(x) = \max_{j \in J_k} (f_j + g^{jT}(x - y^j)),$$

wobei  $l_j$  die lineare Stützfunktion zum Bündel-Eintrag  $(y^j, f_j, g^j)$  ist:

$$l_j(x) = f_j + g^{jT}(x - y^j), \quad f_j = f(y^j), \quad g^j \in \partial f(y^j).$$

Zur Berechnung eines Schrittes  $s^k$  wird nun das folgende Teilproblem gelöst:

$$(2.25) \quad \min_{s \in \mathbb{R}^n} f_k^{\text{se}}(x^k + s) + \frac{1}{2\gamma_k} \|s\|^2,$$

wobei  $\gamma_k > 0$  geeignet gewählt ist. Der Penalty-Term  $\frac{1}{2\gamma_k} \|s\|^2$  sorgt dafür, dass der Schritt  $s^k$  nicht zu weit von  $x^k$  wegführt. Je kleiner  $\gamma_k$ , desto kürzer wird der Schritt ausfallen. Man kann sogar folgendes zeigen:

Ist  $s^k$  Lösung von (2.25), so gibt es  $\Delta_k > 0$ , so dass  $s^k$  das folgende Trust-Region-Schnittebenen-Problem löst:

$$(2.26) \quad \min_s f_k^{\text{se}}(x^k + s) \quad \text{u.d.N.} \quad \|s\| \leq \Delta_k.$$

Ist umgekehrt  $s^k$  eine Lösung von (2.26), in der die Nebenbedingung  $\|s\| \leq \Delta_k$  stark aktiv ist (d.h. nach Vergrößern von  $\Delta_k$  wäre  $s^k$  nicht mehr optimal), so gibt es  $\gamma_k > 0$ , so dass  $s^k$  das Problem (2.25) löst.

Man kann  $\gamma_k$  geeignet anpassen (siehe [SZ92]), aber für ein konvergentes Verfahren ist das nicht nötig. Wir setzen ab jetzt  $\gamma_k = 1$  (wie z.B. auch in [GK02]).

Im folgenden ist es günstig,  $l_j$  folgendermaßen umzuschreiben:

$$(2.27) \quad \begin{aligned} l_j(x) &= f_j + g^{jT}(x - y^j) = f(x^k) + g^{jT}(x - x^k) - (f(x^k) - f_j + g^{jT}(y^j - x^k)) \\ &= f(x^k) + g^{jT}(x - x^k) - \alpha_j^k \end{aligned}$$

mit

$$(2.28) \quad \alpha_j^k = f(x^k) - f_j - g^{jT}(x^k - y^j) = f(x^k) - l_j(x^k).$$

Der Wert  $\alpha_j^k$  ist also die Differenz zwischen  $f(x^k)$  und  $l_j(x^k)$  und somit immer nichtnegativ. Damit lautet das Schnittebenenmodell

$$(2.29) \quad f_k^{\text{se}}(x) = \max_{j \in J_k} l_j(x) = f(x^k) + \max_{j \in J_k} (g^{jT}(x - x^k) - \alpha_j^k) =: f(x^k) + \bar{f}_k^{\text{se}}(x).$$

In dem Problem (2.25) lassen wir den konstanten Offset  $f(x^k)$  weg und erhalten

$$(2.30) \quad \min_s \bar{f}_k^{\text{se}}(x^k + s) + \frac{1}{2} \|s\|^2.$$

Hierbei haben wir wie angekündigt  $\gamma_k = 1$  gesetzt.

Gemäß Lemma 2.6.1 können wir dieses Problem als QP schreiben:

$$(2.31) \quad \min_{s \in \mathbb{R}^n, \xi \in \mathbb{R}} \xi + \frac{1}{2} \|s\|^2 \quad \text{u.d.N.} \quad g^{jT}s - \alpha_j^k - \xi \leq 0, \quad j \in J_k.$$

Das Paar  $(s^k, \xi_k)$  ist genau dann Lösung von (2.31), wenn  $s^k$  Lösung von (2.30) ist und  $\xi_k = \bar{f}_k^{se}(x^k + s^k)$  gilt.

Ist der Schritt  $s^k$  berechnet, so prüfen wir in gewohnter Trust-Region-Manier, ob die tatsächliche Zielfunktionsabnahme

$$(2.32) \quad \text{ared}_k(s^k) = f(x^k) - f(x^k + s^k)$$

hinreichend groß ist im Vergleich zur Modellabnahme (da es stets  $j \in J_k$  mit  $y^j = x^k$  gibt, gilt  $f_k^{se}(x^k) = f(x^k)$ ), der wie folgt definiert ist (predicted reduction):

$$(2.33) \quad \text{pred}_k(s^k) = f_k^{se}(x^k) - f_k^{se}(x^k + s^k) = f(x^k) - f_k^{se}(x^k + s^k) = -\bar{f}_k^{se}(x^k + s^k) = -\xi_k.$$

Hierzu verwenden wir die Bedingung

$$(2.34) \quad \text{ared}_k(s^k) \geq \eta \text{pred}_k(s^k) = -\eta \xi_k.$$

Ist diese erfüllt, so führen wir einen *wesentlichen Schritt* (serious step, dies ist die gängige Terminologie) aus:

$$x^{k+1} = x^k + s^k.$$

Sonst führen wir einen *Nullschritt* aus:

$$x^{k+1} = x^k.$$

In beiden Fällen wird der Punkt  $y^{k+1} = x^k + s^k$  in das neue Bündel aufgenommen.

Es ergibt sich das folgende Verfahren:

**Algorithmus 2.8.1** (Bundle-Verfahren).

0. Wähle  $x^0 \in \mathbb{R}^n$ ,  $\eta \in (0, 1)$  und  $\varepsilon \geq 0$ . Bestimme  $g^0 \in \partial f(x^0)$ , setze  $y^0 = x^0$ ,  $\alpha_0^0 = 0$  und  $J_0 = \{0\}$ .

Für  $k = 0, 1, 2, \dots$ :

1. Berechne ein KKT-Tupel  $(s^k, \xi_k, \lambda^k)$  des Problems (2.31).
2. Berechne  $v^k = -s^k$  und  $\varepsilon_k = \sum_{j \in J_k} \lambda_j^k \alpha_j^k$ .
3. Prüfe auf Abbruch: Falls  $\|v^k\| \leq \varepsilon$  und  $\varepsilon_k \leq \varepsilon$ , STOP.
4. Gilt

$$f(x^k + s^k) - f(x^k) \leq \eta \xi_k,$$

so führe einen *wesentlichen Schritt* durch:

$$y^{k+1} = x^k + s^k, \quad x^{k+1} = y^{k+1}, \quad J_{k+1} = \{j \in J_k; \lambda_j^k > 0\} \cup \{k+1\}.$$

5. Gilt

$$f(x^k + s^k) - f(x^k) > \eta \xi_k,$$

so führe einen *Nullschritt* durch:

$$y^{k+1} = x^k + s^k, \quad x^{k+1} = x^k, \quad J_{k+1} = \{j \in J_k; \lambda_j^k > 0 \text{ oder } y^j = x^k\} \cup \{k+1\}.$$

6. Berechne  $f_{k+1} = f(y^{k+1})$ ,  $g^{k+1} \in \partial f(y^{k+1})$  und

$$\alpha_j^{k+1} = f(x^{k+1}) - f_j - g^{jT}(x^{k+1} - y^j), \quad j \in J_{k+1}.$$

Einige Bemerkungen:

- Die Abbruchbedingung wird erst mit Lemma 2.8.5 klar werden. Sie sichert ein gewisse Form von  $\varepsilon$ -Optimalität zu.
- Die Indexmenge  $J_k$  wird so aktualisiert, dass neben den Schnittebenen zu  $y^{k+1} = x^k + s^k$  und zu  $x^k$  nur die im Punkt  $x^k + s^k$  stark aktiven Schnittebenen im Bündel bleiben. Andere Varianten sind möglich.

## 2.8.2 Eine duale Sichtweise des Bundle-Verfahrens

Wir leiten nun zu dem Teilproblem (2.31) ein duales Problem her.

**Lemma 2.8.2.** a)  $(s^k, \xi_k)$  ist genau dann Lösung von (2.31), wenn die KKT-Bedingungen gelten, d.h. wenn es  $\lambda_j^k \in \mathbb{R}$ ,  $j \in J_k$ , gibt mit

$$(2.35) \quad \begin{aligned} s^k + \sum_{j \in J_k} \lambda_j^k g^j &= 0, \\ \sum_{j \in J_k} \lambda_j^k &= 1, \\ g^{jT} s^k - \alpha_j^k - \xi_k &\leq 0, \quad \lambda_j^k \geq 0, \quad \lambda_j^k (g^{jT} s^k - \alpha_j^k - \xi_k) = 0, \quad j \in J_k. \end{aligned}$$

b) Der Vektor  $\lambda^k$  ist genau dann Lösung des Problems

$$(2.36) \quad \begin{aligned} \min_{\lambda} \quad & \frac{1}{2} \left\| \sum_{j \in J_k} \lambda_j g^j \right\|^2 + \sum_{j \in J_k} \lambda_j \alpha_j^k \\ \text{u.d.N.} \quad & \sum_{j \in J_k} \lambda_j = 1, \quad \lambda_j \geq 0, \quad j \in J_k, \end{aligned}$$

wenn die KKT-Bedingungen gelten. Weiter ist  $(\lambda^k, \mu^k, \xi_k) \in \mathbb{R}^{|J_k|} \times \mathbb{R}^{|J_k|} \times \mathbb{R}$  genau dann ein KKT-Tupel von (2.36), wenn (2.35) für

$$(2.37) \quad s^k = - \sum_{j \in J_k} \lambda_j^k g^j,$$

erfüllt ist und zusätzlich gilt:

$$(2.38) \quad \mu_j^k = -g^{jT} s^k + \alpha_j^k + \xi_k, \quad j \in J_k.$$

gilt.

c) Sei  $\lambda^k$  eine Lösung von (2.36). Dann ist  $\lambda^k$  auch Lösung des folgenden Problems:

$$\begin{aligned}
(2.39) \quad & \min_{\lambda} \quad \frac{1}{2} \left\| \sum_{j \in J_k} \lambda_j g^j \right\|^2 \\
& \text{u.d.N.} \quad \sum_{j \in J_k} \lambda_j = 1, \quad \lambda_j \geq 0, \quad j \in J_k, \quad \sum_{j \in J_k} \lambda_j \alpha_j^k \leq \varepsilon_k
\end{aligned}$$

mit

$$(2.40) \quad \varepsilon_k = \sum_{j \in J_k} \lambda_j^k \alpha_j^k.$$

**Beweis:** zu a):

Für das konvexe QP (2.31) sind die KKT-Bedingungen notwendig und hinreichend. Diese KKT-Bedingungen sind in a) angegeben.

zu b):

Die KKT-Bedingungen sind für das konvexe QP (2.36) notwendig und hinreichend und lauten:

$$\begin{aligned}
(2.41) \quad & g^j{}^T \sum_{i \in J_k} \lambda_i^k g^i + \alpha_j^k + \xi_k - \mu_j^k = 0, \quad j \in J_k \\
& \sum_{j \in J_k} \lambda_j^k = 1, \\
& \lambda_j^k \geq 0, \quad \mu_j^k \geq 0, \quad \mu_j^k \lambda_j^k = 0, \quad j \in J_k.
\end{aligned}$$

Gelte nun (2.35). Die erste Zeile in (2.35) liefert dann (2.37). Definieren wir nun  $\mu_j^k$  gemäß (2.38), und setzen (2.37) ein, so ergibt sich die erste Gleichung in (2.41). Umschreiben der dritten Gleichung von (2.35) auf  $\mu_j^k$  ergibt die dritte Gleichung in (2.41). Die zweite Gleichung in (2.35) und in (2.41) sind identisch.

Gelte nun umgekehrt (2.41). Definieren wir  $s^k$  gemäß (2.37), so folgt die erste Gleichung in (2.35). Einsetzen von (2.37) in die erste Gleichung von (2.41) ergibt (2.38) und Einsetzen von (2.38) in die dritte Zeile von (2.41) liefert die dritte Zeile von (2.35). Die zweiten Zeilen sind wiederum identisch.

zu c):

Für das konvexe QP in (2.39) sind die im folgenden aufgeschriebenen KKT-Bedingungen notwendig und hinreichend:

$$\begin{aligned}
(2.42) \quad & g^j{}^T \sum_{i \in J_k} \lambda_i^k g^i + \xi_k - \mu_j^k + \tau^k \alpha_j^k = 0, \quad j \in J_k \\
& \sum_{j \in J_k} \lambda_j^k = 1, \\
& \lambda_j^k \geq 0, \quad \mu_j^k \geq 0, \quad \mu_j^k \lambda_j^k = 0, \quad j \in J_k, \\
& \sum_{j \in J_k} \lambda_j^k \alpha_j^k \leq \varepsilon_k, \quad \tau^k \geq 0, \quad \tau^k \left( \sum_{j \in J_k} \lambda_j^k \alpha_j^k - \varepsilon_k \right) = 0.
\end{aligned}$$

Ist nun  $\lambda^k$  Lösung von (2.36), so gelten die Bedingungen (2.41). Definieren wir  $\varepsilon_k$  gemäß (2.40) und setzen wir  $\tau^k = 1$ , so folgt unmittelbar (2.42).  $\square$

Wie soeben gezeigt, sind also die Probleme (2.31) und (2.36) äquivalent. Aus einer Lösung von (2.36) läßt sich  $s^k$  aus (2.37) zurückgewinnen. Der Wert  $\xi_k$  kann ebenfalls direkt berechnet werden. Hierzu benutzen wir die Komplementaritätsbedingung in (2.35), (2.37) und (2.40):

$$\begin{aligned}\xi_k &= \xi_k \sum_{j \in J_k} \lambda_j^k = \sum_{j \in J_k} \lambda_j^k \xi_k = \sum_{j \in J_k} \lambda_j^k (g^{jT} s^k - \alpha_j^k) \\ &= \left( \sum_{j \in J_k} \lambda_j^k g^j \right)^T s^k - \sum_{j \in J_k} \lambda_j^k \alpha_j^k = -\|s^k\|^2 - \varepsilon_k.\end{aligned}$$

Damit ergeben sich folgende Formeln:

$$(2.43) \quad s^k = -v^k, \quad \xi_k = -\|s^k\|^2 - \varepsilon_k = -\|v^k\|^2 - \varepsilon_k$$

$$(2.44) \quad \text{mit } v^k = \sum_{j \in J_k} \lambda_j^k g^j, \quad \varepsilon_k = \sum_{j \in J_k} \lambda_j^k \alpha_j^k.$$

Wir kommen nun zu einem wichtigen Punkt: Gemäß Lemma 2.8.2 ist  $\lambda^k$  Lösung von (2.39) und daher gilt

$$v^k = P_{G_{\varepsilon_k}^k}(0),$$

wobei

$$(2.45) \quad G_{\varepsilon}^k = \left\{ \sum_{j \in J_k} \lambda_j g^j; \sum_{j \in J_k} \lambda_j \alpha_j^k \leq \varepsilon, \sum_{j \in J_k} \lambda_j = 1, \lambda_j \geq 0, j \in J_k \right\}$$

Ein wesentlicher Schritt  $s^k$  des Bundle-Verfahrens, d.h.

$$x^{k+1} - x^k = s^k = -v^k = -P_{G_{\varepsilon_k}^k}(0)$$

erinnert also stark an den Schritt  $x^{k+1} - x^k = -\sigma_k P_{\partial_{\varepsilon} f(x^k)}(0)$  im Modellalgorithmus 2.7.8. Dieser Eindruck wird nun erhärtet, indem wir zeigen, dass  $G_{\varepsilon}^k$  eine Teilmenge von  $\partial_{\varepsilon} f(x^k)$  ist.

**Lemma 2.8.3.** *Sei  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  konvex. Weiter seien  $x^k \in \mathbb{R}^n$ ,  $y^j \in \mathbb{R}^n$ ,  $g^j \in \partial f(y^j)$ ,  $j \in J_k$ ,  $\varepsilon \geq 0$  und  $G_{\varepsilon}^k$  definiert gemäß (2.45). Dann gilt*

$$G_{\varepsilon}^k \subset \partial_{\varepsilon} f(x^k).$$

**Beweis:**

Wir zeigen zunächst  $g^j \in \partial_{\alpha_j^k} f(x^k)$ :

Nach Definition von  $\alpha_j^k$  und wegen  $g^j \in \partial f(y^j)$  gilt für alle  $x \in \mathbb{R}^n$

$$\begin{aligned}f(x) - f(x^k) &= f(x) - f(y^j) + f(y^j) - f(x^k) \geq g^{jT}(x - y^j) + f(y^j) - f(x^k) \\ &= g^{jT}(x - x^k) + g^{jT}(x^k - y^j) + f(y^j) - f(x^k) = g^{jT}(x - x^k) - \alpha_j^k,\end{aligned}$$

also  $g^j \in \partial_{\alpha_j^k} f(x^k)$ .

Sei nun  $\lambda_j \geq 0$ ,  $\sum_{j \in J_k} \lambda_j = 1$ . Multiplizieren von

$$f(x) - f(x^k) \geq g^{jT}(x - x^k) - \alpha_j^k$$

mit  $\lambda_j$  und Aufaddieren liefert

$$f(x) - f(x^k) \geq \left( \sum_{j \in J_k} \lambda_j g^j \right)^T (x - x^k) - \sum_{j \in J_k} \lambda_j \alpha_j^k \quad \forall x \in \mathbb{R}^n.$$

Damit ist

$$\sum_{j \in J_k} \lambda_j g^j \in \partial_{(\sum_{j \in J_k} \lambda_j \alpha_j^k)} f(x^k) \stackrel{\sum_{j \in J_k} \lambda_j \alpha_j^k \leq \varepsilon}{\subset} \partial_\varepsilon f(x^k)$$

gezeigt. □

Wir können daher Algorithmus 2.8.1 auch so formulieren:

**Algorithmus 2.8.4** (Bundle-Verfahren 2.8.1 in dualer Formulierung).

0. Wähle  $x^0 \in \mathbb{R}^n$ ,  $\eta \in (0, 1)$  und  $\varepsilon \geq 0$ . Bestimme  $g^0 \in \partial f(x^0)$ , setze  $y^0 = x^0$ ,  $\alpha_0^0 = 0$  und  $J_0 = \{0\}$ .

Für  $k = 0, 1, 2, \dots$ :

1. Berechne  $\lambda^k$  durch Lösen des Problems (2.36).
2. Berechne

$$v^k = \sum_{j \in J_k} \lambda_j^k g^j, \quad s^k = -v^k, \quad \varepsilon_k = \sum_{j \in J_k} \lambda_j^k \alpha_j^k, \quad \xi_k = -\|v^k\|^2 - \varepsilon_k.$$

3. Prüfe auf Abbruch: Falls  $\|v^k\| \leq \varepsilon$  und  $\varepsilon_k \leq \varepsilon$ , STOP.
4. Gilt

$$f(x^k + s^k) - f(x^k) \leq \eta \xi_k,$$

so führe einen *wesentlichen Schritt* durch:

$$y^{k+1} = x^k + s^k, \quad x^{k+1} = y^{k+1}, \quad J_{k+1} = \{j \in J_k; \lambda_j^k > 0\} \cup \{k+1\}.$$

5. Gilt

$$f(x^k + s^k) - f(x^k) > \eta \xi_k,$$

so führe einen *Nullschritt* durch:

$$y^{k+1} = x^k + s^k, \quad x^{k+1} = x^k, \quad J_{k+1} = \{j \in J_k; \lambda_j^k > 0 \text{ oder } y^j = x^k\} \cup \{k+1\}.$$

6. Berechne  $f_{k+1} = f(y^{k+1})$ ,  $g^{k+1} \in \partial f(y^{k+1})$  und

$$\alpha_j^{k+1} = f(x^{k+1}) - f_j - g^{jT}(x^{k+1} - y^j), \quad j \in J_{k+1}.$$

Wir sehen uns nun die Abbruchbedingung näher an.

**Lemma 2.8.5.** In Algorithmus 2.8.1 (bzw. Algorithmus 2.8.4) sei für  $\varepsilon > 0$  die Abbruchbedingung

$$\|v^k\| \leq \varepsilon, \quad \varepsilon_k \leq \varepsilon.$$

erfüllt. Dann ist  $x_k$   $\varepsilon$ -optimal im folgenden Sinne:

$$f(x^k) \leq f(x) + \varepsilon \|x - x^k\| + \varepsilon \quad \forall x \in \mathbb{R}^n.$$

**Beweis:** Für alle  $j \in J_k$  und  $x \in \mathbb{R}^n$  gilt nach (2.27)

$$g^j{}^T(x - x^k) = l_j(x) - f(x^k) + \alpha_j^k \leq f(x) - f(x^k) + \alpha_j^k.$$

Multiplizieren mit  $\lambda_j^k$  und Summieren ergibt:

$$\begin{aligned} v^k{}^T(x - x^k) &= \sum_{j \in J_k} \lambda_j^k g^j{}^T(x - x^k) \leq \sum_{j \in J_k} \lambda_j^k (f(x) - f(x^k) + \alpha_j^k) \\ &= f(x) - f(x^k) + \sum_{j \in J_k} \lambda_j^k \alpha_j^k = f(x) - f(x^k) + \varepsilon_k. \end{aligned}$$

Dies ergibt mit der Cauchy-Schwarz-Ungleichung

$$f(x_k) \leq f(x) - v^k{}^T(x - x^k) + \varepsilon_k \leq f(x) + \|v^k\| \|x - x^k\| + \varepsilon_k \leq f(x) + \varepsilon \|x - x_k\| + \varepsilon.$$

□

### 2.8.3 Globale Konvergenz

Wir weisen nun die globale Konvergenz des Verfahrens nach. Dies ist relativ aufwendig.

Die Iterationen, in denen wesentliche Schritte erfolgen, werden in der Menge  $\mathcal{K}$  zusammengefasst:

$$\mathcal{K} = \{k \geq 0; t_k = 1\},$$

wobei  $t_k = 1$ , falls  $x^k \rightarrow x^{k+1}$  ein wesentlicher Schritt ist und  $t_k = 0$ , sonst.

Wir beginnen mit einem technischen Resultat:

**Lemma 2.8.6.** *Sei  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  konvex. Sei  $\varepsilon = 0$  und die Folgen  $(x^k)$ ,  $(s^k)$  usw. seien von Algorithmus 2.8.1 erzeugt (insbesondere terminiere das Verfahren also nicht endlich). Weiter sei die Folge  $(f(x^k))$  durch  $f^* \in \mathbb{R}$  nach unten beschränkt. Dann gilt:*

- a)  $\lim_{k \rightarrow \infty} f(x^{k+1}) - f(x^k) = 0.$
- b)  $\sum_{k=0}^{\infty} t_k (\|v^k\|^2 + \varepsilon_k) \leq \frac{f(x^0) - f^*}{\eta}.$
- c) *Erzeugt der Algorithmus unendlich viele wesentliche Schritte, d.h. gilt  $|\mathcal{K}| = \infty$ , so folgt*

$$\lim_{\mathcal{K} \ni k \rightarrow \infty} \|v^k\| = 0, \quad \lim_{\mathcal{K} \ni k \rightarrow \infty} \varepsilon_k = 0.$$

**Beweis:**

zu a):

Die Folge  $(f(x^k))$  ist wegen der Bedingung für wesentliche Schritte in Schritt 4 monoton fallend und durch  $f^* \in \mathbb{R}$  nach unten beschränkt. Daher folgt a) unmittelbar.

zu b):

Für alle wesentlichen Schritte gilt  $t_k = 1$  und

$$f(x^{k+1}) - f(x^k) \leq \eta t_k \xi_k$$

Für alle Nullschritte ist dies wegen  $x^{k+1} = x^k$  und  $t_k = 0$  ebenfalls erfüllt. Damit haben wir

$$\begin{aligned} f(x^0) - f^* &\geq f(x^0) - \lim_{k \rightarrow \infty} f(x^k) = \sum_{k=0}^{\infty} (f(x^k) - f(x^{k+1})) \geq -\eta \sum_{k=0}^{\infty} t_k \xi_k \\ &= \eta \sum_{k=0}^{\infty} t_k (\|v^k\|^2 + \varepsilon_k) \geq \eta \sum_{k=0}^{\infty} t_k (\gamma^- \|v^k\|^2 + \varepsilon_k), \end{aligned}$$

wobei wir (2.43) benutzt haben.

zu c):

Nach b) gilt

$$\sum_{k \in \mathcal{K}} (\gamma^- \|v^k\|^2 + \varepsilon_k) = \sum_{k=0}^{\infty} t_k (\gamma^- \|v^k\|^2 + \varepsilon_k) \leq \frac{f(x^0) - f^*}{\eta} < \infty.$$

daher ist  $(\gamma^- \|v^k\|^2 + \varepsilon_k)_{\mathcal{K}}$  eine Nullfolge und somit (wegen  $\gamma^- > 0$  und  $\varepsilon_k \geq 0$ ) auch  $(\|v^k\|)_{\mathcal{K}}$  und  $(\varepsilon_k)_{\mathcal{K}}$ . □

Wir zeigen nun ein erstes Konvergenzresultat für den Fall, dass unendlich viele wesentliche Schritte durchgeführt werden:

**Lemma 2.8.7.** *Algorithmus 2.8.1 mit  $\varepsilon = 0$  erzeuge unendlich viele wesentliche Schritte. Dann ist jeder Häufungspunkt von  $(x^k)$  ein (globales) Minimum von  $f$ .*

**Beweis:**

Sei  $x^*$  ein Häufungspunkt von  $(x^k)$ . Die Folge  $(f(x^k))$  ist monoton fallend und hat, da  $f$  stetig ist,  $f(x^*)$  als Häufungspunkt. Daraus folgt  $f(x^k) \downarrow f(x^*) =: f^* \in \mathbb{R}$ . Wegen  $|\mathcal{K}| = \infty$  gilt

$$\{x^k; k \geq 0\} = \{x^0\} \cup \{x^{k+1}; k \in \mathcal{K}\} = \{x^k; k \in \mathcal{K}\}$$

und daher ist dann  $x^*$  auch ein Häufungspunkt von  $(x^k)_{\mathcal{K}}$ . Es gibt somit eine Teilfolge  $(x^k)_{\mathcal{K}'}$ ,  $\mathcal{K}' \subset \mathcal{K}$ , mit

$$(x^k)_{\mathcal{K}'} \rightarrow x^*.$$

Gemäß Lemma 2.8.3 gilt  $v^k \in \partial_{\varepsilon_k} f(x^k)$  und aus Lemma 2.8.6 c) folgt

$$(\|v^k\|)_{\mathcal{K}} \rightarrow 0, \quad (\varepsilon_k)_{\mathcal{K}} \rightarrow 0.$$

Weiter haben wir für alle  $x \in \mathbb{R}^n$

$$f(x) \geq f(x^k) + v^{kT} (x - x^k) - \varepsilon_k \xrightarrow{\mathcal{K}' \ni k \rightarrow \infty} f(x^*) + 0^T (x - x^*) + 0 = f(x^*).$$

Damit ist  $x^*$  globales Minimum von  $f$ . □

Wir untersuchen nun den Fall, dass nur endlich viele wesentliche Schritte auftreten. Hierzu treffen wir, um uns das Leben etwas zu erleichtern, die folgende Annahme:

$$(2.46) \quad \begin{aligned} &\exists m > 0 : \forall k \geq m : \\ &\{k - m, k - m + 1, \dots, k - 1\} \cap \mathcal{K} = \emptyset \implies \gamma_k = \gamma^-. \end{aligned}$$

In Worten: Nach einer Serie von  $m$  Nullschritten gilt stets  $\gamma_k = \gamma^-$ .



**Lemma 2.8.8.** *Algorithmus 2.8.1 mit  $\varepsilon = 0$  erzeuge eine unendliche Folge  $(x^k)$  und es gelte (2.46). Werden nur endlich viele wesentliche Schritte durchgeführt, d.h. gibt es  $l \geq 0$  mit  $x^k = x^l$  für alle  $k \geq l$ , so ist  $x^l$  globales Minimum von  $f$ .*

**Beweis:** Sei

$$J_k^+ = \left\{ j \in J_k ; \lambda_j^k > 0 \right\}.$$

Wegen  $x^k = x^l$  für alle  $k \geq l$  gilt

$$\alpha_j^{k+1} = \alpha_j^k \quad \forall j \in J_k^+, \quad k \geq l.$$

Ohne Einschränkung können wir wegen (2.46) annehmen, dass gilt:

$$\gamma_k = \gamma^- \quad \forall k \geq l.$$

Wir setzen nun

$$(2.47) \quad \theta_k := \sum_{j \in J_k^+} \lambda_j^k \alpha_j^{k+1} = \sum_{j \in J_k^+} \lambda_j^k \alpha_j^k = \varepsilon_k \quad \forall k \geq l.$$

Bezeichne weiter

$$Q_k(\lambda) = \frac{\gamma_k}{2} \left\| \sum_{j \in J_k} \lambda_j g^j \right\|^2 + \sum_{j \in J_k} \lambda_j \alpha_j^k$$

die Zielfunktion in (2.36).

Im folgenden sei  $k > l$  beliebig.

Wir wählen nun zu beliebigem  $\mu \in [0, 1]$  den Vektor  $\lambda^{k\mu}$  in folgender Weise:

$$\lambda_k^{k\mu} = \mu, \quad \lambda_j^{k\mu} = (1 - \mu) \lambda_j^{k-1}, \quad j \in J_{k-1}^+, \quad \lambda_j^{k\mu} = 0, \quad j \in J_k \setminus (J_{k-1}^+ \cup \{k\}).$$

Hierbei sei  $\lambda^{k-1}$  der durch Lösen des  $(k-1)$ -ten Teilproblems erhaltenene Vektor. Offensichtlich gilt dann

$$\lambda^{k\mu} \geq 0, \quad \sum_{j \in J_k} \lambda_j^{k\mu} = 1.$$

Somit ist  $\lambda^{k\mu}$  zulässig für (2.36) und daraus folgt

$$Q_k(\lambda^k) \leq Q_k(\lambda^{k\mu}).$$

Zur Vereinfachung von  $Q_k(\lambda^{k\mu})$  berechnen wir:

$$\sum_{j \in J_k} \lambda_j^{k\mu} g^j = \mu g^k + (1 - \mu) \sum_{j \in J_{k-1}^+} \lambda_j^{k-1} g^j = \mu g^k + (1 - \mu) \sum_{j \in J_{k-1}} \lambda_j^{k-1} g^j = \mu g^k + (1 - \mu) v^{k-1}.$$

Ebenso ergibt sich

$$\sum_{j \in J_k} \lambda_j^{k\mu} \alpha_j^k = \mu \alpha_k^k + (1 - \mu) \sum_{j \in J_{k-1}^+} \lambda_j^{k-1} \alpha_j^k = \mu \alpha_k^k + (1 - \mu) \theta_{k-1}$$

mit  $\theta_{k-1}$  wie in (2.47) definiert.

Damit erhalten wir

$$Q_k(\lambda^{k\mu}) = \frac{\gamma^-}{2} \|\mu g^k + (1 - \mu)v^{k-1}\|^2 + \mu\alpha_k^k + (1 - \mu)\theta_{k-1} = q_k(\mu)$$

mit

$$q_k(\mu) = \frac{\gamma^-}{2} \|\mu g^k + (1 - \mu)v^{k-1}\|^2 + \mu\alpha_k^k + (1 - \mu)\theta_{k-1}.$$

Bezeichne nun  $\mu^k$  das Minimum von  $q_k$  auf  $[0, 1]$ . Weiter sei  $\nu_k = q_k(\mu_k)$ . Dann ergibt sich

$$\begin{aligned} (2.48) \quad \nu_k &= q_k(\mu_k) \leq q_k(0) = \frac{\gamma^-}{2} \|v^{k-1}\|^2 + \theta_{k-1} = \frac{\gamma^-}{2} \|v^{k-1}\|^2 + \varepsilon_{k-1} \\ &= Q_{k-1}(\lambda^{k-1}) \leq Q_{k-1}(\lambda^{k-1, \mu_{k-1}}) = q_{k-1}(\mu_{k-1}) = \nu_{k-1}. \end{aligned}$$

Da  $q_k$  quadratisch ist, ergibt sich

$$q_k(\mu) = q_k(0) + \mu q'_k(0) + \frac{\mu^2}{2} q''_k(0) \leq \nu_{k-1} + \mu q'_k(0) + \frac{\mu^2}{2} q''_k(0).$$

Wir schätzen nun  $q'_k(0)$  ab:

$$q'_k(0) = \gamma^-(g^k - v^{k-1})^T v^{k-1} + \alpha_k^k - \theta_{k-1} = -\gamma^- \|v^{k-1}\|^2 + \gamma^- g^{kT} v^{k-1} + \alpha_k^k - \theta_{k-1}.$$

Wir benutzen  $x^k = x^{k-1}$ ,  $y^k = x^{k-1} + s^k$ ,  $\text{ared}_{k-1}(s^{k-1}) = f(x^{k-1}) - f^k < -\eta \xi_{k-1}$  (sonst wäre  $k-1 \in \mathcal{K}$ ) sowie (2.43) und erhalten

$$\begin{aligned} \alpha_k^k &= f(x^k) - f_k - g^{kT}(x^k - y^k) = f(x^k) - f_k - g^{kT}(x^{k-1} - y^k) \\ &= f(x^{k-1}) - f_k + g^{kT} s^{k-1} < -\eta \xi_{k-1} + g^{kT} s^{k-1} = \eta(\varepsilon_{k-1} + \gamma^- \|v^{k-1}\|^2) + g^{kT} s^{k-1} \\ &= \eta(\theta_{k-1} + \gamma^- \|v^{k-1}\|^2) - \gamma^- g^{kT} v^{k-1}. \end{aligned}$$

Daraus folgt

$$q'_k(0) = -\gamma^- \|v^{k-1}\|^2 + \gamma^- g^{kT} v^{k-1} + \alpha_k^k - \theta_{k-1} \leq -(1 - \eta)(\theta_{k-1} + \gamma^- \|v^{k-1}\|^2).$$

Als nächstes schätzen wir  $q''_k(0)$  ab. Zunächst gilt

$$\gamma^- \|v^{k-1}\|^2 \leq \gamma^- \|v^{k-1}\|^2 + 2\theta_{k-1} \leq 2\nu_{k-1} \leq 2\nu_l.$$

Daraus wiederum folgt mit (2.48)

$$\begin{aligned} \|y^k\| &= \|x^{k-1} + s^{k-1}\| \leq \|x^{k-1}\| + \|s^{k-1}\| = \|x^{k-1}\| + \gamma_{k-1} \|v^{k-1}\| \\ &= \|x^l\| + \gamma^- \|v^{k-1}\| \leq \|x^l\| + \sqrt{2\gamma^- \nu_l}. \end{aligned}$$

Somit ist die Folge  $(y^k)_{k>l}$  beschränkt und daher auch die Folge  $(g^k)_{k>l}$ , siehe Satz 2.4.3 und Lemma 2.5.7. Insgesamt gibt es daher eine Konstante  $C > 0$  mit

$$\|v^{k-1}\| \leq C, \quad \|g^k\| \leq C \quad \forall k > l.$$

Nun folgt

$$q''_k(0) = \gamma^- \|g^k - v^{k-1}\|^2 \leq \gamma^- (\|g^k\| + \|v^{k-1}\|)^2 \leq 4\gamma^- C^2 \quad \forall k > l,$$

Für alle  $\mu \geq 0$  ergibt sich

$$q_k(\mu) \leq \nu_{k-1} + \mu q'_k(0) + \frac{\mu^2}{2} q''_k(0) \leq \nu_{k-1} + \mu q'_k(0) + 2\mu^2 \gamma^- C^2 =: \bar{q}_k(\mu).$$

Bezeichne  $\bar{\mu}_k$  das unrestringierte globale Minimum von  $\bar{q}_k$ . Dann gilt

$$\bar{\mu}_k = \frac{-q'_k(0)}{4\gamma^- C^2}$$

Im Fall  $\bar{\mu}_k > 1$  erhalten wir

$$\nu_k = q_k(\mu_k) \leq q_k(1) \leq \bar{q}_k(1) = \nu_{k-1} + q'_k(0) + 2\gamma^- C^2 < \nu_{k-1} + q'_k(0) - \frac{q'_k(0)}{2} = \nu_{k-1} + \frac{q'_k(0)}{2}.$$

Im Fall  $\bar{\mu}_k \leq 1$  ergibt sich

$$\nu_k = q_k(\mu_k) \leq q_k(\bar{\mu}_k) \leq \bar{q}_k(\bar{\mu}_k) = \nu_{k-1} + \bar{\mu}_k q'_k(0) + 2\bar{\mu}_k^2 \gamma^- C^2 = \nu_{k-1} - \frac{q'_k(0)^2}{8\gamma^- C^2}.$$

Die nichtnegative, monoton fallende Folge  $(\nu_k)$  ist konvergent und daher eine Cauchy-Folge. Insbesondere ist  $(\nu_{k-1} - \nu_k)$  eine Nullfolge. Wegen

$$\nu_{k-1} - \nu_k \geq \min \left\{ \frac{-q'_k(0)}{2}, \frac{q'_k(0)^2}{8\gamma^- C^2} \right\} > 0$$

folgt, dass  $(q'_k(0))$  eine Nullfolge ist, und daraus wiederum

$$\lim_{k \rightarrow \infty} (\theta_{k-1} + \gamma^- \|v^{k-1}\|^2) = 0.$$

Daher haben wir wegen  $\varepsilon_k = \theta_k$ ,  $k \geq l$ :

$$\varepsilon_k \rightarrow 0, \quad v^k \rightarrow 0, \quad k \rightarrow \infty.$$

Wir können nun fortfahren wie im zweiten Teil des Beweises von Lemma 2.8.7, um zu zeigen, dass jeder Häufungspunkt von  $(x^k)$  ein globales Minimum von  $f$  ist. Die stationäre Folge besitzt genau einen Grenzwert, nämlich  $x^l$ .  $\square$

Nehmen wir Lemma 2.8.7 und Lemma 2.8.8 zusammen, dann erhalten wir den folgenden Konvergenzsatz:

**Satz 2.8.9.** *Algorithmus 2.8.1 mit  $\varepsilon = 0$  erzeuge die Folge  $(x^k)$  und es gelte (2.46). Dann ist jeder Häufungspunkt von  $(x^k)$  ein globales Minimum von  $f$ .*

Durch eine kleine Modifikation des Beweises von Lemma 6.79 und Satz 6.80 in [GK02] kann man außerdem noch zeigen:

**Satz 2.8.10.** *Die Funktion  $f$  besitze globale Minima und Algorithmus 2.8.1 mit  $\varepsilon = 0$  erzeuge die Folge  $(x^k)$  und es gelte (2.46). Dann konvergiert  $(x^k)$  gegen ein globales Minimum  $x^*$  von  $f$ .*



## Kapitel 3

# Verfahren für Nichtglatte Gleichungssysteme

Ziel dieses Kapitels ist die Entwicklung einer Klasse von Newton-Verfahren zur Lösung des folgenden Problems:

**Nichtglattes Gleichungssystem:**

(P)  $F(x) = 0$ ,  $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$  lokal Lipschitz-stetig, nicht überall differenzierbar.

Wir werden sehen, dass wir uns hierbei auf eine geeignete Unterklasse der lokal Lipschitz-stetigen Funktionen einschränken müssen.

In der Einführung wurde bereits gezeigt, dass Komplementaritätsprobleme mit Hilfe von NCP-Funktionen in äquivalente nichtglatte Gleichungssysteme überführt werden können. Wir hatten auch gezeigt, dass sich KKT-Systeme in nichtglatte Gleichungssysteme überführen lassen.

### 3.1 Ein allgemeines Newton-artiges Verfahren

Wir werden nun minimale Voraussetzungen erarbeiten, unter denen ein Newton-artiges Verfahren lokal q-superlinear bzw. q-quadratisch konvergiert.

Wir suchen eine Lösung des Gleichungssystems

$$(3.1) \quad F(x) = 0,$$

wobei  $F : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$  definiert ist auf der offenen, nichtleeren Menge  $U$ . Hierzu betrachten wir das folgende Verfahren:

**Algorithmus 3.1.1** (Newton-artiges Verfahren).

0. Wähle  $x^0 \in \mathbb{R}^n$ .

Für  $k = 0, 1, 2, \dots$ :

1. Falls  $F(x^k) = 0$ : STOP mit Ergebnis  $x^k$ .
2. Sonst wähle eine (geeignete) Matrix  $M_k \in \mathbb{R}^{n \times n}$ , löse

$$M_k s^k = -F(x^k)$$

und setze  $x^{k+1} = x^k + s^k$ .

**Bemerkung.** Das Verfahren ist durchführbar, solange  $x^k$  in  $U$  liegt und  $M_k$  invertierbar ist.

**Bemerkung.** Beim gewöhnlichen Newton-Verfahren wählt man  $M_k = F'(x^k)$  (Jacobi-Matrix von  $F$  bei  $x^k$ ).

Wir wollen uns nun eine notwendige und hinreichende Bedingung für  $q$ -superlineare bzw.  $q$ -quadratische Konvergenz des Verfahrens überlegen.

Sei hierzu  $x^* \in U$  eine Lösung von (3.1) und die Folgen  $(M_k)$ ,  $(s^k)$  und  $(x^k)$  seien von Algorithmus 3.1.1 erzeugt. Wir erhalten dann

$$(3.2) \quad \begin{aligned} x^* - x^{k+1} &= x^* - x^k - s^k = x^* - x^k + M_k^{-1} F(x^k) \\ &= M_k^{-1} (F(x^k) - F(x^*) - M_k(x^k - x^*)), \end{aligned}$$

wobei wir  $F(x^*) = 0$  verwendet haben. Daraus folgt:

**Satz 3.1.2.** Sei  $x^*$  eine Lösung von (3.1). Algorithmus 3.1.1 erzeuge die Folgen  $(M_k)$ ,  $(s^k)$  und  $(x^k)$ ,  $x^k \rightarrow x^*$ . Dann gilt:

(a) Die Konvergenzrate ist  $q$ -superlinear genau dann, wenn gilt

$$(3.3) \quad \|M_k^{-1} (F(x^k) - F(x^*) - M_k(x^k - x^*))\| = o(\|x^k - x^*\|) \quad \text{für } k \rightarrow \infty.$$

(b) Die Konvergenzrate ist  $q$ -quadratisch genau dann, wenn gilt

$$(3.4) \quad \|M_k^{-1} (F(x^k) - F(x^*) - M_k(x^k - x^*))\| = O(\|x^k - x^*\|^2) \quad \text{für } k \rightarrow \infty.$$

**Beweis:** Gemäß (3.2) ist die linke Seite in (3.3) und (3.4) gleich  $\|x^{k+1} - x^*\|$ , woraus direkt die Aussagen (a) und (b) folgen.  $\square$

In Algorithmus 3.1.1 ist die Wahl von  $M_k$  nicht näher spezifiziert. Wir wollen uns nun am gewöhnlichen Newton-Verfahren orientieren: Hier gilt  $M_k = F'(x^k)$ , d.h. die Wahl von  $M_k$  ist *punktbasiert*. Da wir auf nicht überall differenzierbare Funktionen  $F$  abzielen, betrachten wir die folgende allgemeinere punktbasierte Wahl von  $M_k$ :

**Voraussetzung 3.1.3.** In Algorithmus 3.1.1 werden die Matrizen  $M_k$  gewählt gemäß

$$M_k \in M_F(x^k) \quad \forall k.$$

Hierbei sei

$$M_F : U \subset \mathbb{R}^n \rightrightarrows \mathbb{R}^{n \times n}$$

eine *mengenwertige Abbildung* mit nichtleeren Bildern, d.h.

$$M_F(x) \subset \mathbb{R}^{n \times n}, \quad M_F(x) \neq \emptyset \quad \forall x \in U.$$

**Bemerkung.** Später werden wir für  $M_F$  verallgemeinerte Differentiale von  $F$  verwenden. Dies sind mengenwertige Abbildungen, was unsere Wahl eines mengenwertigen  $M_F$  begründet.

**Satz 3.1.4.** Sei  $x^* \in U$  eine Lösung von (3.1) und es gelte Voraussetzung 3.1.3. Weiter gebe es  $\eta > 0$ , so dass gilt:

$$(3.5) \quad M \text{ invertierbar} \quad \forall M \in M_F(x) \quad \forall x \in B_\eta(x^*).$$

a) *Gilt*

$$(3.6) \quad \sup_{M \in M_F(x^* + s)} \|M^{-1}(F(x^* + s) - F(x^*) - Ms)\| = o(\|s\|) \quad \text{für } s \rightarrow 0,$$

dann gibt es  $\delta > 0$ , so dass für alle  $x^0 \in B_\delta(x^*)$  der Algorithmus entweder mit  $x^k = x^*$  terminiert oder eine Folge  $(x^k)$  erzeugt, die  $q$ -superlinear gegen  $x^*$  konvergiert.

b) *Gilt in (a) die stärkere Bedingung*

$$(3.7) \quad \sup_{M \in M_F(x^* + s)} \|M^{-1}(F(x^* + s) - F(x^*) - Ms)\| = O(\|s\|^2) \quad \text{für } s \rightarrow 0,$$

so ist die Konvergenzrate  $q$ -quadratisch.

**Beweis:**

zu a):

Gemäß (3.6) gibt es  $0 < \delta \leq \eta$  mit

$$(3.8) \quad \sup_{M \in M_F(x)} \|M^{-1}(F(x) - F(x^*) - M(x - x^*))\| \leq \frac{1}{2} \|x - x^*\|.$$

für alle  $x \in B_\delta(x^*)$  (im Fall  $x = x^*$  ist dies trivial).

Wegen (3.5) ist der Algorithmus wohldefiniert, solange  $\|x^k - x^*\| < \eta$  gilt. Für  $x^k \in B_\delta(x^*)$  mit  $F(x^k) \neq 0$  haben wir  $x^k \neq x^*$  und erhalten aus (3.2), (3.6) sowie (3.8)

$$(3.9) \quad \begin{aligned} \|x^{k+1} - x^*\| &= \|M_k^{-1}(F(x^k) - F(x^*) - M_k(x^k - x^*))\| \\ &\leq \sup_{M \in M_F(x^k)} \|M^{-1}(F(x^k) - F(x^*) - M(x^k - x^*))\| \\ &\begin{cases} = o(\|x^k - x^*\|) & (\text{für } x^k \rightarrow x^*), \\ \leq \frac{1}{2} \|x^k - x^*\|. \end{cases} \end{aligned}$$

Für beliebiges  $x^0 \in B_\delta(x^*)$  ergibt sich dann induktiv für alle Iterierten  $x^k$ :

$$(3.10) \quad x^k \in B_\delta(x^*), \quad \|x^k - x^*\| \leq 2^{-k} \|x^0 - x^*\|.$$

Terminiert der Algorithmus endlich, so haben wir  $F(x^k) = 0$  und daher wie oben gezeigt  $x^k = x^*$ .

Erzeugt der Algorithmus unendlich viele Iterierte, so folgt aus (3.10), dass  $x^k \rightarrow x^*$  und (3.9) liefert  $q$ -superlineare Konvergenzrate.

zu b):

Aus (3.7) ergibt sich unmittelbar die Ordnung  $O(\|x^k - x^*\|^2)$  für  $x^k \rightarrow x^*$  in (3.9). Daher ist die Konvergenzrate  $q$ -quadratisch.  $\square$

Häufig ist es bequem, die Bedingung (3.6) (und entsprechend (3.7)) in zwei Bedingungen aufzuspalten:

**Regularitätsbedingung:**

$$(3.11) \quad \exists \eta > 0, C > 0 : \quad M \text{ nichtsingulär, } \|M^{-1}\| \leq C \quad \forall M \in M_F(x), \quad \forall x \in B_\eta(x^*).$$

**Approximationsbedingung:**

$$(3.12) \quad \sup_{M \in M_F(x^*+s)} \|F(x^*+s) - F(x^*) - Ms\| = o(\|s\|) \quad \text{für } s \rightarrow 0.$$

Für q-quadratische Konvergenz benötigen wir die stärkere

**Quadratische Approximationsbedingung:**

$$(3.13) \quad \sup_{M \in M_F(x^*+s)} \|F(x^*+s) - F(x^*) - Ms\| = O(\|s\|^2) \quad \text{für } s \rightarrow 0.$$

**Definition 3.1.5.** Erfüllt  $M_F$  die Approximationsbedingung (3.12), so nennen wir  $M_F$  *punkt-basierte Approximation* von  $F$  bei  $x^*$ . Gilt sogar (3.13), dann nennen wir  $M_F$  *punktbasierte Approximation der Ordnung 1* von  $F$  bei  $x^*$ .

**Lemma 3.1.6.** Die Regularitätsbedingung (3.11) und die Approximationsbedingung (3.12) zusammen implizieren die Bedingungen (3.5) und (3.6). Ebenso folgen (3.5) und (3.7) aus (3.11) und (3.13).

**Beweis:** (3.11) ist stärker als (3.5). Weiter haben wir für alle  $s \in B_\eta(0)$  und alle  $M \in M_F(x^*+s)$ :

$$\begin{aligned} \|M^{-1}(F(x^*+s) - F(x^*) - Ms)\| &\leq \|M^{-1}\| \|F(x^*+s) - F(x^*) - Ms\| \\ &\leq C \|F(x^*+s) - F(x^*) - Ms\|, \end{aligned}$$

so dass (3.6) nun aus (3.12) und (3.7) aus (3.13) folgt.  $\square$

Als unmittelbare Konsequenz erhalten wir:

**Korollar 3.1.7.** Die Aussagen von Satz 3.1.4 gelten auch, wenn wir die Bedingungen (3.5), (3.6) und (3.7) durch die Bedingungen (3.11), (3.12) und (3.13) ersetzen.

**3.1.1 Spezialfall: Das gewöhnliche Newton-Verfahren**

Sei nun  $F$  stetig differenzierbar. Mit der Wahl  $M_F(x) = \{F'(x)\}$  liefert dann Algorithmus 7 das klassische Newton-Verfahren. Nach Definition der Ableitung ist

$$F(x^*+s) - F(x^*) - F'(x^*+s)s = o(\|s\|) + (F'(x^*) - F'(x^*+s))s = o(\|s\|),$$

die Approximationsbedingung (3.12) ist also erfüllt. Ist  $F'$  lokal Lipschitz-stetig nahe  $x^*$ , dann gilt auch die quadratische Approximationsbedingung (3.13). Ist also zudem die Regularitätsbedingung (3.11) erfüllt dann folgt die schnelle lokale Konvergenz aus Korollar 3.1.7.

Wir wollen nun verallgemeinerte Differentiale angeben, die sich als punktbasierte Approximationen  $M_F$  eignen.

**3.2 Verallgemeinerte Differentiale**

Wir führen nun einige Begriffe der nichtglatten Analysis ein und betrachten durchgehend eine auf der offenen Menge  $U \neq \emptyset$  definierte lokal Lipschitz-stetige Funktion  $F : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ .

Für lokal Lipschitz-stetige Funktionen gilt der grundlegende (und schwer zu beweisende)



**Satz 3.2.1** (Rademacher). Sei  $F : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$  lokal Lipschitz-stetig auf der offenen, nicht-leeren Menge  $U$ . Dann ist  $F$  fast überall auf  $U$  differenzierbar, d.h. das Komplement der Menge  $U_d \subset U$  aller Punkte, in denen  $F'$  existiert, ist eine Lebesgue-Nullmenge:  $\mu(U \setminus U_d) = 0$ .

Die folgenden verallgemeinerten Differentiale sind von fundamentaler Bedeutung:

**Definition 3.2.2.** Sei  $F : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$  lokal Lipschitz-stetig und bezeichne  $U_d \subset U$  die Menge aller Differenzierbarkeitspunkte von  $F$ . Das *Bouligand-* (oder *B-*) *Subdifferential*  $\partial_B F(x) \subset \mathbb{R}^{m \times n}$  von  $F$  in  $x \in U$  ist die Menge

$$\partial_B F(x) \stackrel{\text{def}}{=} \{M \in \mathbb{R}^{m \times n}; \exists (x_k) \subset U_d : x_k \rightarrow x, F'(x_k) \rightarrow M (k \rightarrow \infty)\}.$$

Clarke's verallgemeinerte Ableitung  $\partial^{cl} F(x) \subset \mathbb{R}^{m \times n}$  von  $F$  in  $x \in U$  ist die konvexe Hülle von  $\partial_B F(x)$ :

$$\partial^{cl} F(x) \stackrel{\text{def}}{=} \text{conv } \partial_B F(x).$$

**Beispiel 3.2.3.** Wir betrachten  $F : \mathbb{R} \rightarrow \mathbb{R}$ ,  $F(x) = |x|$ .

$F$  ist offensichtlich Lipschitz-stetig mit  $L = 1$ . Es gilt  $U_d = \mathbb{R} \setminus \{0\}$  und  $F'$  ist stetig auf  $U_d$ . Somit haben wir für alle  $x \neq 0$ :

$$x_k \rightarrow x \implies F'(x_k) \rightarrow F'(x) \implies \partial_B F(x) = \{F'(x)\} = \begin{cases} \{-1\} & \text{für } x < 0, \\ \{1\} & \text{für } x > 0. \end{cases}$$

Weiter gilt für  $(x_k) \subset U_d$ ,  $x_k \rightarrow 0$ ,

$$F'(x_k) = \begin{cases} -1 & \text{für } x_k < 0, \\ 1 & \text{für } x_k > 0. \end{cases}$$

Im Falle  $F'(x_k) \rightarrow M$  gilt  $M = -1$  oder  $M = 1$  und beide Limiten lassen sich offensichtlich erzeugen (durch  $x_k \rightarrow 0^-$  bzw.  $x_k \rightarrow 0^+$ ). Somit gilt  $\partial_B F(0) = \{-1, 1\}$ .

Damit folgt:

$$\partial^{cl} F(x) = \text{conv } \partial_B F(x) = \begin{cases} \{-1\} & \text{für } x < 0, \\ [-1, 1] & \text{für } x = 0, \\ \{1\} & \text{für } x > 0. \end{cases}$$

**Lemma 3.2.4.** Ist  $F : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$  stetig differenzierbar in einer Umgebung von  $x \in U$ , so gilt  $\partial^{cl} F(x) = \partial_B F(x) = \{F'(x)\}$ .

**Beweis:**

Für  $x_k \rightarrow x$  existiert  $F'(x_k)$  nahe bei  $x$  und es gilt  $F'(x_k) \rightarrow F'(x)$  (Stetigkeit). Daher  $\partial_B F(x) = \{F'(x)\}$ . Offensichtlich ist  $\partial^{cl} F(x) = \partial_B F(x)$ .  $\square$

**Beispiel 3.2.5.** Wir berechnen die verallgemeinerte Ableitung der Fischer–Burmeister-Funktion  $\phi_{FB}(a, b) = a + b - \sqrt{a^2 + b^2}$ . Wir prüfen zunächst die (hier sogar globale) Lipschitz-Stetigkeit: Für  $(a, b) \neq (0, 0)$  gilt

$$\phi'_{FB}(a, b) = \left(1 - \frac{a}{\sqrt{a^2 + b^2}}, 1 - \frac{b}{\sqrt{a^2 + b^2}}\right).$$

Offensichtlich haben wir  $\|\phi'_{FB}(a, b)\| \leq \|(1, 1)\| + 1 = 1 + \sqrt{2}$  und somit ist  $\phi_{FB}$  nach dem Mittelwertsatz Lipschitz-stetig mit  $L = 1 + \sqrt{2}$ , denn entlang jeder Geraden ist  $\phi_{FB}$  stetig und darüber hinaus differenzierbar bis auf höchstens einen Punkt mit  $\|\phi'_{FB}(a, b)\| \leq 1 + \sqrt{2}$ .

In  $(a, b) \neq (0, 0)$  haben wir also

$$\begin{aligned}\partial^{cl}\phi_{FB}(a, b) &= \partial_B\phi_{FB}(a, b) = \{\phi'_{FB}(a, b)\} \\ &= \left\{ \left( 1 - \frac{a}{\sqrt{a^2 + b^2}}, 1 - \frac{b}{\sqrt{a^2 + b^2}} \right) \right\}.\end{aligned}$$

Wegen

$$\phi'_{FB}(a, b) = (1, 1) - \frac{(a, b)}{\|(a, b)\|}$$

ist klar, dass die durch  $\phi'_{FB}(a_k, b_k)$  erzielbaren Limiten für  $(a_k, b_k) \rightarrow (0, 0)$  genau alle Vektoren der Form  $(1, 1) - v^T$ ,  $\|v\| = 1$ , sind. Somit gilt

$$\partial_B\phi_{FB}(0, 0) = \{(1, 1) - v^T; \|v\| = 1\}.$$

**Beispiel 3.2.6.** Ist  $F : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$  stetige Auswahl der  $C^1$ -Funktionen  $F^r : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ ,  $r = 1, \dots, l$ , ist also  $F$  stetig mit

$$F(x) \in \{F^1(x), \dots, F^l(x)\} \quad \forall x \in \mathbb{R}^n,$$

so gilt

$$\partial_B F(x) = \{F^{r'}(x); r \in I_e(x)\},$$

wobei  $I_e(x)$  die Menge der in  $x$  *essentiell aktiven Indizes* ist:

$$I_e(x) = \{r; x \in \text{cl}(\text{int}\{y \in U; F^r(y) = F(y)\})\}.$$

Der Nachweis erfordert einige Arbeit, siehe Scholtes [Sch94] und wird hier nicht geführt.

Das Clarksche Differential verallgemeinert das konvexe Subdifferential:

**Satz 3.2.7.** Sei  $f : U \rightarrow \mathbb{R}$  konvex auf der offenen konvexen Menge  $U \subset \mathbb{R}^n$ . Dann gilt  $\partial^{cl} f = \partial f^T$  auf  $U$ .

**Beweis:** Siehe Clarke [Cla83], Thm 2.5.1.

Wir beweisen nun einige Eigenschaften der angegebenen Differentiale.

**Definition 3.2.8.** Die mengenwertige Abbildung  $\Gamma : U \subset \mathbb{R}^n \rightrightarrows \mathbb{R}^{m \times n}$  heißt *oberhalbstetig* in  $x^* \in U$  wenn es zu jedem  $\varepsilon > 0$  ein  $\delta > 0$  gibt mit

$$\Gamma(B_\delta(x^*)) \subset \Gamma(x^*) + B_\varepsilon(0),$$

d.h. zu jedem  $x \in B_\delta(x^*)$  und jedem  $M_x \in \Gamma(x)$  ein  $M_{x^*} \in \Gamma(x^*)$  existiert mit  $\|M_x - M_{x^*}\| < \varepsilon$ .

**Satz 3.2.9.** Sei  $F : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$  eine lokal Lipschitz-stetige Funktion. Dann gilt: Die Differentiale  $\partial_B F$  und  $\partial^{cl} F$  haben nichtleere, kompakte Bilder und sind oberhalbstetig (also auch lokal beschränkt). Darüber hinaus hat  $\partial^{cl} F$  konvexe Bilder.

*Zusatz:* Ist  $L$  eine lokale Lipschitz-Konstante von  $F$  bei  $x^*$ , dann gilt  $\|M\| \leq L$  für alle  $M \in \partial^{cl} F(x^*) \subset \partial_B F(x^*)$ .

**Beweis:** Sei  $x^* \in U$  beliebig fest.

- $\partial_B F$  ist lokal beschränkt (und somit ist auch  $\partial_B F(x^*)$  beschränkt):

$F$  ist auf einer  $\delta$ -Umgebung  $B_\delta(x^*)$  von  $x^*$  Lipschitz-stetig mit Konstante  $L$ . Für alle  $x \in U_d \cap B_\delta(x^*)$  und alle  $s \in \mathbb{R}^n$  ergibt sich:

$$\|F'(x)s\| = \lim_{t \rightarrow 0} \left\| \frac{F(x+ts) - F(x)}{t} \right\| \leq \lim_{t \rightarrow 0} \frac{L\|x+ts - x\|}{|t|} = L\|s\|.$$

Somit gilt

$$(3.14) \quad \|F'(x)\| \leq L \quad \forall x \in U_d \cap B_\delta(x^*).$$

Zu  $x \in B_\delta(x^*)$  und  $M \in \partial_B F(x)$  gibt es  $x_k \in U_d \cap B_\delta(x^*)$  mit  $x_k \rightarrow x$  und  $F'(x_k) \rightarrow M$ , also

$$(3.15) \quad \|M\| = \lim_{k \rightarrow \infty} \|F'(x_k)\| \leq L.$$

- $\partial_B F(x^*) \neq \emptyset$ :

Da  $U_d$  dicht in  $U$  ist, gibt es  $(x_k) \subset U_d \cap B_\delta(x^*)$  mit  $x_k \rightarrow x^*$ . Wegen (3.14) gibt es eine Teilfolge  $(x_{k'})$  mit  $F'(x_{k'}) \rightarrow M$  und wir haben dann  $M \in \partial_B F(x^*)$ .

- $\partial_B F(x^*)$  ist abgeschlossen:

Sei  $M_k \in \partial_B F(x^*)$  mit  $M_k \rightarrow M$ . Zu  $M_k$  finden wir nach Definition von  $\partial_B F$  ein  $x_k \in U_d$  mit  $\|x_k - x^*\| < 1/k$  und  $\|F'(x_k) - M_k\| < 1/k$ . Nach Konstruktion gilt  $U_d \ni x_k \rightarrow x^*$  und

$$\|F'(x_k) - M\| \leq \|F'(x_k) - M_k\| + \|M_k - M\| \rightarrow 0 \quad \text{für } k \rightarrow \infty.$$

Daher haben wir  $M \in \partial_B F(x^*)$  und  $\partial_B F(x^*)$  ist somit abgeschlossen.

Aus Beschränktheit und Abgeschlossenheit folgt Kompaktheit.

- $\partial_B F$  ist oberhalbstetig in  $x^*$ :

Angenommen,  $\partial_B F$  ist nicht oberhalbstetig in  $x^*$ . Dann gibt es  $\varepsilon > 0$  und  $B_\delta(x^*) \ni y_k \rightarrow x^*$  mit  $M_k \in \partial_B F(y_k)$ ,  $\|M_k - M\| \geq \varepsilon$  für alle  $M \in \partial_B F(x^*)$ . Nun gibt es  $x_k \in U_d \cap B_\delta(x^*)$  mit  $\|x_k - y_k\| < 1/k$  und  $\|F'(x_k) - M_k\| < 1/k$ . Wegen (3.14) gibt es eine Teilfolge mit  $F'(x_{k'}) \rightarrow M$ . Nun gilt

$$\|x_{k'} - x^*\| \leq \|x_{k'} - y_{k'}\| + \|y_{k'} - x^*\| \rightarrow 0 \quad \text{für } k' \rightarrow \infty,$$

also  $M \in \partial_B F(x^*)$ . Weiter gilt

$$\|M_{k'} - M\| \leq \|M_{k'} - F'(x_{k'})\| + \|F'(x_{k'}) - M\| \rightarrow 0 \quad \text{für } k' \rightarrow \infty.$$

Dies widerspricht  $\|M_k - M\| \geq \varepsilon$ .

Nun zu  $\partial^{cl} F$ :

Wegen  $\partial^{cl} F(x^*) = \text{conv } \partial_B F(x^*) \supset \partial_B F(x^*)$  ist  $\partial^{cl} F(x^*)$  nichtleer und nach Satz ?? auch kompakt.

- $\partial^{cl} F$  ist oberhalbstetig in  $x^*$ :

Wegen der Oberhalbstetigkeit von  $\partial_B F$  gibt es zu  $\varepsilon > 0$  ein  $\rho > 0$  mit  $\partial_B F(B_\rho(x^*)) \subset \partial_B F(x^*) + B_\varepsilon$ . Ist nun  $x \in B_\rho(x^*)$  und  $M \in \partial^{cl} F(x)$ , so gibt es  $M_i \in \partial_B F(x)$  und  $\lambda_i \geq 0$ ,  $\sum_i \lambda_i = 1$ ,  $i = 1, \dots, l$ , mit  $M = \sum_i \lambda_i M_i$ . Weiter gibt es  $M_i^* \in \partial_B F(x^*)$  mit  $\|M_i - M_i^*\| < \varepsilon$  und für  $M^* = \sum_i \lambda_i M_i^* \in \partial^{cl} F(x^*)$  gilt:

$$\|M - M^*\| \leq \sum_i \lambda_i \|M_i - M_i^*\| < \varepsilon.$$

Zusatz:

Zu  $M \in \partial^{cl} F(x^*)$  gibt es Matrizen  $M_i \in \partial_B F(x^*)$  und  $\lambda_i \geq 0$ ,  $\sum_{i=1}^l \lambda_i = 1$ , mit  $M = \sum_i \lambda_i M_i$ . Aus (3.15) folgt nun

$$\|M\| \leq \sum_i \lambda_i \|M_i\| \leq L.$$

□

Wir stellen nun einige Rechenregeln für verallgemeinerte Ableitungen bereit:

**Satz 3.2.10 (Kettenregel).** Seien  $F : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$  und  $G : V \subset \mathbb{R}^m \rightarrow \mathbb{R}^l$  lokal Lipschitz-stetig und es gelte  $F(U) \subset V$ . Dann ist  $G \circ F : U \rightarrow \mathbb{R}^l$  lokal Lipschitz-stetig und es gilt für alle  $x \in U$  und  $v \in \mathbb{R}^n$ :

$$(3.16) \quad \partial^{cl}(G \circ F)(x)v \subset \text{conv}(\partial^{cl} G(F(x)) \partial^{cl} F(x)v) = \text{conv}(\partial^{cl} G(F(x)) \partial^{cl} F(x))v.$$

Ist  $G$  stetig differenzierbar in einer Umgebung von  $F(x)$ , so gilt

$$(3.17) \quad \partial^{cl}(G \circ F)(x)v = G'(F(x)) \partial^{cl} F(x)v.$$

Ist  $F$  stetig differenzierbar in einer Umgebung von  $x$ , so gilt

$$(3.18) \quad \partial^{cl}(G \circ F)(x)v \subset \partial^{cl} G(F(x)) F'(x)v.$$

Im Falle  $l = 1$  kann der rechts stehende Vektor  $v$  in (3.16)–(3.18) entfallen.

**Beweis:** Siehe z.B. Clarke [Cla83], §2.6.

□

### 3.3 Semiglattheit

Wir nutzen nun die beschriebenen verallgemeinerten Ableitungen zur Entwicklung nichtglatter Newton-Verfahren. Zum Nachweis der schnellen lokalen Konvergenz muss die Approximationsbedingung verifiziert werden. Dies führt auf den Begriff der Semiglattheit.

**Definition 3.3.1.** Die Funktion  $F : U \rightarrow \mathbb{R}^m$ ,  $U \subset \mathbb{R}^n$  offen, heißt *semiglat* in  $x \in U$ , falls sie Lipschitz-stetig nahe  $x$  und richtungsdifferenzierbar in  $x$  ist und falls gilt:

$$(3.19) \quad \max_{M \in \partial^{cl} F(x+s)} \|F(x+s) - F(x) - Ms\| = o(\|s\|) \quad \text{für } s \rightarrow 0.$$

Die Bedingung (3.19) ist nichts anderes als die Forderung, dass  $\partial^{cl} F$  eine punktbasierende Approximation von  $F$  in  $x$  ist.

Mit Blick auf q-quadratische Konvergenz ergibt sich daraus unmittelbar auch das folgende Konzept:

**Definition 3.3.2.** Die Funktion  $F : U \rightarrow \mathbb{R}^m$ ,  $U \subset \mathbb{R}^n$  offen, heißt *stark semiglat* (oder semiglat der Ordnung 1) in  $x \in U$ , falls sie Lipschitz-stetig nahe  $x$  und richtungsdifferenzierbar in  $x$  ist und falls gilt:

$$(3.20) \quad \max_{M \in \partial^{cl} F(x+s)} \|F(x+s) - F(x) - Ms\| = O(\|s\|^2) \quad \text{für } s \rightarrow 0.$$

Die Definition 3.3.1 von Semiglattheit wird zunehmend beliebter, weil sie sehr praxisnah ist. Historisch gesehen wurde Semiglattheit zunächst in der folgenden abstrakten, aber äquivalenten, Form definiert:

**Definition 3.3.3** (Mifflin [Mif77], Qi [Qi93], Qi und Sun [QS93]). Die Funktion  $F : U \rightarrow \mathbb{R}^m$ ,  $U \subset \mathbb{R}^n$  offen, heißt *semiglat* in  $x \in U$ , falls sie Lipschitz-stetig nahe  $x$  ist und falls der folgende Grenzwert für alle  $s \in \mathbb{R}^n$  existiert und endlich ist:

$$(3.21) \quad \ell(s) = \lim_{\substack{M \in \partial^{cl} F(x+ts') \\ s' \rightarrow s, \ t \rightarrow 0^+}} Ms'.$$

Zum besseren Verständnis geben wir (3.21) nochmals mithilfe von Folgen an:

Zu jedem  $s \in \mathbb{R}^n$  gibt es  $\ell(s) \in \mathbb{R}^m$ , so dass für alle Folgen  $(t_k) \subset (0, \infty)$ ,  $(s^k) \subset \mathbb{R}^n$  und  $(M_k) \subset \mathbb{R}^{m \times n}$  mit

$$\lim_{k \rightarrow \infty} t_k = 0, \quad \lim_{k \rightarrow \infty} s^k = s \quad \text{und} \quad M_k \in \partial^{cl} F(x + t_k s^k)$$

gilt:

$$\lim_{k \rightarrow \infty} M_k s^k = \ell(s).$$

Wir stellen zunächst fest, dass  $C^1$ -Funktionen semiglat sind.

**Satz 3.3.4.** Die Funktion  $F : U \rightarrow \mathbb{R}^m$ ,  $U \subset \mathbb{R}^n$  offen, sei stetig differenzierbar in einer Umgebung von  $x \in U$ . Dann ist  $F$  semiglat im Punkt  $x$ . Ist  $F'$  Lipschitz-stetig nahe  $x$ , so ist  $F$  stark semiglat in  $x$ .

**Beweis:**  $F$  ist  $C^1$  auf einer Umgebung  $B_\delta(x)$ . Auf  $B_\delta(x)$  gilt daher  $\partial^{cl} F = \{F'\}$  nach Lemma 3.2.4. Daraus folgt

$$\begin{aligned} \max_{M \in \partial^{cl} F(x+s)} \|F(x+s) - F(x) - Ms\| &= \|F(x+s) - F(x) - F'(x+s)s\| \\ &\leq \int_0^1 \|(F'(x+ts) - F'(x+s))s\| dt = o(\|s\|). \end{aligned}$$

Im Fall, dass  $F'$  Lipschitz-stetig nahe  $x$  ist, folgt sogar

$$\int_0^1 \|(F'(x+ts) - F'(x+s))s\| dt \leq \int_0^1 L(1-t)\|s\|^2 dt = \frac{L}{2}\|s\|^2 = O(\|s\|^2),$$

also starke Semiglattheit. □

**Satz 3.3.5.** Die Funktion  $F : U \rightarrow \mathbb{R}^m$ ,  $U \subset \mathbb{R}^n$  offen, ist semiglat in  $x \in U$  genau dann, wenn alle Komponentenfunktionen  $F_i$  semiglat in  $x$  sind.

**Beweis:** Sei  $F$  semiglatt in  $x$ . Dann ist  $F$  Lipschitz-stetig nahe  $x$  und richtungsdifferenzierbar in  $x$  und dies gilt dann auch für alle  $F_i$ . Aus der Kettenregel (3.17) mit  $G(z) = z_i$  folgt

$$\partial^{cl} F_i(y) = \partial^{cl}(G \circ F)(y) = G'(F(y)) \partial^{cl} F(y) = e_i^T \partial^{cl} F(y)$$

für alle  $y$  nahe  $x$ , wobei  $e_i \in \mathbb{R}^m$  den  $i$ -ten Einheitsvektor bezeichnet. Somit gilt:

$$\begin{aligned} \max_{v \in \partial^{cl} F_i(x+s)} |F_i(x+s) - F_i(x) - vs| &= \max_{M \in \partial^{cl} F(x+s)} |F_i(x+s) - F_i(x) - e_i^T Ms| \\ &= \max_{M \in \partial^{cl} F(x+s)} |e_i^T (F(x+s) - F(x) - Ms)| \leq \max_{M \in \partial^{cl} F(x+s)} \|F(x+s) - F(x) - Ms\| \\ &= o(\|s\|). \end{aligned}$$

Somit ist  $F_i$  semiglatt in  $x$ .

Sind umgekehrt alle Komponenten  $F_i$  semiglatt in  $x$ , so sind alle  $F_i$  Lipschitz-stetig nahe  $x$  und richtungsdifferenzierbar in  $x$ , und dies gilt dann auch für  $F$ . Weiter gilt

$$\begin{aligned} \max_{M \in \partial^{cl} F(x+s)} \|F(x+s) - F(x) - Ms\| &\leq \sum_{i=1}^n \max_{M \in \partial^{cl} F(x+s)} |F_i(x+s) - F_i(x) - e_i^T Ms| \\ &\stackrel{\text{s.o.}}{=} \sum_{i=1}^n \max_{v \in \partial^{cl} F_i(x+s)} |F_i(x+s) - F_i(x) - vs| = \sum_{i=1}^n o(\|s\|) = o(\|s\|). \end{aligned}$$

□

Wir stellen nun noch einige Aussagen bereit, die hilfreich sind, um die Semiglattheit von Funktionen nachzuweisen:

**Satz 3.3.6.**

- a) *Stückweise  $C^1$ -Funktionen sind semiglatt.*
- b) *Konvexe Funktionen sind semiglatt.*
- c) *Summen, Produkte und Kompositionen semiglatte Funktionen sind semiglatt.*

**Folgerungen:**

1. Die Fischer-Burmeister-Funktion und die min-NCP-Funktion sind semiglatt, da konvex.
2. Betrachte für stetig differenzierbares  $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$  das NCP

$$x \geq 0, \quad F(x) \geq 0, \quad x_i F_i(x) = 0, \quad i = 1, \dots, n.$$

Dann ist in der Reformulierung

$$(\phi(x_i, F_i(x)))_{1 \leq i \leq n} = 0$$

mit der Fischer-Burmeister-Funktion oder min-NCP-Funktion  $\phi$  semiglatt.

3. Reformulierungen von KKT-Systemen mit Hilfe von obigen NCP-Funktionen für Optimierungsprobleme mit zweimal stetig differenzierbaren Nebenbedingungen und Zielfunktion sind semiglatt.

### 3.4 Semiglatte Newton-Verfahren

Zur Lösung des Gleichungssystems

$$(3.22) \quad F(x) = 0,$$

wobei  $F : U \rightarrow \mathbb{R}^n$ ,  $U \subset \mathbb{R}^n$ , semiglatt in der Lösung  $x^* \in U$  sei, betrachten wir nun unser allgemeines nichtglatte Newton-Verfahren 3.1.1 mit der punktbasierten Wahl  $M_k \in \partial^{cl} F(x^k)$ , setzen also  $M_F(x) = \partial^{cl} F(x)$ :

**Algorithmus 3.4.1** (Semiglatte Newton-Verfahren von Qi und Sun [Qi93, QS93]).

0. Wähle  $x^0 \in \mathbb{R}^n$ .

Für  $k = 0, 1, 2, \dots$ :

1. Falls  $F(x^k) = 0$ : STOP mit Ergebnis  $x^k$ .
2. Wähle  $M_k \in \partial^{cl} F(x^k)$  und berechne den Schritt  $s^k \in \mathbb{R}^n$  durch Lösen von

$$M_k s^k = -F(x^k).$$

3. Setze  $x^{k+1} = x^k + s^k$ .

Die zentrale Bedingung (3.19) für Semiglattheit bzw. (3.20) für starke Semiglattheit liefert:

**Lemma 3.4.2.** *Sei  $F : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$  semiglatt in  $x^* \in U$ . Dann ist  $\partial^{cl} F$  eine punktbasierte Approximation von  $F$  in  $x^* \in U$ . Ist  $F$  stark semiglatt in  $x^*$ , so ist  $\partial^{cl} F$  eine punktbasierte Approximation von  $F$  in  $x^*$  der Ordnung 1.*

**Beweis:** Wegen der Semiglattheit von  $F$  in  $x^*$  gilt (3.19). Dies ist genau die Approximationsbedingung (3.12) mit  $M_F = \partial^{cl} F$ . Entsprechend gilt im Fall der starken Semiglattheit die Bedingung (3.20) und diese ist äquivalent zu (3.13) mit  $M_F = \partial^{cl} F$ .  $\square$

Neben der durch Lemma 3.4.2 bereitgestellten Approximationsbedingung fehlt uns nun nur noch der Nachweis der Regularitätsbedingung (3.11), um die lokale q-superlineare Konvergenz von Algorithmus 3.4.1 nachzuweisen.

Wir fordern hierzu:

**Voraussetzung 3.4.3.**  $\partial^{cl} F$  ist in der Lösung  $x^*$  CD-regulär (CD für Clarke-Differential), d.h. alle  $M \in \partial^{cl} F(x^*)$  sind invertierbar.

**Lemma 3.4.4.** *Die Funktion  $F : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$  sei Lipschitz-stetig in einer Umgebung von  $x^* \in U$  und es gelte Voraussetzung 3.4.3. Dann gibt es  $\eta > 0$  und  $C > 0$  mit*

$$(3.23) \quad M \text{ invertierbar, } \|M^{-1}\| \leq C \quad \forall M \in \partial^{cl} F(x), \quad \forall x \in B_\eta(x^*).$$

**Beweis:** Nach Satz 3.2.9 ist  $\partial^{cl} F(x^*)$  nichtleer, kompakt und  $\partial^{cl} F(x^*)$  ist oberhalbstetig in  $x^*$ . Die Funktion  $M \in \partial^{cl} F(x^*) \rightarrow \|M^{-1}\|$  ist nach Voraussetzung 3.4.3 wohldefiniert und bekanntlich stetig, nimmt also auf dem Kompaktum  $\partial^{cl} F(x^*)$  ein Maximum  $\hat{C}$  an. Es gilt also

$$\|M^{-1}\| \leq \hat{C} \quad \forall M \in \partial^{cl} F(x^*).$$

Für jede Störung  $E \in \mathbb{R}^{n,n}$  mit  $\|E\| \leq \frac{1}{2\hat{C}} =: \varepsilon$  gilt nach dem Banach-Lemma also auch

$$(3.24) \quad \|(M + E)^{-1}\| \leq 2\hat{C} \quad \forall M + E \in \partial^{cl}F(x^*) + B_\varepsilon(0).$$

Wegen der Oberhalbstetigkeit von  $\partial^{cl}F(x^*)$  existiert  $\eta > 0$  mit  $\partial^{cl}F(x) \subset \partial^{cl}F(x^*) + B_\varepsilon(0)$  für alle  $x \in B_\eta(x^*)$ . Somit gilt (3.23) nach (3.24) mit  $C = 2\hat{C}$ .  $\square$

Wir erhalten den folgenden Konvergenzsatz:

**Satz 3.4.5.** *Sei  $x^* \in U$  eine Lösung von (3.22), in der  $F : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$  semiglatt ist, und es gelte Voraussetzung 3.4.3. Dann gibt es  $\delta > 0$ , so dass Algorithmus 3.4.1 für alle  $x_0 \in \mathbb{R}^n$  mit  $\|x_0 - x^*\| < \delta$  entweder mit  $x_k = x^*$  terminiert oder eine Folge  $(x_k)$  erzeugt, die  $q$ -superlinear gegen  $x^*$  konvergiert.*

*Ist  $F$  stark semiglatt in  $x^*$ , so ist die Konvergenzrate  $q$ -quadratisch.*

**Beweis:** Algorithmus 3.4.1 ist identisch zu Algorithmus 3.1.1 mit punktbasierter Wahl  $M_k \in \partial^{cl}F(x_k)$ , d.h.  $M_F = \partial^{cl}F$ . In den Lemmas 3.4.2 und 3.4.4 haben wir die (quadratische) Approximationsbedingung (3.12) bzw. (3.13) und die Regularitätsbedingung (3.11) nachgewiesen. Somit ist Korollar 3.1.7 anwendbar und liefert die Behauptung.  $\square$

**Bemerkung.** Die Aussagen dieses Abschnittes 3.4 lassen sich unmittelbar auch auf andere Differentiale übertragen:

Ohne jede Änderung können wir anstelle von  $\partial^{cl}F$  mit dem kleineren Differential  $\partial_B F$  arbeiten. Die Regularitätsbedingung 3.4.3 stellt dann nur noch Anforderungen an die Elemente von  $\partial_B F(x^*)$  und wird daher BD-Regularität genannt (BD für B-Differential).



## Kapitel 4

# Konzepte und Methoden für erweitert-reellwertige Funktionen

Wir betrachten nun konvexe erweitert-reellwertige Funktionen  $f : \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$  und geben eine kurze Einführung in die Analysis und mögliche Optimierungsmethoden für diese Funktionenklasse. Umfassende Lektüre zu den hier zusammengefassten Themen findet sich in [BC11,Ro70,RW09].

### 4.1 Konvexe Analysis für erweitert-reellwertige Funktionen

#### 4.1.1 Wichtige Begriffe

Die Definition der Konvexität kann unverändert übertragen werden, wobei die Konvention verwendet wird, dass eine Ungleichung auch dann gilt, wenn auf beiden Seiten  $+\infty$  steht.

Das Subdifferential  $\partial f(x)$  einer erweitert-reellwertigen konvexen Funktion kann in Punkten mit  $f(x) < \infty$  genauso definiert werden wie im  $\mathbb{R}$ -wertigen Fall. Man setzt  $\partial f(x) = \emptyset$ , falls  $f(x) = \infty$ .

Es gilt weiterhin  $g \in \partial f(x)$  genau dann, wenn der Vektor  $\begin{pmatrix} g \\ -1 \end{pmatrix}$  die Normale einer stützenden Hyperebene von  $\text{epi}(f)$  im Punkt  $\begin{pmatrix} x \\ f(x) \end{pmatrix}$  ist.

Der *wesentliche Definitionsbereich* von  $f$  wird definiert durch  $\text{dom}(f) = f^{-1}(\mathbb{R})$  und  $f$  heißt *echt*, falls  $\text{dom}(f) \neq \emptyset$  gilt.

Es kann passieren, dass  $\partial f(x) = \emptyset$  gilt, obwohl  $x \in \text{dom}(f)$ . Zum Beispiel ist  $f(x) = -\sqrt{x}$  für  $x \geq 0$  und  $f(x) = +\infty$  für  $x < 0$  konvex und  $\partial f(0) = \emptyset$ .

Man kann aber zeigen:

$$\partial f(x) \neq \emptyset \quad \forall x \in \text{int}(\text{dom}(f)).$$

In einer Umgebung von  $x \in \text{int}(\text{dom}(f))$  ist  $f$  zudem Lipschitz-stetig.

Weiterhin gilt dass  $\bar{x} \in \text{dom}(f)$  ein Minimum von  $f$  ist genau dann, wenn  $0 \in \partial f(\bar{x})$ .

Die Funktion  $f : \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$  heißt *unterhalbstetig* in  $x$ , falls für jede Folge  $x^k \rightarrow x$  gilt

$$\liminf_{k \rightarrow \infty} f(x^k) \geq f(x).$$

Äquivalent dazu ist die Abgeschlossenheit des Epigraphs oder auch die Abgeschlossenheit der Niveaumengen  $\{x; f(x) \leq c\}$ .

Unterhalbstetigkeit stellt sicher, dass im Falle  $f(x^k) \rightarrow \inf_x f(x)$  jeder Häufungspunkt von  $(x^k)$  ein globales Minimum von  $f$  ist.

#### 4.1.2 Die Proximalabbildung

Für die echte, konvexe, unterhalbstetige Funktion  $f : \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$  ist die Proximalabbildung (der prox-Operator)  $\text{prox}_f : \mathbb{R}^n \rightarrow \mathbb{R}^n$  definiert durch

$$\text{prox}_f(x) = \arg \min_y f(y) + \frac{1}{2} \|y - x\|^2 \quad \forall x \in \mathbb{R}^n.$$

Die Zielfunktion des prox-Problems ist gleichmäßig konvex und unterhalbstetig, besitzt also eine eindeutige Lösung.

**Lemma 4.1.1.** *Sei  $g : \mathbb{R}^n \rightarrow \mathbb{R}$  konvex und stetig differenzierbar sowie  $h : \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$  echt und konvex. Dann ist  $\bar{x}$  ein Minimum von  $g + h$  genau dann, wenn gilt:*

$$(4.1) \quad \nabla g(\bar{x})^T (x - \bar{x}) + h(x) - h(\bar{x}) \geq 0 \quad \forall x \in \mathbb{R}^n.$$

*Dies ist äquivalent zu*

$$(4.2) \quad -\nabla g(\bar{x}) \in \partial h(\bar{x}).$$

**Beweis:** Sei  $\bar{x}$  ein Minimum von  $g + h$ . Für  $x \in \mathbb{R}^n$  und  $t \in (0, 1]$  haben wir dann mit  $y = (1 - t)\bar{x} + tx$ :

$$g(\bar{x}) + h(\bar{x}) \leq g(y) + h(y) \leq g(\bar{x} + t(x - \bar{x})) + (1 - t)h(\bar{x}) + th(x).$$

Dies liefert

$$0 \leq (g(\bar{x} + t(x - \bar{x})) - g(\bar{x})) + t(h(x) - h(\bar{x})).$$

Multiplizieren mit  $1/t$  und  $t \rightarrow 0^+$  liefert

$$0 \leq \nabla g(\bar{x})^T (x - \bar{x}) + h(x) - h(\bar{x}).$$

Der Rest folgt aus der Subgradientenungleichung.

Gilt umgekehrt (4.1) (oder (4.2)), dann folgt

$$g(x) + h(x) - (g(\bar{x}) + h(\bar{x})) \geq \nabla g(\bar{x})^T (x - \bar{x}) + h(x) - h(\bar{x}) \geq 0.$$

□

**Satz 4.1.2.** *Sei  $f : \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$  echt, konvex und unterhalbstetig. Dann gilt:*

a)  $\text{prox}_f$  ist wohldefiniert.

b)  $y = \text{prox}_f(x)$  ist die eindeutige Lösung von

$$x \in y + \partial f(y).$$

c)  $\text{prox}_f$  ist (firmly) nichtexpansiv:

$$\|\text{prox}_f(x) - \text{prox}_f(\hat{x})\|^2 \leq (\text{prox}_f(x) - \text{prox}_f(\hat{x}))^T (x - \hat{x}) \leq \|\text{prox}_f(x) - \text{prox}_f(\hat{x})\| \|x - \hat{x}\|.$$

d) Die Funktion (Moreau-Envelope)

$$m_f(x) = \min_y f(y) + \frac{1}{2} \|y - x\|^2$$

ist differenzierbar mit Lipschitz-stetigem Gradienten

$$\nabla m_f(x) = x - \text{prox}_f(x).$$

e) Es gilt die Moreau-Dekomposition

$$\text{prox}_f(x) + \text{prox}_{f^*}(x) = x \quad \forall x \in \mathbb{R}^n.$$

Hierbei ist  $f^*$  die konvexe Konjugierte von  $f$ :

$$f^*(y) = \sup_{x \in \mathbb{R}^n} (y^T x - f(x)).$$

Diese ist echt, konvex und unterhalbstetig.

**Beweis:** Wir zeigen nur b) und c).

zu b): Wir können Lemma 4.1.1 mit  $h = f$  und  $g(y) = \frac{1}{2} \|y - x\|^2$  anwenden und erhalten für  $y = \text{prox}_f(x)$ :

$$x - y = -\nabla g(y) \in \partial h(y).$$

zu a): Für  $y = \text{prox}_f(x)$  und  $\hat{y} = \text{prox}_f(\hat{x})$  gilt wegen b):  $x - y \in \partial f(y)$ ,  $\hat{x} - \hat{y} \in \partial f(\hat{y})$  und somit:

$$\begin{aligned} f(\hat{y}) - f(y) &\geq (x - y)^T (\hat{y} - y), \\ f(y) - f(\hat{y}) &\geq (\hat{x} - \hat{y})^T (y - \hat{y}). \end{aligned}$$

Addieren liefert

$$0 \geq (x - y)^T (\hat{y} - y) + (\hat{x} - \hat{y})^T (y - \hat{y}) = (y - \hat{y})^T (\hat{x} - \hat{y} - (x - y)) = -(y - \hat{y})^T (x - \hat{x}) + \|y - \hat{y}\|^2.$$

Dies zeigt

$$\|y - \hat{y}\|^2 \leq (y - \hat{y})^T (x - \hat{x}) \leq \|y - \hat{y}\| \|x - \hat{x}\|.$$

□

## 4.2 Das proximale Gradientenverfahren

Wie betrachten das Problem

$$\min_x g(x) + h(x),$$

wobei  $g : \mathbb{R}^n \rightarrow \mathbb{R}$  konvex und stetig differenzierbar und  $h : \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$  echt, konvex und unterhalbstetig ist.

Mit Lemma 4.1.1 erhält man die notwendige und hinreichende Optimalitätsbedingung

$$0 \in \nabla g(\bar{x}) + \partial h(\bar{x}).$$

Mit  $t > 0$  können wir dies schreiben als

$$\bar{x} - t\nabla g(\bar{x}) \in \bar{x} + t\partial h(\bar{x}) = \bar{x} + \partial(th)(\bar{x}).$$

Betrachten wir Satz 4.1.2, b) dann ist dies gleichbedeutend mit

$$\bar{x} = \text{prox}_{th}(\bar{x} - t\nabla g(\bar{x})).$$

Die Lösung ist also Fixpunkt der Abbildung

$$F(x) = \text{prox}_{th}(x - t\nabla g(x)).$$

Das *proximale Gradientenverfahren* mit fester Schrittweite  $t$  ist nun definiert durch die Fixpunktiteration

$$x^{k+1} = \text{prox}_{th}(x^k - t\nabla g(x^k)).$$

Der Schritt

$$x^k \mapsto x^{k+\frac{1}{2}} := x^k - t\nabla g(x^k)$$

ist ein Gradientenschritt für den glatten Teil  $g$  mit Schrittweite  $t$ .

Weiter ist  $x^{k+1} = \text{prox}_{th}(x^{k+\frac{1}{2}})$  Lösung von

$$\min_y th(y) + \frac{1}{2}\|y - x^{k+\frac{1}{2}}\|^2$$

Wieder ist  $t$  mit einer Schrittweite vergleichbar, denn wir können auch mit  $1/t$  multiplizieren und erhalten

$$\min_y h(y) + \frac{1}{2t}\|y - x^{k+\frac{1}{2}}\|^2$$

Je kleiner  $t$ , desto stärker wird die Entfernung zu  $x^{k+\frac{1}{2}}$  penalisiert.

Nach Konstruktion des prox-Problems gilt

$$h(x^{k+1}) \leq h(x^{k+\frac{1}{2}}) - \frac{1}{2t}\|x^{k+1} - x^{k+\frac{1}{2}}\|^2 \leq h(x^{k+\frac{1}{2}}),$$

d.h. es handelt sich um einen Abstiegschritt für  $h$ .

Wir betrachten diesen Schritt für den Fall, dass  $h$   $C^1$  ist. Dann gilt

$$0 = t\nabla h(x^{k+1}) + x^{k+1} - x^{k+\frac{1}{2}},$$

also

$$x^{k+1} = x^{k+\frac{1}{2}} - t\nabla h(x^{k+1}).$$

Es handelt sich hier also um einen „impliziten“ Gradientenschritt (vergleichbar mit implizitem Euler für gew. DGLn). Der normale Gradientenschritt ist explizit (wie expliziter Euler für DGLn).

Man nennt die proximale Gradientenmethode daher auch *Forward-Backward Splitting*.

Wir geben noch eine andere Herleitung des gleichen Verfahrens:

Zur Schrittberechnung  $x^k \rightarrow x^{k+1}$

- linearisieren wir  $g$  in  $x^k$ ,
- lassen  $h$  unverändert
- addieren einen Proximalterm  $\frac{1}{2t}\|x - x^k\|^2$  zur Schrittweitensteuerung.

Das liefert für  $x^{k+1}$  das Teilproblem

$$(4.3) \quad \min_x \nabla g(x^k)^T(x - x^k) + h(x) + \frac{1}{2t}\|x - x^k\|^2.$$

Dies wandeln wir in ein prox-Problem für  $th$  um:

$$\frac{1}{2t}\|x - (x^k - t\nabla g(x^k))\|^2 = \frac{1}{2t}\|x - x^k\|^2 + (x - x^k)^T \nabla g(x^k) + \frac{t}{2}\|\nabla g(x^k)\|^2$$

zeigt, dass (4.3) auch geschrieben werden kann als

$$\min_x h(x) + \frac{1}{2t}\|x - (x^k - t\nabla g(x^k))\|^2.$$

Wir haben lediglich den  $x$ -unabhängigen Offset  $-\frac{t}{2}\|\nabla g(x^k)\|^2$  weggelassen.

Wir sind bei einem prox-Problem für  $th$  angelangt. Dessen Lösung  $x^{k+1}$  erfüllt:

$$x^{k+1} = \text{prox}_{th}(x^k - t\nabla g(x^k)).$$

Dies ist genau die Iteration des proximalen Gradientenverfahrens.

#### 4.2.1 Konvergenz des Verfahrens

**Voraussetzung 4.2.1.** a)  $g$  ist konvex und stetig differenzierbar.

b)  $g$  ist echt, konvex und unterhalbstetig.

c) Es gibt  $\mu \geq 0$  so dass  $g(x) - \frac{\mu}{2}\|x\|^2$  konvex ist.

d) Es gibt  $L > 0$  so dass  $\frac{L}{2}\|x\|^2 - g(x)$  konvex ist.

e) Die Funktion  $f = g + h$  nimmt ihr Minimum in einem (nicht notwendig eindeutigen) Punkt  $\bar{x}$  an.

**Bemerkung.** Bedingung d) ist gleichbedeutend damit, dass  $\nabla g$  Lipschitz-stetig ist mit Konstante  $L$ : Ist  $\nabla g$  Lipschitz-stetig, dann folgt für alle  $x, y \in \mathbb{R}^n$ :

$$\begin{aligned} \frac{L}{2}\|y\|^2 - g(y) - \frac{L}{2}\|x\|^2 + g(x) &= - \int_0^1 \nabla g(x + t(y-x))^T(y-x) dt + \frac{L}{2}(\|y\|^2 - \|x\|^2) \\ &= -\nabla g(x)^T(y-x) + \int_0^1 (\nabla g(x) - \nabla g(x + t(y-x)))^T(y-x) dt + \frac{L}{2}(\|y\|^2 - \|x\|^2) \\ &\geq -\nabla g(x)^T(y-x) - \frac{L}{2}\|y-x\|^2 + \frac{L}{2}(\|y\|^2 - \|x\|^2) = Lx^T(y-x) - \nabla g(x)^T(y-x). \end{aligned}$$

Damit ist  $(L/2)\|x\|^2 - g(x)$  konvex.

Ist umgekehrt  $(L/2)\|x\|^2 - g(x)$  konvex, dann gilt für alle  $x, y \in \mathbb{R}^n$ :

$$\begin{aligned} Lx^T(y-x) - \nabla g(x)^T(y-x) &\leq \frac{L}{2}\|y\|^2 - g(y) - \frac{L}{2}\|x\|^2 + g(x), \\ Ly^T(x-y) - \nabla g(y)^T(x-y) &\leq \frac{L}{2}\|x\|^2 - g(x) - \frac{L}{2}\|y\|^2 + g(y), \end{aligned}$$

Addieren liefert

$$-L\|y-x\|^2 + (\nabla g(y) - \nabla g(x))^T(y-x) \leq 0.$$

Daraus folgt

$$\|\nabla g(y) - \nabla g(x)\|\|y-x\| \leq (\nabla g(y) - \nabla g(x))^T(y-x) \leq L\|y-x\|^2,$$

also die L-Stetigkeit von  $\nabla g$  mit Konstante  $L$ .

Wir führen nun Überlegungen durch, die die Konvergenz des Verfahrens zeigen. Zunächst tun wir dies für eine konstante Schrittweite  $t$ . Insbes. ist  $t \in (0, 1/L]$  zugelassen, wobei  $L$  aus d) stammt.

Aus c) folgt für alle  $x, y \in \mathbb{R}^n$ :

$$g(y) - \frac{\mu}{2}\|y\|^2 - g(x) + \frac{\mu}{2}\|x\|^2 \geq \nabla g(x)^T(y-x) - \mu x^T(y-x).$$

Wegen

$$\frac{\mu}{2}\|y\|^2 - \frac{\mu}{2}\|x\|^2 - \mu x^T(y-x) = \frac{\mu}{2}(\|y\|^2 - x^T y + \|x\|^2) = \frac{\mu}{2}\|y-x\|^2$$

erhalten wir:

$$(4.4) \quad g(y) \geq g(x) + \nabla g(x)^T(y-x) + \frac{\mu}{2}\|y-x\|^2.$$

Entsprechend erhalten wir aus d):

$$\frac{L}{2}\|y\|^2 - g(y) - \frac{L}{2}\|x\|^2 + g(x) \geq Lx^T(y-x) - \nabla g(x)^T(y-x).$$

Daraus folgt

$$(4.5) \quad g(y) \leq g(x) + \nabla g(x)^T(y-x) + \frac{L}{2}\|y-x\|^2.$$

Wir definieren den proximalen Gradienten-Schritt durch

$$s_t(x) = -\frac{1}{t}(x - \text{prox}_{th}(x - t\nabla g(x))).$$

Es gilt dann (zur Abkürzung sei  $x = x^k$  und  $x^+ = x^{k+1}$ ):

$$x^+ = \text{prox}_{th}(x - t\nabla g(x)) = x + ts_t(x).$$

Aus Satz 4.1.2 b) folgt wegen  $x + ts_t(x) = \text{prox}_{th}(x - t\nabla g(x))$ :

$$(x - t\nabla g(x)) - (x + ts_t(x)) \in \partial(th)(x + ts_t(x)).$$

Umstellen ergibt

$$-ts_t(x) \in t\nabla g(x) + \partial(th)(x + ts_t(x)),$$

also

$$(4.6) \quad -s_t(x) \in \nabla g(x) + \partial h(x + ts_t(x)).$$

Wir verwenden die Abschätzungen (4.4) und (4.5) für  $y = x + ts_t(x)$  und erhalten:

$$\frac{\mu t^2}{2} \|s_t(x)\|^2 \leq g(x + ts_t(x)) - g(x) - t\nabla g(x)^T s_t(x) \leq \frac{Lt^2}{2} \|s_t(x)\|^2.$$

Für  $0 < t \leq 1/L$  folgt

$$(4.7) \quad \frac{\mu t^2}{2} \|s_t(x)\|^2 \leq g(x + ts_t(x)) - g(x) - t\nabla g(x)^T s_t(x) \leq \frac{t}{2} \|s_t(x)\|^2.$$

Im Fall  $s_t(x) \neq 0$  folgt dann auch  $\mu t \leq 1$ .

Sei ab jetzt  $t > 0$  so dass die rechte Ungleichung in (4.7) erfüllt ist.

Wir haben für alle  $z \in \mathbb{R}^n$

$$\begin{aligned} f(x + ts_t(x)) &= g(x + ts_t(x)) + h(x + ts_t(x)) \\ &\stackrel{(4.7), 2. \text{Ungl.}}{\leq} g(x) + t\nabla g(x)^T s_t(x) + \frac{t}{2} \|s_t(x)\|^2 + h(x + ts_t(x)) \\ &\stackrel{(4.4)}{\leq} g(z) - \nabla g(x)^T (z - x) - \frac{\mu}{2} \|z - x\|^2 + t\nabla g(x)^T s_t(x) + \frac{t}{2} \|s_t(x)\|^2 + h(x + ts_t(x)) \\ &\leq g(z) - \nabla g(x)^T (z - x) - \frac{\mu}{2} \|z - x\|^2 + t\nabla g(x)^T s_t(x) + \frac{t}{2} \|s_t(x)\|^2 \\ &\quad + h(z) + (s_t(x) + \nabla g(x))^T (z - (x + ts_t(x))) \\ &= g(z) + h(z) + s_t(x)^T (z - x) - \frac{t}{2} \|s_t(x)\|^2 - \frac{\mu}{2} \|z - x\|^2. \end{aligned}$$

In der letzten Ungleichung wurde  $-s_t(x) - \nabla g(x) \in \partial h(x + ts_t(x))$  verwendet, siehe (4.6).

Damit ergibt sich

$$(4.8) \quad f(x + ts_t(x)) \leq f(z) + s_t(x)^T (z - x) - \frac{t}{2} \|s_t(x)\|^2 - \frac{\mu}{2} \|z - x\|^2.$$

Wir setzen  $x^+ := x + ts_t(x)$ .

Für Schrittweiten  $t > 0$ , die die 2. Ungleichung in (4.7) erfüllen, folgt mit  $z = x$  aus (4.8):

$$f(x^+) \leq f(x) - \frac{t}{2} \|s_t(x)\|^2.$$

Einsetzen von  $z = x^*$  in (4.8) ergibt:

$$\begin{aligned} f(x^+) - f(x^*) &\leq -s_t(x)^T (x - x^*) - \frac{t}{2} \|s_t(x)\|^2 - \frac{\mu}{2} \|x - x^*\|^2 \\ &= \frac{1}{2t} (\|x - x^*\|^2 - \|x - x^* + ts_t(x)\|^2) - \frac{\mu}{2} \|x - x^*\|^2 \\ (4.9) \quad &= \frac{1}{2t} ((1 - \mu t) \|x - x^*\|^2 - \|x^+ - x^*\|^2) \\ (4.10) \quad &\leq \frac{1}{2t} (\|x - x^*\|^2 - \|x^+ - x^*\|^2). \end{aligned}$$

Wir wenden die Ungleichung (4.10) mit  $x = x^j$ ,  $x^+ = x^{j+1}$  sowie konstanter Scheitweite  $t > 0$ , die die 2. Ungleichung in (4.7) erfüllt, an und erhalten:

$$\begin{aligned} \sum_{j=1}^k (f(x^j) - f(x^*)) &\leq \frac{1}{2t} \sum_{j=0}^{k-1} (\|x^j - x^*\|^2 - \|x^{j+1} - x^*\|^2) \leq \frac{1}{2t} (\|x^0 - x^*\|^2 - \|x^k - x^*\|^2) \\ &\leq \frac{1}{2t} \|x^0 - x^*\|^2. \end{aligned}$$

Damit folgt  $f(x^k) \rightarrow f(x^*)$ . Wegen  $f(x^{j+1}) \leq f(x^j)$  ergibt sich:

$$f(x^k) - f(x^*) \leq \frac{1}{k} \sum_{j=1}^k (f(x^j) - f(x^*)) \leq \frac{1}{2kt} \|x^0 - x^*\|^2.$$

Ist  $g$  gleichmäßig konvex, d.h.  $\mu > 0$ , dann folgt aus (4.9):

$$\|x^+ - x^*\|^2 \leq f(x^+) - f(x^*) + (1 - \mu t) \|x - x^*\|^2 \leq (1 - \mu t) \|x - x^*\|^2.$$

Damit haben wir lineare Konvergenz  $x^k \rightarrow x^*$  mit Rate  $\sqrt{1 - \mu t}$ .

Zusammenfassend erhalten wir:

**Satz 4.2.2.** *Es gelte die Voraussetzung 4.2.1. Das proximale Gradientenverfahren verwende eine konstante Schrittweite  $t \in (0, 1/L]$ . Dann gilt  $f(x^k) - f(x^*) = O(1/k)$  für  $k \rightarrow \infty$ , wobei  $x^*$  ein Minimum von  $f$  bezeichnet. Ist  $\mu > 0$ , dann gilt*

$$\|x^{k+1} - x^*\| \leq \sqrt{1 - \mu t} \|x^k - x^*\|.$$

Man muss zudem  $L$  nicht kennen, um etwa  $0 < t \leq 1/L$  zu wählen, sondern kann mit Schrittweiten  $t_k$  arbeiten, die die rechte Ungleichung in (4.7) erfüllen:

$$(4.11) \quad g(x^k + t_k s_{t_k}(x^k)) - g(x^k) - t_k \nabla g(x^k)^T s_{t_k}(x^k) \leq \frac{t_k}{2} \|s_{t_k}(x^k)\|^2.$$

Dies gilt insbesondere, sobald  $t_k \leq 1/L$  gilt.

Man kann also z.B. folgende Regel verwenden (mit festen Parametern  $t_{\max} > 0$  und  $\beta \in (0, 1)$ ):

**Schrittweitenregel:**

Bestimme das größte  $t_k \in \{t_{\max}, \beta t_{\max}, \beta^2 t_{\max}\}$ , so dass (4.11) gilt.

Man hat dann  $t_k \geq \min\{t_{\max}, \beta/L\} =: t_{\min}$  und kann zeigen:

$$(0 \leq ) f(x^{k+1}) - f(x^*) \leq \frac{1}{2t_k} (\|x^k - x^*\|^2 - \|x^{k+1} - x^*\|^2) \leq \frac{1}{2t_{\min}} (\|x^k - x^*\|^2 - \|x^{k+1} - x^*\|^2).$$

Daraus ergeben sich dann die gleichen Aussagen wie in Satz 4.2.2, allerdings mit  $t$  ersetzt durch  $t_{\min}$ .



# Literaturverzeichnis

- [BC11] H. H. Bauschke, P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Springer, 2011.
- [Ber99] D. P. Bertsekas, *Nonlinear Programming*, Athena Scientific, Belmont, MA, 1999.
- [Cla83] F. H. Clarke, *Optimization and nonsmooth analysis*, John Wiley & Sons, Inc., New York, 1983.
- [Cla98] F. H. Clarke, Yu. S. Ledyaeu, R. J. Stern und P. R. Wolenski, *Nonsmooth analysis and control theory*, Graduate Texts in Mathematics, 178, Springer-Verlag, New York, 1998.
- [GK99] C. Geiger, C. Kanzow, *Numerische Verfahren zur Lösung unrestringierter Optimierungsaufgaben*, Springer-Verlag, 1999.
- [GK02] C. Geiger, C. Kanzow, *Theorie und Numerik restringierter Optimierungsaufgaben*, Springer-Verlag, 2002.
- [HU93a] J.-B. Hiriart-Urruty und C. Lemaréchal, *Convex analysis and minimization algorithms I*, Grundlehren der Mathematischen Wissenschaften, 305, Springer-Verlag, Berlin, 1993.
- [HU93b] J.-B. Hiriart-Urruty und C. Lemaréchal, *Convex analysis and minimization algorithms II*, Grundlehren der Mathematischen Wissenschaften, 306, Springer-Verlag, Berlin, 1993.
- [KS86] M. Kojima und S. Shindo, *Extension of Newton and quasi-Newton methods to systems of  $PC^1$  equations*, J. Oper. Res. Soc. Japan 29 (1986), 352–375.
- [Lem89] C. Lemaréchal, *Nondifferentiable optimization*, in: G.L. Nemhauser, A. H. G. Rinnooy Kan, M. J. Todd (eds.), *Optimization, Handbooks in Operations Research and Management Science*, 1, North-Holland Publishing Co., Amsterdam, 1989, 529–572.
- [Mif77] R. Mifflin, *Semismooth and semiconvex functions in constrained optimization*, SIAM J. Control Optim. 15 (1977) 957–972.
- [Ro70] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, 1970.
- [RW09] R. T. Rockafellar, R. J.-B. Wets, *Variational Analysis*, Springer, 2009.
- [Qi93] L. Qi, *Convergence analysis of some algorithms for solving nonsmooth equations*, Math. Oper. Res. 18 (1993), 227–244.
- [QS93] L. Qi und J. Sun, *A nonsmooth version of Newton’s method*, Math. Programming 58 (1993), 353–367.

- [Sch94] S. Scholtes, *Introduction to piecewise differentiable equations*, Preprint No. 53/1994, Universität Karlsruhe, Institut f. Statistik u. Math. Wirtschaftstheorie, 1994.
- [SZ92] H. Schramm und J. Zowe, *A version of the bundle idea for minimizing a nonsmooth function: conceptual idea, convergence analysis, numerical results*, SIAM J. Optim. 2 (1992) 121–152.