Nichtglatte Optimierung und Anwendungen

Stefan Ulbrich

Fachbereich Mathematik

TU Darmstadt

Wintersemester 2009/2010

Inhaltsverzeichnis

1 Einführung **3**

1.1 Nichtglatte Optimierungsprobleme . . . 4

1.1.1 ℓ^∞-Approximation . . . 4

1.1.2 Dekomposition . . . 5

1.1.3 Minimax . . . 6

1.2 Nichtglatte Gleichungssysteme . . . 7

1.2.1 Das quadratische Penalty-Verfahren . . . 7

1.2.2 Reformulierung von Komplementaritätsproblemen . . . 8

2 Nichtglatte Optimierungsprobleme **9**

2.1 Beispiel . . . 9

2.2 Vorüberlegungen . . . 10

2.3 Das Subdifferential konvexer Funktionen . . . 13

2.4 Das Subgradienten-Verfahren . . . 23

2.4.1 Subgradienten-Verfahren bei unbekanntem Optimalwert . . . 26

2.4.2 Subgradienten-Verfahren bei bekanntem Optimalwert . . . 28

2.5 Schnittebenen-Verfahren . . . 32

2.6 Das ε-Subdifferential . . . 35

2.7 Bundle Methoden . . . 40

2.7.1 Das Bundle-Verfahren aus Sicht der Schnittebenenmethode . . . 41

2.7.2 Eine duale Interpretation des Bundle-Verfahrens . . . 44

2.7.3 Globale Konvergenz . . . 49

2.8 Anwendungen 56
2.8.1 Eigenwertoptimierung 56
2.8.2 SDP-Relaxierungen kombinatorischer Optimierungsprobleme . . . 57

3 Verfahren für Nichtglatte Gleichungssysteme 61
3.1 Beispiele 61
3.2 Ein allgemeines Newton-artiges Verfahren 61
3.2.1 Spezialfall: Das gewöhnliche Newton-Verfahren 65
3.3 Verallgemeinerte Differentiale 65
3.3.1 Einige wichtige Hilfsmittel 66
3.3.2 Einige Verallgemeinerte Differentiale 67
3.4 Semiglattheit 71
3.5 Semiglatte Newton-Verfahren 74

Literaturverzeichnis 76

Kapitel 1

Einführung

Diese Vorlesung gibt eine Einführung in Verfahren zur Lösung nichtlinearer Optimierungsprobleme sowie nichtlinearer Gleichungssysteme, die nichtdifferenzierbare Funktionen enthalten. Die behandelten Verfahren beruhen auf Resultaten der nichtglatten Analysis, die wir im erforderlichen Maße ebenfalls einführen.

Wir betrachten zwei Klassen nichtglatter Probleme:

Nichtglatte Optimierungsprobleme:

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

mit lokal Lipschitz-stetiger, aber i.a. nicht überall differenzierbarer Zielfunktion $f : X \to \mathbb{R}$ und abgeschlossenem zulässigem Bereich $X \subset \mathbb{R}^n$.

Wir werden überwiegend (aber nicht ausschließlich) konvexe Probleme (d.h. f und X konvex) betrachten und beschränken uns häufig auf den unrestringierten Fall $X = \mathbb{R}^n$.

Nichtglatte Gleichungssysteme:

$$F(x) = 0, \tag{1.2}$$

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

Bemerkung: Die Nichtdifferenzierbarkeit der Funktion f in (1.1) ist schwerwiegender als die Nichtdifferenzierbarkeit von F in (1.2). Denn ist z.B. im Fall $X = \mathbb{R}^n$ die Zielfunktion f in (1.1) sogar noch stetig differenzierbar, aber nicht zweimal differenzierbar, dann ist ∇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 ∇f lokal L-stetig ist) mit einem Problem vom Typ (1.2) zu tun.

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. 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 C^1, aber nicht C^2 ist (wichtiges Beispiel: Quadratische Penalty-Funktion).

1.1 Nichtglatte Optimierungsprobleme

Die Aufgabenstellung besteht darin, eine Funktion $f : X \to \mathbb{R}$ auf dem zulässigen Bereich $X \subset \mathbb{R}^n$ zu *minimieren*, d.h.

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

wobei f lokal Lipschitz-stetig, aber i.a. nicht differenzierbar sei. Wir wiederholen zunächst die Definition der lokalen Lipschitz-Stetigkeit:

Definition 1.1.1 Die Funktion $F : X \subset \mathbb{R}^n \to \mathbb{R}^m$ heißt *lokal Lipschitz-stetig in* $x \in X$, falls es Zahlen $L = L(x) > 0$ und $\delta = \delta(x) > 0$ gibt mit

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

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

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

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

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

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

Wir betrachten nun einige Beispiele für nichtglatte Optimierungsprobleme.

1.1.1 ℓ^∞-Approximation

Ein System $S : \mathbb{R}^m \to \mathbb{R}$ soll modelliert werden, das bei Eingabe der Daten y die Ausgabe $S(y)$ liefert. Hierbei ist die Funktion S nicht bekannt, es stehen jedoch eine große Zahl von Messwerte $z_i \approx S(y_i)$ für Eingabedaten $y_i \in \mathbb{R}^m$, $i = 1, \dots, l$, zur Verfügung.

Eine mathematische Modellierung liefere als Modell für S eine von dem Parametervektor $x \in \mathbb{R}^n$ abhängige Funktion

$$y \in \mathbb{R}^m \mapsto M(x, y) \in \mathbb{R}.$$

Zur Eichung des Modells M bestimmen wir x so, dass die maximale Abweichung zwischen Modell und Messdaten minimal ist:

$$\min_x f(x), \quad f(x) := \max_{1 \le i \le l} |M(x, y_i) - z_i|. \tag{1.3}$$

Die Funktion f ist offensichtlich nichtglatt.

Ist M stetig differenzierbar bezüglich x, so können wir (1.3) in ein restringiertes, glattes Optimierungsproblem umwandeln:

$$\min_{\alpha, x} \alpha \quad \text{u.d.N.} \quad -\alpha \le M(x, y_i) - z_i \le \alpha, \;\; i = 1, \ldots, l.$$

Bei großem l ergibt sich der Nachteil, dass das Problem sehr viele Ungleichungsnebenbedingungen hat. In diesem Fall ist es oft vorteilhaft, das nichtglatte Problem (1.3) zu lösen.

Ist $x \mapsto M(x, y)$ bereits nichtglatt, so ist in aller Regel ebenfalls (1.3) vorzuziehen.

1.1.2 Dekomposition

Häufig lassen sich in einem Optimierungsproblem die Unbekannten in zwei Klassen (x, y) gruppieren, so dass für festes x das Problem bezüglich y leicht zu lösen ist. Sei zum Beispiel das Problem

$$\min_{x,y} \; q(x, y) \quad \text{u.d.N.} \quad g(x, y) \le 0. \tag{1.4}$$

mit $x \in \mathbb{R}^n$, $y \in \mathbb{R}^l$ und glatten Funktionen $q : \mathbb{R}^n \times \mathbb{R}^l \to \mathbb{R}$, $g : \mathbb{R}^n \times \mathbb{R}^l \to \mathbb{R}^m$ gegeben. Wir nehmen an, dass für alle x das Problem

$$\min_y q(x, y) \quad \text{u.d.N.} \quad g(x, y) \le 0 \tag{1.5}$$

eine nichtleere Lösungsmenge $Y(x)$ besitzt und wesentlich einfacher zu lösen ist als (1.4). Zum Beispiel könnte (1.5) ein konvexes quadratisches Optimierungsproblem sein. Gleichzeitig könnte q und/oder g in sehr nichtlinearer Weise von x abhängen.

Den Optimalwert von (1.5) bezeichnen wir mit $f(x)$, d.h. wir haben

$$f(x) = q(x, y) \quad \forall\, y \in Y(x).$$

Hierdurch wird eine Funktion $f : \mathbb{R}^n \to \mathbb{R}$ definiert. Es ist nun häufig attraktiv, anstelle von (1.4) das Problem

$$\min_x f(x) \tag{1.6}$$

zu lösen. Ist dann x^* eine Lösung von (1.6) und $y^* \in Y(x^*)$ eine zugehörige Lösung von (1.5), so löst (x^*, y^*) das Ausgangsproblem (1.4).

Denn natürlich gilt $g(x^*, y^*) \leq 0$ und für alle x, y mit $g(x, y) \leq 0$ gilt

$$q(x, y) \geq \min_{\tilde{y}} \{q(x, \tilde{y}) \ : \ g(x, \tilde{y}) \leq 0\} = f(x) \geq f(x^*) = q(x^*, y^*).$$

Dieses Vorgehen hat jedoch einen Preis: Während (1.4) ein glattes nichtlineares Optimierungsproblem ist, haben wir es bei (1.6) mit einem im allgemeinen *nichtglatten* Problem zu tun.

Wir zeigen dies an dem Beispiel

$$\min_{x, y \in \mathbb{R}} \psi(x) + y^2 \quad \text{u.d.N.} \quad y - x \geq 0, \ y \geq 1.$$

Es gilt dann

$$\begin{aligned} f(x) &= \min_y \left\{\psi(x) + y^2 \ : \ y - x \geq 0, \ y \geq 1\right\} = \psi(x) + \min_y \left\{y^2 \ : \ y \geq \max\{x, 1\}\right\} \\ &= \psi(x) + \max\{x, 1\}^2. \end{aligned}$$

Die Funktion f hat also einen Knick bei $x = 1$.

1.1.3 Minimax

Viele Probleme der Praxis (Spiele, Robuste Optimierung unter Unsicherheit, Equilibriumsprobleme) 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 $\psi : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$.

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 $x = 0$.

Wie wir später zeigen werden, ist f konvex, falls $\psi(\cdot, y)$ konvex ist für alle $y \in Y$.

Weitere Beispiele werden später behandelt.

1.2 Nichtglatte Gleichungssysteme

Im zweiten Teil der Vorlesung behandeln wir Verfahren zur Lösung nichtglatter Gleichungssysteme

$$F(x) = 0$$

mit $F : \mathbb{R}^n \to \mathbb{R}^n$, aber i.a. nicht differenzierbar.

Die Bedeutung nichtglatter Gleichungssysteme beruht wesentlich auf der Tatsache, dass viele Ungleichungssysteme in äquivalente nichtglatte Gleichungssysteme überführt werden können.

Das Newton-Verfahren ist die wichtigste Basis für die Entwicklung schnell konvergenter Verfahren für nichtlineare Probleme, erfordert aber für q-superlineare Konvergenz die stetige Differenzierbarkeit von F. Wir werden jedoch sehen, dass auch für nichglatte Gleichungssysteme schnell lokal konvergente Verfahren existieren.

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

$$\min f(x) \quad \text{u.d.N.} \quad g(x) \le 0, \;\; h(x) = 0 \tag{1.7}$$

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

$$\min_{x \in \mathbb{R}^n} P_\alpha(x), \quad P_\alpha(x) = f(x) + \frac{\alpha}{2} \sum_{i=1}^m (g_i(x))_+^2 + \frac{\alpha}{2} \sum_{i=1}^p h_i(x)^2 \tag{1.8}$$

für eine Folge $(\alpha_k) \to \infty$ von α-Werten gelöst. Hierbei ist $(t)_+ = \max(t, 0)$ und $(t)_+^2 = ((t)_+)^2$. Zur Lösung des Problems (1.8) 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 (ist z.B. $g_i(x) = x$, dann gilt $(g_i(x))_+ \nabla g_i(x) = (x)_+$ und wir haben einen Knick bei $x = 0$). Das Newton-Verfahren ist also auf (1.8) nicht anwendbar, weil $P_\alpha(x)$ zwar C^1, aber nicht C^2 ist. Wir werden aber sehen, dass die in dieser Vorlesung entwickelten semiglatten Newtonverfahren trotzdem schnell konvergieren.

1.2.2 Reformulierung von Komplementaritätsproblemen

Ein Beispiel hierfür bildet die folgende wichtige Problemklasse:

Nichtlineares Komplementaritätsproblem NCP(F):

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

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

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

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

in der Form NCP(∇f) schreiben: in einem lokalen Minimum gelten die Kuhn–Tucker-Bedingungen

$$x_i \geq 0, \quad \nabla f_i(x) \geq 0, \quad x_i \nabla f_i(x) = 0, \quad i = 1, \ldots, n,$$

Die *grundlegende Idee* besteht nun darin, NCP(F) in folgender Form zu schreiben:

(1.9)
$$\Phi(x) := \begin{pmatrix} \phi(x_1, F_1(x)) \\ \vdots \\ \phi(x_n, F_n(x)) \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix},$$

wobei ϕ eine *NCP-Funktion* ist, d.h.

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

Beispiele für NCP-Funktionen:

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

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

Die Äquivalenz von NCP(F) und (1.9) ist offensichtlich. Sowohl ϕ_{FB} als auch ϕ_M sind Lipschitz-stetig, aber nicht überall differenzierbar. Bei (1.9) handelt es sich somit um ein nichtglattes Gleichungssystem der Form (1.2).

Kapitel 2

Nichtglatte Optimierungsprobleme

Wir betrachten in diesem Kapitel das folgende Optimierungsproblem:

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

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 beginnen mit einem weiteren Beispiel.

2.1 Beispiel

In der Einführung haben wir bereits drei Beispiele für nichtglatte Optimierungsprobleme kennengelernt (Dekomposition, ℓ^∞-Approximation und Minimax-Probleme). Wir geben noch ein Beispiel an.

Beispiel 2.1.1 (Exakte Penalisierung) Wir betrachten das Optimierungsproblem

$$\min \ f(x) \quad \text{u.d. N.} \quad g(x) \leq 0 \tag{2.2}$$

mit konvexen Funktionen $f, g_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \dots, m$. Dieses kann durch exakte Penalisierung in ein unrestringiertes, nichtglattes Problem überführt werden:

$$\min p_\mu(x), \quad p_\mu(x) := f(x) + \mu \sum_{i=1}^{m} \max\{g_i(x), 0\}. \tag{2.3}$$

Besitzt (2.2) eine Lösung, so stimmen unter geeigneten Voraussetzungen für hinreichend großen Penalty-Parameter $\mu > 0$ die Lösungsmengen von (2.2) und (2.3) überein.

Wir weisen dies für den Fall nach, daß f und g_i stetig differenzierbar sind und ein Punkt x_0 existiert mit $g_i(x_0) < 0$ für alle i (Slater-Bedingung). Dann ist $\bar{x}$ genau dann optimale Lösung von (2.2), wenn die Kuhn–Tucker-Bedingungen erfüllt sind:

$$\nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x) = 0, \quad \lambda_i \geq 0, \quad g_i(x) \leq 0, \quad \lambda_i g_i(x) = 0 \quad i = 1, \ldots, m.$$

Für zulässige Punkte x, d.h. $g(x) \leq 0$, gilt

$$p_\mu(x) = f(x).$$

Daher genügt es, zu zeigen, daß für hinreichend großes μ das Problem (2.3) keine Lösung mit $g(x) \not\leq 0$ besitzt. Sei hierzu $\bar{x}$ eine Lösung von (2.2) mit zugehörigen Lagrange-Multiplikatoren λ_i. Weiter gelte

$$\mu > \max_{i=1,\ldots,m} \lambda_i.$$

Sei nun x beliebig mit $g(x) \not\leq 0$. Für alle $i \in I = \{i : \lambda_i > 0\}$ gilt $g_i(\bar{x}) = 0$ und daher wegen der Konvexität von g_i:

$$\max\{g_i(x), 0\} \geq g_i(x) \geq \nabla g_i(\bar{x})^T (x - \bar{x}).$$

Damit erhalten wir:

$$\begin{aligned} p_\mu(x) &= f(x) + \mu \sum_{i=1}^{m} \max\{g_i(x), 0\} > f(x) + \sum_{i=1}^{m} \lambda_i \max\{g_i(x), 0\} \\ &\geq f(\bar{x}) + \nabla f(\bar{x})^T (x - \bar{x}) + \sum_{i \in I} \lambda_i \nabla g_i(\bar{x})^T (x - \bar{x}) \\ &= f(\bar{x}) + \left(\nabla f(\bar{x}) + \sum_{i=1}^{m} \lambda_i \nabla g_i(\bar{x}) \right)^T (x - \bar{x}) = f(\bar{x}) = p_\mu(\bar{x}). \end{aligned}$$

2.2 Vorüberlegungen

Ein naheliegender, aber wie wir sehen werden ungeeigneter, Ansatz für ein nichtglattes Optimierungsverfahren ist eine Verallgemeinerung des Verfahrens des steilsten Abstiegs.

Wir betrachten das Problem

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

und nehmen an, dass die Zielfunktion $f : \mathbb{R}^n \to \mathbb{R}$ auf ganz $\mathbb{R}^n$ richtungsdifferenzierbar und lokal Lipschitz-stetig ist. Wie wir sehen werden, trifft dies für konvexe Funktionen zu.

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

$$F'(x,s) := \lim_{t \to 0^+} \frac{F(x+ts) - F(x)}{t} \in \mathbb{R}^m$$

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

Bemerkung 2.2.2 Nicht jede Lipschitz-stetige Funktion ist in jedem Punkt richtungsdifferenzierbar. Wie wir aber gleich sehen werden, sind konvexe Funktionen auf offenen Mengen richtungsdifferenzierbar.

Die Richtungsableitung hat folgende Eigenschaften:

Lemma 2.2.3 *Die Funktion $F : U \subset \mathbb{R}^n \to \mathbb{R}^m$, $U \subset \mathbb{R}^n$ offen, sei richtungsdifferenzierbar im Punkt x. Dann gilt*

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

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 $\lambda > 0$ ergibt sich

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

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

$$\begin{aligned} \|F'(x,s) - F'(x,\hat{s})\| &= \lim_{t \to 0^+} \left\| \frac{F(x+ts) - F(x)}{t} - \frac{F(x+t\hat{s}) - F(x)}{t} \right\| \\ &= \lim_{t \to 0^+} \left\| \frac{F(x+ts) - F(x+t\hat{s})}{t} \right\| \leq L\|s - \hat{s}\|. \end{aligned}$$

□

Für eine reellwertigen Ziefunktion 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 2.2.4

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 2.2.5 a) Wegen der Lipschitz-Stetigkeit von $f'(x,\cdot)$, siehe Lemma 2.2.3, 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 1 (Verfahren des steilsten Abstiegs)

Wähle $x^0 \in \mathbb{R}^n$.

Für $k = 0, 1, 2, \ldots$:

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

Ist f differenzierbar, dann handelt es sich um das Gradientenverfahren mit exakter Schrittweitensuche, und man kann zeigen, dass jeder Häufungspunkt von x^k ein stationärer Punkt ist.

Wir machen uns nun klar, dass der Algorithmus für nichtglattes f versagen kann: Betrachte zunächst den Fall der konvexen glatten Zielfunktion

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

Man kann zeigen, dass das Verfahren zum Startpunkt $x^0 = (2,1)^T$ die Folge

$$x^{2k} = 9^{-k} x^0, \quad x^{2k+1} = 9^{-k} x^1,$$

erzeugt, die gegen das Minimum $(0,0)^T$ von f_1 konvergiert (siehe Übung).

Man kann weiter folgendes zeigen (siehe Übung):

a) Das Verfahren produziert für die nichtglatte konvexe Zielfunktion

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

dieselbe Folge x^k, konvergiert also gegen das Minimum von f_2.

b) Das Verfahren erzeugt auch für die nichtglatte konvexe Zielfunktion

$$f_3(x) := \begin{cases} f_2(x) & x_1 \geq |x_2|, \\ \frac{1}{\sqrt{3}}(x_1 + 2|x_2|) & x_1 < |x_2|, \end{cases}$$

dieselbe Folge x^k, aber x^k konvergiert nicht gegen ein Minimum von f_3, denn f_3 fällt entlang des Strahls $\left\{(-\sigma, 0)^T \;:\; \sigma \geq 0\right\}$ unbeschränkt.

Das Verfahren des steilsten Abstiegs mit exakter Schrittweitensuche ist also ungeeignet für nichtglatte Probleme! Wir werden das Verfahren später geeignet modifizieren.

2.3 Das Subdifferential konvexer Funktionen

Wir werden in diesem Abschnitt eine verallgemeinerte Ableitung für konvexe Funktionen einführen. Diese wird die Basis von Verfahren für nichtglatte konvexe Minimierungsprobleme

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

bilden, wobei $X \subset \mathbb{R}^n$ konvex und $f : X \to \mathbb{R}$ konvex, aber im Allgemeinen nicht differenzierbar.

Wir behandlen zunächst einige wichtige Eigenschaften konvexer Funktionen.

Satz 2.3.1 *Sei $f : X \to \mathbb{R}$ 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*

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

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 Quader

$$S := \{y \in \mathbb{R}^n \;:\; |y_i - \bar{x}_i| \leq 2\delta\}$$

mit Seitenlänge 4δ und Mittelpunkt $\bar{x}$ eine Teilmenge von X ist. Wir zeigen zunächst, dass Zahlen $m, M \in \mathbb{R}$ existieren mit

$$m \le f(x) \le M \quad \forall\, x \in S.$$

Seien $v_1, \ldots, v_p$, $p = 2^n$, die Ecken von S. Dann ist S die konvexe Hülle seiner Ecken, kurz $S = \text{conv}(v_1, \ldots, v_p)$, und somit ist jedes $x \in S$ eine Konvexkombination

$$x = \sum_{i=1}^{p} \lambda_i v_i, \quad \lambda_i \ge 0, \quad \sum_{i=1}^{p} \lambda_i = 1$$

der Ecken. Die Konvexität von f liefert

$$f(x) = f\left(\sum_{i=1}^{p} \lambda_i v_i\right) \le \sum_{i=1}^{p} \lambda_i f(v_i) \le \max_{i=1,\ldots,p} f(v_i) =: M.$$

Um eine untere Schranke zu gewinnen, sei $x \in S$ beliebig und z die Spiegelung von x an $\bar{x}$, also

$$z = x + 2(\bar{x} - x) = 2\bar{x} - x.$$

Dann ist $z \in S$ und es gilt

$$\bar{x} = \frac{1}{2}x + \frac{1}{2}z,$$

also mit der Konvexität von f

$$f(\bar{x}) \le \frac{1}{2}f(x) + \frac{1}{2}f(z).$$

Dies ergibt

$$f(x) \ge 2f(\bar{x}) - f(z) \ge 2f(\bar{x}) - M =: m$$

für alle $x \in S$.

Zum Nachweis der lokalen Lipschitz-Stetigkeit zeigen wir nun die Ungleichung

$$|f(x) - f(y)| \le \frac{M - m}{\delta}\|x - y\| \quad forall\, x, y \in B_\delta(\bar{x}). \tag{2.4}$$

Seien also $x, y \in B_\delta(\bar{x})$ beliebig und definiere z zu

$$z = y + \delta \frac{y - x}{\|y - x\|}.$$

Wegen $y \in B_\delta(\bar{x})$ gilt dann $z \in S$ und

$$y = \frac{\|y - x\|}{\delta + \|y - x\|}z + \frac{\delta}{\delta + \|y - x\|}x = \left(1 - \frac{\|y - x\|}{\delta + \|y - x\|}\right)x + \frac{\|y - x\|}{\delta + \|y - x\|}z.$$

Die Konvexität von f ergibt nun

$$f(y) \leq \left(1 - \frac{\|y-x\|}{\delta + \|y-x\|}\right) f(x) + \frac{\|y-x\|}{\delta + \|y-x\|} f(z),$$

also

$$f(y) - f(x) \leq \frac{\|y-x\|}{\delta + \|y-x\|}(f(z) - f(x)) \leq \frac{M-m}{\delta}\|y-x\|.$$

Vertauschen wir die Rollen von x und y, dann ergibt sich

$$f(x) - f(y) \leq \frac{M-m}{\delta}\|x-y\|$$

und somit insgesamt (2.4).

Da $\bar{x} \in X$ beliebig war, folgt aus (2.4) die lokale Lipschitz-Stetigkeit von f auf X.

zu b):

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

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

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

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

für kleine $t > 0$. Daher ergibt sich

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

und das Infimum existiert.

zu c):

Für $x, y \in X$ und $s = y - x$ gilt

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

wobei Teile des Beweises von b) benutzt wurden. □

Bemerkung 2.3.2 Ist X abgeschlossen, dann muss eine konvexe Funktion $f : X \to \mathbb{R}$ nicht einmal stetig sein: Zum Besipiel ist $f : [0,1] \to \mathbb{R}$, $f(0) = 1$, $f(x) = 0$, $x > 0$ konvex, aber nicht stetig.

In der konvexen Analysis kommt dem Subdifferential die tragende Rolle zu, die im differenzierbaren Fall der Gradient spielt:

Definition 2.3.3 (Subgradient, Subdifferential) Sei $f : X \to \mathbb{R}$ 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:

$$f(y) - f(x) \geq g^T(y - x) \quad \forall\, y \in X. \tag{2.5}$$

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 graph(l) stützt graph(f) in $\bar{x}$ von unten.

Wir weisen nun einige wichtige Eigenschaften des Subdifferentials nach.

Satz 2.3.4 *Sei $f : X \to \mathbb{R}$ konvex auf der nichtleeren offenen konvexen Menge $X \subset \mathbb{R}^n$. Weiter sei $x \in X$. Dann gilt:*

a) $\partial f(x) = \left\{g \in \mathbb{R}^n \;:\; g^T s \leq f'(x, s) \;\; \forall\, s \in \mathbb{R}^n\right\}$.

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

d) *f ist nicht nur richtungsdifferenzierbar, sondern sogar Bouligand-differenzierbar, d.h.*

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

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

Für einen Teil des Beweises benötigen wir den Trennungssatz (siehe Satz A.1 Anhang)

Satz 2.3.5 (Trennungssatz) *Seien $X_1, X_2 \subset \mathbb{R}^n$ nichtleer, disjunkt und konvex. Dann sind X_1 und X_2 durch eine Hyperebene trennbar:*

Es gibt $v \in \mathbb{R}^n$, $v \neq 0$, mit

$$v^T x_1 \leq v^T x_2 \quad \forall\, x_1 \in X_1,\ x_2 \in X_2.$$

Bemerkung 2.3.6 Der Vektor v ist die Normale der trennenden Hyperebene $\{x\ :\ v^T x = \alpha\}$ mit (z.B.) $\alpha = \inf\limits_{x \in X_2} v^T x$. Sie zeigt in den Halbraum, in dem X_2 enthalten ist.

Beweis: (von Satz 2.3.4)

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 (2.5)

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

Daraus folgt

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

und der Grenzübergang $t \to 0^+$ liefert

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

"$K \subset \partial f(x)$":

Sei $g \in K$. Nach Satz 2.3.1 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 $g \in \partial f(x)$, da X offen ist.

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 gilt dann auch für den Durchschnitt

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

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

Sei $g \in \partial f(x)$ beliebig. Ist nun L die Lipschitz-Konstante von f nahe x, dann gilt

$$|f'(x,g)| = \lim_{t \to 0^+} \frac{|f(x+tg) - f(x)|}{t} \le \lim_{t \to 0^+} \frac{L\|tg\|}{t} = L\|g\|.$$

Wir wissen aus a)

$$\|g\|^2 = g^T g \le f'(x,g) \le L\|g\|,$$

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

$$K_1 = \left\{ \begin{pmatrix} y \\ z \end{pmatrix} \;:\; y \in X,\; z \in \mathbb{R},\; f(y) < z \right\},$$

$$K_2 = \left\{ \begin{pmatrix} y \\ z \end{pmatrix} \;:\; y \in X,\; y = x + ts,\; z = f(x) + tf'(x,s),\; t \ge 0 \right\}$$

Beide Mengen sind nichtleer und konvex. Sie sind auch disjunkt, da für $\binom{y}{z} \in K_2$ gilt:

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

Daher lassen sich K_1 und K_2 nach Satz 2.3.5 durch eine Hyperebene trennen, d.h. es gibt $\binom{w}{\alpha} \in \mathbb{R}^{n+1} \setminus \{0\}$ mit

$$w^T y_2 + \alpha z_2 \le w^T y_1 + \alpha z_1 \quad \forall \begin{pmatrix} y_1 \\ z_1 \end{pmatrix} \in K_1,\; \begin{pmatrix} y_2 \\ z_2 \end{pmatrix} \in K_2.$$

Aus Stetigkeitsgründen gilt dann auch

$$w^T y_2 + \alpha z_2 \le w^T y_1 + \alpha z_1 \quad \forall \begin{pmatrix} y_1 \\ z_1 \end{pmatrix} \in \mathrm{epi}(f),\; \begin{pmatrix} y_2 \\ z_2 \end{pmatrix} \in K_2, \tag{2.6}$$

wobei

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

der Epigraph von f ist.

Wir zeigen nun: $\alpha > 0$ und $-w/\alpha \in \partial f(x)$.

Wegen $\binom{x}{f(x)+1} \in \mathrm{epi}(f)$ und $\binom{x}{f(x)} \in K_2$ folgt zunächst $\alpha \ge 0$.

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 $\binom{y}{f(y)} \in \mathrm{epi}(f)$ sowie $\binom{x}{f(x)} \in K_2$ und daher (wir benutzen $\alpha = 0$)

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

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

Durchmultiplizieren von (2.6) mit $1/\alpha$ ergibt für $g = -w/\alpha$:

$$-g^T y_2 + z_2 \le -g^T y_1 + z_1 \quad \forall \begin{pmatrix} y_1 \\ z_1 \end{pmatrix} \in \text{epi}(f),\ \begin{pmatrix} y_2 \\ z_2 \end{pmatrix} \in K_2, \tag{2.7}$$

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

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

und somit ist g ein Subgradient.

Zum Nachweis von c) nutzen wir, dass wegen (2.7) für kleine $t > 0$ gilt (X ist offen)

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

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

zu d):

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

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

Durch Übergang auf eine Teilfolge erzielen wir $v_k \to v$. 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:

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

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 $s \in \mathbb{R}^n$:

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

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

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

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:

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

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

Weiter gilt die folgende notwendige und hinreichende Optimalitätsbedingung:

Satz 2.3.7 *Sei $X \subset \mathbb{R}^n$ offen und konvex und $f : X \to \mathbb{R}$ sei eine konvexe Funktion. Weiter sei $\bar{x} \in \mathbb{R}^n$. Dann sind die folgenden Aussagen äquivalent:*

a) *Der Punkt $\bar{x}$ ist (globales = lokales) Minimum von f auf X.*

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

c) *Es gilt* $\quad 0 \in \partial f(\bar{x})$.

Beweis:

a) $\Longrightarrow$ b):

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

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

b) $\Longrightarrow$ c):

Wegen

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

folgt $0 \in \partial f(\bar{x})$ gemäß Satz 2.3.4 a).

c) $\Longrightarrow$ a):

Da 0 ein Subgradient ist, folgt nach Definition

$$f(x) - f(\bar{x}) \geq 0^T (x - \bar{x}) = 0 \quad \forall\, x \in X.$$

Somit ist $\bar{x}$ globales Minimum von f auf X. $\square$

Man beachte, dass X in Satz 2.3.7 als offen vorausgesetzt wurde. Wir können Satz 2.3.4 c) und Satz 2.3.7 zu einer Charakterisierung der Richtungen des steilsten Abstiegs nutzen:

Satz 2.3.8 (Richtung des steilsten Abstiegs) *Sei $f : X \to \mathbb{R}$ konvex auf der offenen und konvexen Menge $X \subset \mathbb{R}^n$. Im Punkt $x \in X$ gelte $0 \notin \partial f(x)$. Ist nun d die Lösung des Problems*

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

dann ist $s = -d$ 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:

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

Daraus folgt

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

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. $\square$

Wir sehen, dass zur Berechnung der Richtung des steilsten Abstiegs die Kenntnis des gesamten Subdifferentials erforderlich ist.

Wir geben zum Abschluss einige nützliche Rechenregeln an:

Satz 2.3.9 *Seien die Funktionen $f_i : X \to \mathbb{R}$ konvex auf der offenen konvexen Menge $X \subset \mathbb{R}^n$. Dann ist die Funktion*

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

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

$$f'(x, s) = \max_{i\in I(x)} f_i'(x, s), \tag{2.8}$$

$$\partial f(x) = \textit{conv}\left(\cup_{i\in I(x)} \partial f_i(x)\right). \tag{2.9}$$

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

Beweis: (Für Interessierte)
Die Konvexität folgt aus Optimierung 1 (oder einfache Übung). Nach Blatt 1, H2 gilt

$$f'(x,s) = \max_{i \in I(x)} f_i'(x,s).$$

Sei $K = \text{conv}\left(\cup_{i \in I(x)} \partial f_i(x)\right)$.

"$K \subset \partial f(x)$":

Für $i \in I(x)$ und $g \in \partial f_i(x)$ gilt wegen Satz 2.3.4 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 $\cup_{i \in I(x)} \partial f_i(x) \subset \partial f(x)$ 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 (Satz A.2 Anhang) 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 $i \in I(x)$:

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

Daraus wiederum folgt

$$f'(x,v) < \alpha < g^T v,$$

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

Abschließend geben wir noch ohne Beweis ein paar Rechenregeln an:

Satz 2.3.10 a) *Sei $f : \mathbb{R}^m \to \mathbb{R}$ konvex, $A \in \mathbb{R}^{m \times n}$ und $b \in \mathbb{R}^m$. Dann ist die Funktion $g : \mathbb{R}^n \ni x \to f(Ax+b)$ konvex und es gilt:*

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

b) *Seien die Funktionen $f_1, \ldots, f_m : \mathbb{R}^n \to \mathbb{R}$ konvex. Weiter sei $g : \mathbb{R}^m \to \mathbb{R}$ konvex und es gelte*

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

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:

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

Beweis: Siehe [HU93a], Thm. VI.4.2.1 und VI.4.3.1.

Aus Satz 2.3.7 kann man den folgenden Mittelwertsatz ableiten:

Satz 2.3.11 *Sei $f : U \to \mathbb{R}$ 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: (Für Interessierte)
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 ≤ 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 ϕ ein lokales Minimum in einem $t^* \in (0, 1)$ an. Dort gilt $0 \in \partial\phi(t^*)$ nach Satz 2.3.7 und wir können daher $z = x + t^* s$ wählen. □

2.4 Das Subgradienten-Verfahren

Wir haben gesehen, dass das Verfahren des steilsten Abstiegs mit exakter Schrittweitensuche für nichtglatte Probleme nicht immer konvergent ist. Zudem 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 2.4.1 Sei $f(x) = \frac{x_1^2}{2} + 2|x_2|$. Betrachte den Punkt $x = (1, 0)^T$. Nach Blatt 1, H1 ist dann $g = (1, 2)^T \in \partial f(x)$, aber $s = -g$ ist keine Abstiegsrichtung, da für jedes $t > 0$ gilt

$$f(x - tg) > f(x).$$

□

Aber Subgradienten haben, wie wir sehen werden, eine andere wertvolle Eigenschaft: Die Richtung $s = -g$ ist eine Abstiegsrichtung für $\|x - \bar{x}\|$, falls $f(x) > f(\bar{x})$ gilt. Dies werden wir ausnutzen.

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

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

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

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

Bemerkung: Man kann zeigen, dass P_X nichtexpansiv ist, also

$$\|P_X(x) - P_X(y)\| \le \|x - y\| \quad \forall\, x, y \in \mathbb{R}^n.$$

□

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

Algorithmus 2 (Projiziertes Subgradienten-Verfahren)

Wähle $x^0 \in X$.

Für $k = 0, 1, 2, \ldots$:

1. Bestimme einen Subgradienten $g^k \in \partial f(x^k)$.
2. Falls $g^k = 0$ oder ein geeignetes Abbruchkriterium erfüllt ist, STOP mit Ergebnis x^k.
3. Setze $s^k = -g^k / \|g^k\|$ und
$$x^{k+1} = P_X(x^k + \sigma_k s^k).$$
mit einer Schrittweite $\sigma_k > 0$.

Wir wissen aus §2.2, dass es keinen Sinn macht, stets Abstieg, d.h. $f(x^{k+1}) < f(x^k)$, erzielen zu wollen. Vielmehr zeigt der folgende Satz, dass wir stattdessen versuchen sollten, $\sigma_k > 0$ so zu wählen, dass gilt:

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

Dies ist tatsächlich möglich:

Satz 2.4.1 *Sei $\bar{x}$ eine Lösung von (PX). Weiter sei $x \in X$ und $g \in \partial f(x)$. Dann gilt im Fall $g \neq 0$ mit $s = -\frac{g}{\|g\|}$:*

$$\|P_X(x+\sigma s) - \bar{x}\| \leq \|x + \sigma s - \bar{x}\| < \|x - \bar{x}\| \quad \forall\, 0 < \sigma < \frac{2(f(x) - f(\bar{x}))}{\|g\|}. \tag{2.10}$$

Genauer haben wir:

$$\begin{aligned} &\|P_X(x+\sigma s) - \bar{x}\|^2 \leq \|x + \sigma s - \bar{x}\|^2 \leq \|x - \bar{x}\|^2 - \rho(\sigma) \\ &\textit{mit} \quad \rho(\sigma) = 2\sigma \frac{f(x) - f(\bar{x})}{\|g\|} - \sigma^2. \end{aligned} \tag{2.11}$$

Beweis: Wir verwenden die Ungleichung für Subgradienten (Satz 2.3.4 a)):

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

Der Term $\rho(\sigma)$ ist positiv genau für die in (2.10) angegebenen Werte von σ.

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

$$\|P_X(x - \sigma\frac{g}{\|g\|}) - \bar{x}\| = \|P_X(x - \sigma\frac{g}{\|g\|}) - P_X(\bar{x})\| \leq \|x - \sigma\frac{g}{\|g\|} - \bar{x}\|.$$

□

Nach Satz 2.4.1 gilt also

$$\left\|x^{k+1} - \bar{x}\right\|^2 \leq \left\|x^k - \bar{x}\right\|^2 - \rho_k(\sigma_k)$$

mit

$$\rho_k(\sigma) = 2\sigma \frac{f(x^k) - f(\bar{x})}{\|g^k\|} - \sigma^2.$$

Um $\bar{x}$ möglichst nahe zu kommen, wäre es am besten, σ_k als Maximum von $\rho_k(\sigma)$ zu wählen. Dies führt auf die Wahl

$$\rho_k'(\sigma_k) = 2\frac{f(x^k) - f(\bar{x})}{\|g^k\|} - 2\sigma_k = 0,$$

also

$$\sigma_k = \frac{f(x^k) - f(\bar{x})}{\|g^k\|}. \tag{2.12}$$

Ist $f(\bar{x})$ nicht bekannt, dann ist diese Schrittweitenwahl leider nicht möglich.

2.4.1 Subgradienten-Verfahren bei unbekanntem Optimalwert

Wir betrachten zunächst den Fall, dass der Optimalwert $f(\bar{x})$ von (PX) unbekannt ist. Dann ist die Wahl (2.12) nicht möglich und wir wissen nicht, was eine gute Wahl von σ_k ist. Es gibt hier viele, oft heuristische Verfahrensvarianten, diesem Problem beizukommen.

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

$$\sigma_k \to 0, \quad \sum_{k=0}^{\infty} \sigma_k = \infty. \tag{2.13}$$

Die zweite Bedingung ist deshalb sinnvoll, da gilt

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

Wäre also (σ_k) summierbar, dann wäre $\|x^k - x^0\| \le M$ mit einer Konstante $M > 0$ und das Minimum könnte unter Umständen nicht erreicht werden.

Wir weisen nun die Konvergenz des Verfahrens nach:

Satz 2.4.2 *Die Funktion $f : U \to \mathbb{R}$ sei konvex auf der offenen konvexen Menge $U \subset \mathbb{R}^n$. Die Menge $X \subset U$ sei abgeschlossen und konvex und $\bar{x} \in X$ sei eine Lösung von (PX). Algorithmus 2 verwende Schrittweiten, die* (2.13) *genügen. Dann terminiert das Verfahren entweder mit einer Lösung x^k von (PX), oder er erzeugt eine Folge, für die gilt:*

a)

$$\liminf_{k\to\infty} f(x^k) = f(\bar{x}). \tag{2.14}$$

b) *Gilt*

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

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

Bemerkung 2.4.3 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, (2.14) gilt nicht. Dann gibt es $\varepsilon > 0$ und $K > 0$ mit

$$f(x^k) > f(\bar{x}) + \varepsilon \quad \forall\, k \ge K.$$

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

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

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

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

Wegen $\sigma_k \to 0$ gilt nach eventuellem Vergrößern von K:

$$\|x^{k+1} - \bar{x}\|^2 - \|x^k - \bar{x}\|^2 \leq \sigma_k^2 - 2\delta\sigma_k \leq -\delta\sigma_k.$$

Dies ergibt durch Aufsummieren

$$-\|x^K - \bar{x}\|^2 \leq \|x^{K+l} - \bar{x}\|^2 - \|x^K - \bar{x}\|^2 \leq \sum_{k=K}^{K+l-1} (-\delta\sigma_k) \to -\infty \quad (l \to \infty),$$

also einen Widerspruch.

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

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

□

Leider ist die Konvergenzgeschwindigkeit recht unbefriedigend:

Satz 2.4.4 *Wir betrachten den unrestringierten Fall $X = \mathbb{R}^n$ (d.h. ohne Projektion) unter den Voraussetzungen von Satz 2.4.2. Die Folge (x^k) konvergiere gegen den Punkt $\bar{x}$. Dann ist die Konvergenzgeschwindigkeit schlechter als r-linear:*

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

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

$$\|x^k - \bar{x}\| \leq C\gamma^k \quad \forall\, k \geq K.$$

Dies zeigt für alle $k \geq K$

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

Damit ergibt sich aus (2.13):

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

Dies ist ein Widerspruch. $\square$

Bessere Konvergenzgeschwindigkeiten ergeben sich, falls der Optimalwert bekannt ist, da dann die "optimale" Schrittweite (2.12) verwendet werden kann.

2.4.2 Subgradienten-Verfahren bei bekanntem Optimalwert

Wir betrachten nun den Fall, dass der optimale Zielfunktionswert $f(\bar{x})$ von (PX) bekannt ist. Dies kann durchaus auftreten:

Beispiel 2.4.2 Gesucht sei ein zulässiger Punkt bezüglich der Menge

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

Hierzu können wir die Funktion

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

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

Beispiel 2.4.3 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 (PX) 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.

Wir hatten bereits beobachtet, dass es wegen der nach Satz 2.4.1 gültigen Abschätzung

$$\|x^{k+1}-\bar{x}\|^2 \le \|x^k-\bar{x}\|^2 - \rho_k(\sigma_k)$$

mit

$$\rho_k(\sigma) = 2\sigma\frac{f(x^k)-f(\bar{x})}{\|g^k\|} - \sigma^2$$

zu wählen. Wir hatten gesehen, dass dies auf die Wahl führt

$$\sigma_k = \frac{f(x^k)-f(\bar{x})}{\|g^k\|}. \tag{2.12}$$

Wir betrachten nun das Subgradientenverfahren mit der Schrittweitenwahl (2.12). Dann gilt

$$\rho_k(\sigma_k) = \frac{(f(x^k)-f(\bar{x}))^2}{\|g^k\|^2} \tag{2.15}$$

und

$$\|x^{k+1}-\bar{x}\|^2 \le \|x^k-\bar{x}\|^2 - \rho_k(\sigma_k) = \|x^k-\bar{x}\|^2 - \frac{(f(x^k)-f(\bar{x}))^2}{\|g^k\|^2}. \tag{2.16}$$

Wir erhalten das folgende Konvergenzresultat:

Satz 2.4.5 *Die Funktion $f : U \to \mathbb{R}$ sei konvex auf der offenen konvexen Menge $U \subset \mathbb{R}^n$. Die Menge $X \subset U$ sei abgeschlossen und konvex und $\bar{x} \in X$ sei eine Lösung von (PX). Die Folge $(x^k) \subset X$ werde erzeugt durch Algorithmus 2, wobei die Schrittweiten (σ_k) gemäß* (2.12) *gewählt werden, und wir abbrechen, sobald $f(x^k) = f(\bar{x})$ gilt. Der Algorithmus erzeuge unendlich viele Iterierte. Dann gilt:*

a) *Die Folge $(\|g^k\|)$ ist beschränkt.*

b) $\lim_{k\to\infty} f(x^k) = f(\bar{x})$.

c) *Die Folge (x^k) konvergiert gegen eine Lösung von (PX).*

Beweis: Sei $R = \|x^0-\bar{x}\|$. Induktiv ergibt sich aus (2.16) für alle $k \ge 0$:

$$\begin{aligned}\|x^{k+1}-\bar{x}\|^2 &\le \|x^k-\bar{x}\|^2 - \rho_k(\sigma_k) \le \cdots \le \|x^0-\bar{x}\|^2 - \sum_{l=0}^{k}\rho_l(\sigma_l)\\ &= R^2 - \sum_{l=0}^{k}\rho_l(\sigma_l) < R^2.\end{aligned} \tag{2.17}$$

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

zu a): Wir benötigen nun folgendes Hilfsresultat:

Lemma 2.4.6 *Sei $F : X \to \mathbb{R}^m$ lokal Lipschitz-stetig auf der abgeschlossenen konvexen Menge $X \subset \mathbb{R}^n$. Dann ist F Lipschitz-stetig auf jedem Kompaktum $K \subset X$.*

Da f als konvexe Funktion nach Satz 2.3.1 lokal Lipschitz-stetig auf U und somit auf X ist, ist also f auf K Lipschitz-stetig (Konstante L) und nach Satz 2.3.4 b) gilt

$$\|g^k\| \le L \quad \forall\, k.$$

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

zu b): Wir erhalten mit (2.17)

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

also folgt wegen $\rho_k(\sigma_k) > 0$, dass $(\rho_k(\sigma_k))$ eine Nullfolge ist. Nun ergibt sich aus (2.15):

$$0 < f(x^k) - f(\bar{x}) = \sqrt{\rho_k(\sigma_k)}\|g^k\| \le C\sqrt{\rho_k(\sigma_k)} \to 0.$$

zu c):

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

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

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

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

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

$$\|\tilde{x} - \bar{x}\| = \alpha = \|\bar{x} - \bar{x}\|.$$

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

Beweis von Lemma 2.4.6: (Für Interessierte)

Es genügt, die Aussage für alle konvexen Kompakta der Form $K = \overline{B}_R \cap X$, $R > 0$, zu zeigen (denn jedes Kompaktum in X lässt sich in solch eine Menge „packen"). Zu jedem Punkt $x \in K$ existieren $\delta(x) > 0$ und $L(x) > 0$, so dass F auf $B_{\delta(x)}(x)$ Lipschitz-stetig ist mit Konstante $L(x)$. Nun gibt es eine endliche Teilüberdeckung $B_{\delta(x_i)}(x_i)$, $i = 1, \dots, r$, des Kompaktums K. Sei $L = \max_{1 \le i \le r} L(x_i)$. Für beliebige $x, y \in K$ ist $[x, y] \subset K$ und es gibt daher $q + 1$ Punkte

$$x_j = (1 - t_j)x + t_j y, \quad 0 = t_0 < t_1 < \dots < t_q = 1, \quad q \le r,$$
$$\text{mit} \quad x_j, x_{j+1} \in B_{\delta(x_i)}(x_i) \quad \text{für mindestens ein } i, \quad j = 0, \dots, q-1.$$

Nun gilt:

$$\|F(y)-F(x)\| \leq \sum_{j=0}^{q-1} \|F(x_{j+1})-F(x_j)\| \leq L \sum_{j=0}^{q-1} \|x_{j+1}-x_j\| = L\|y-x\|.$$

□

Wir wenden uns nun der Konvergenzgeschwindigkeit zu:

Satz 2.4.7 *Unter den Voraussetzungen von Satz 2.4.5 gilt:*

a) $\liminf_{k\to\infty} \sqrt{k}\big(f(x^k)-f(\bar{x})\big) = 0.$

b) *Gibt es ein $\gamma > 0$ mit*

$$f(x^k)-f(\bar{x}) \geq \gamma\|x^k-\bar{x}\| \quad \forall\, k,$$

dann konvergiert x^k q-linear gegen $\bar{x}$:

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

Bemerkung 2.4.8 Natürlich muss man in Teil b) für $\bar{x}$ den Grenzwert von (x^k) wählen. Dieser existiert nach Satz 2.4.5 c) und ist ein globales Minimum von f.

Beweis: Wir setzen auf dem Beweis von Satz 2.4.5 auf.

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

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

Nun ergibt sich aus (2.15) und (2.17)

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

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

zu b):

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

□

2.5 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

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

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

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

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:

$$\min_x f_k^{\text{se}}(x) \quad \text{u.d.N.} \quad x \in X. \tag{2.18}$$

Natürlich ist dies nur sinnvoll, wenn das Problem (2.18) (es ist selbst ein nichtglattes Optimierungsproblem) wesentlich einfacher lösbar ist als das Ausgangsproblem (PX). Ob dies so ist, hängt von der Menge X ab. In jedem Fall können wir (2.18) als glattes Problem schreiben (siehe Lemma 2.5.1):

$$\min_{x \in \mathbb{R}^n, \alpha \in \mathbb{R}} \alpha \quad \text{u.d.N.} \quad x \in X, \ \ f^j + {g^j}^T(x - x^j) - \alpha \leq 0, \ \ 0 \leq j \leq k. \tag{2.19}$$

Im Fall $X = \mathbb{R}^n$ oder $X = \{x : Ax = b, \ x \geq 0\}$ oder $X = \{x : Ax \leq b\}$ ist (2.19) ein lineares Optimierungsproblem und daher relativ leicht zu lösen.

Lemma 2.5.1 *Gegeben seien Funktionen $f_i : X \to \mathbb{R}$, $i = 1, \dots, m$, mit $X \subset \mathbb{R}^n$. Wir betrachten das (nichtglatte) Problem*

$$\min_{x \in X} f(x) \quad \text{mit} \quad f(x) = \max_{1 \leq i \leq m} f_i(x) \tag{2.20}$$

und parallel dazu das Problem

$$\min_{x \in \mathbb{R}^n, \alpha \in \mathbb{R}} \alpha \quad \text{u.d.N.} \quad x \in X, \quad f_i(x) \leq \alpha, \ \ i = 1, \dots, m. \tag{2.21}$$

Dann gilt:

a) *Ist $\bar{x}$ eine globale Lösung von* (2.20)*, so ist* $(\bar{x}, f(\bar{x}))$ *eine globale Lösung von* (2.21).

b) *Ist* $(\bar{x}, \alpha^*)$ *eine globale Lösung von* (2.21)*, so ist $\bar{x}$ eine globale Lösung von* (2.20) *und es gilt* $f(\bar{x}) = \alpha^*$.

Beweis: Sei $\bar{x}$ Lösung von (2.20) und $\alpha^* = f(\bar{x})$. Dann ist natürlich $(\bar{x}, \alpha^*)$ zulässig für (2.21). Ist nun (x, α) zulässig für (2.21), so gilt $x \in X$ und daher

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

Damit ist $(\bar{x}, \alpha^*)$ optimale Lösung von (2.21).

Sei nun $(\bar{x}, \alpha^*)$ optimale Lösung von (2.21). Dann gilt $\alpha^* = f(\bar{x})$, denn sonst wäre $(\bar{x}, f(\bar{x}))$ ein zulässiger Punkt mit kleinerem Zielfunktionswert.

Ist nun $x \in X$ beliebig, so ist $(x, f(x))$ zulässig für (2.21) und daher folgt

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

so dass $\bar{x}$ optimale Lösung von (2.20) ist. $\square$

Der Name Schnittebenen-Verfahren leitet sich von den Teilproblemen (2.19) ab:

Jede der linearen Ungleichungen

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

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

Der zulässige Bereich von (2.19) wird also in jeder Iteration kleiner, weil eine neue Schnittebene hinzukommt. Wie wir in Satz 2.5.3 sehen werden, ist der Punkt x^k optimale Lösung von (PX), falls die Hinzunahme der k-ten Schnittebene die Lösungsmenge von (2.19) nicht ändert.

Wir betrachten nun das folgende Verfahren:

Algorithmus 3 (Schnittebenen-Verfahren)

0. Wähle $x^0 \in X$.

Für $k = 0, 1, 2, \ldots$:

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 (2.18) (oder, äquivalent dazu, von (2.19)).

Satz 2.5.2 *Algorithmus 3 erzeuge eine unendliche Folge (x^k). Dann ist jeder Häufungspunkt von (x^k) eine optimale Lösung von (PX).*

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

$$\begin{aligned} f(x) &\geq f^{\text{se}}_{k-1}(x) \geq f^{\text{se}}_{k-1}(x^k) \geq f^j + {g^j}^T(x^k - x^j) \\ &\overset{K\ni k\to\infty}{\longrightarrow} f^j + {g^j}^T(\bar{x} - x^j) \overset{K\ni j\to\infty}{\longrightarrow} f(\bar{x}). \end{aligned}$$

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

Aus $f(x) \geq f(\bar{x})$ für alle $x \in X$ folgt, dass $\bar{x}$ globale Lösung von (PX) ist. □

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

Satz 2.5.3 *Algorithmus 3 erzeuge die Iterierten x^k. Ist dann $\bar{x}$ optimale Lösung von (PX) und gilt*

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

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

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

Beweis: Es gilt

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

□

Wir können also in Schritt 2 das folgende Abbruchkriterium verwenden:

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

Bemerkung 2.5.4 Definitionsgemäß ist x^{k+1} optimale Lösung von (2.18). Tritt der Fall ein, dass die alte Iterierte x^k bereits Lösung von (2.18) ist, dann kann $x^{k+1} = x^k$ gewählt werden und wir erhalten mit Satz 2.5.3 (für $\varepsilon = 0$)

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

Damit ist Satz 2.5.3 für $\varepsilon = 0$ anwendbar und liefert

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

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

Dies zeigt auch, dass jede neue Schnittebene den zulässigen Bereich von (2.19) echt verkleinert, solange x^k nicht bereits eine optimale Lösung ist.

Eine grundsätzliche Problematik des Schnittebenen-Verfahrens besteht darin, dass das Problem (2.19) 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:

$$\min_x f_k^{\text{se}}(x) + \frac{1}{2\gamma_k}\|x - x^k\|^2 \quad \text{u.d.N.} \quad x \in X. \tag{2.22}$$

Der Parameter $\gamma_k > 0$ muss natürlich geeignet gewählt werden. Die Teilprobleme (2.22) bilden die Basis der Bundle-Verfahren, auf die wir nun hinarbeiten.

2.6 Das ε-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 2.6.1 (ε-Subgradient, ε-Subdifferential) Sei $f : \mathbb{R}^n \to \mathbb{R}$ konvex und $\varepsilon \geq 0$. Der Vektor $g \in \mathbb{R}^n$ heißt ε-*Subgradient* von f im Punkt $x \in \mathbb{R}^n$, wenn gilt:

$$f(y) - f(x) \geq g^T(y - x) - \varepsilon \quad \forall\, y \in \mathbb{R}^n. \tag{2.23}$$

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

Anschauliche Interpretation:

Der Vektor g ist ε-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 2.6.2 (ε-Richtungsableitung) Sei $f : \mathbb{R}^n \to \mathbb{R}$ konvex und $\varepsilon \geq 0$. Die ε-*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 $\phi : \mathbb{R}_+ \to \mathbb{R}$, $\phi(t) = f(x+ts)$, $t \geq 0$. Der Wert $a = f'_\varepsilon(x, s)$ ist nun die Steigung jener Geraden $g : \mathbb{R}_+ \to \mathbb{R}$, $g(t) = f(x) + at - \varepsilon$, die durch den Punkt $(0, f(x) - \varepsilon)$ verläuft und den Graphen der Funktion ϕ von unten berührt.

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

Satz 2.6.3 *Sei $f : \mathbb{R}^n \to \mathbb{R}$ 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 2.3.1 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 $\inf\limits_{t>0}$ liefert nun die Behauptung. □

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

Wir formulieren nun eine ε-Entsprechung zu Satz 2.3.4:

Satz 2.6.4 *Sei $f : \mathbb{R}^n \to \mathbb{R}$ konvex und $x \in \mathbb{R}^n$. Dann gilt:*

a) $\partial_\varepsilon f(x) = \left\{ g \in \mathbb{R}^n \ :\ g^T s \leq f'_\varepsilon(x, s) \ \ \forall\, s \in \mathbb{R}^n \right\}$.

b) $\partial_\varepsilon f(x)$ *ist nichtleer, konvex und kompakt.*

c) $f'_\varepsilon(x,s) = \max\limits_{g \in \partial_\varepsilon f(x)} g^T s \quad \forall\, s \in \mathbb{R}^n$.

Beweis: Erfolgt in ganz ähnlicher Weise wie der Nachweis der entsprechenden Aussagen in Satz 2.3.4.

Satz 2.6.5 *Sei $f : \mathbb{R}^n \to \mathbb{R}$ eine konvexe Funktion, $\bar{x} \in \mathbb{R}^n$ und $\varepsilon \geq 0$. Dann sind die folgenden Aussagen äquivalent:*

a) *Der Punkt $\bar{x}$ ist ε-optimal, d.h.*

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

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

c) *Es gilt* $\quad 0 \in \partial_\varepsilon f(\bar{x})$.

Beweis:

a) $\Longrightarrow$ b):

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

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

b) $\Longrightarrow$ c):

Wegen

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

folgt $0 \in \partial_\varepsilon f(\bar{x})$ gemäß Satz 2.6.4 a).

c) $\Longrightarrow$ a):

Da 0 ein ε-Subgradient ist, folgt nach Definition

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

Somit ist $\bar{x}$ ε-optimal. $\Box$

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

Satz 2.6.6 *Sei $f : \mathbb{R}^n \to \mathbb{R}$ 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 2.3.1 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 $2L\delta \le \varepsilon$. Der Rest des Beweises wird gemeinsam mit b) geführt.

Fall b): Die Funktion f ist nach Satz 2.3.1 a) und Lemma 2.4.6 Lipschitz-stetig auf $\bar{B}_\delta(x)$ mit Konstante $L > 0$. Wähle nun $\varepsilon \ge 2L\delta$.

Fall a) und b) gemeinsam:

Für alle $y \in B_\delta(x)$ und alle $g \in \partial f(y)$ gilt $\|g\| \le L$ nach Satz 2.3.4 b) und daher ergibt sich für alle $z \in \mathbb{R}^n$:

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

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

Im Gegensatz zu f' können wir mit f'_ε robuste Abstiegsrichtungen berechnen. Genauer gilt:

Satz 2.6.7 *Sei $f : \mathbb{R}^n \to \mathbb{R}$ konvex, $\varepsilon \ge 0$ 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 $\dfrac{s}{\|s\|}$ Lösung des Problems*

$$\min_{\|d\|=1} f'_\varepsilon(x,d) \tag{2.24}$$

und es gilt

$$f'_\varepsilon(x,s) = -\|s\|^2 < 0, \tag{2.25}$$

$$\inf_{t>0} f(x+ts) < f(x) - \varepsilon. \tag{2.26}$$

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

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

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 (2.24) und es gilt

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

Weiter folgt

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

Daher gibt es $t^* > 0$ mit

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

Daraus erhalten wir

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

□

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

Algorithmus 4 (Modellalgorithmus)

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

Für $k = 0, 1, 2, \ldots$:

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 2.6.8 Es würde genügen, wenn die Schrittweite σ_k so gewählt wird, dass sie einen Teil des maximal möglichen Abstiegs realisiert.

Dieses Verfahren hat sehr schöne Konvergenzeigenschaften. Es gilt nämlich:

Satz 2.6.9 *Sei $f : \mathbb{R}^n \to \mathbb{R}$ konvex und nach unten beschränkt. Weiter sei $\varepsilon > 0$. Dann terminiert Algorithmus 4 nach endlich vielen Iterationen mit einem ε-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 2.6.7:

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

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 ε nicht mehr möglich ist. Umgekehrt folgt aus $g^k = 0$, dass x^k ε-optimal ist, siehe Satz 2.6.5. $\square$

Da $\partial_\varepsilon f(x)$ in der Praxis schwer oder gar nicht zu berechnen ist, hat Algorithmus 4 nur theoretischen Wert. Er legt aber nahe, $\partial_\varepsilon f(x^k)$ geeignet durch eine Menge G_ε^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.

2.7 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 für ein gewisses ε_k durch geeignete Konvexkombinationen der Subgradienten eine innere Approximation $G_{\varepsilon_k}^k$ des ε_k-Subdifferentials $\partial_{\varepsilon_k} f(x^k)$ zu erzeugen und dieses mittels $s^k = -P_{G_{\varepsilon_k}^k}(0)$ im Sinne des Modellalgorithmus 4 zur Berechnung von Suchrichtungen zu verwenden.

2.7.1 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:

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

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

$$f^{\mathrm{se}}_{J_k}(x) = \max_{j \in J_k} l_j(x) = \max_{j \in J_k} (f^j + {g^j}^T(x - y^j)),$$

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

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

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

$$\min_{s \in \mathbb{R}^n} \; f^{\mathrm{se}}_{J_k}(x^k + s) + \frac{1}{2\gamma_k}\|s\|^2, \tag{2.27}$$

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 γ_k, desto kürzer wird der Schritt ausfallen. Man kann sogar folgendes zeigen:

Ist s^k Lösung von (2.27), so gibt es $\Delta_k > 0$, so dass s^k das folgende Trust-Region-Schnittebenen-Problem löst:

$$\min_{s} \; f^{\mathrm{se}}_{J_k}(x^k + s) \quad \text{u.d.N.} \quad \|s\| \le \Delta_k. \tag{2.28}$$

Ist umgekehrt s^k eine Lösung von (2.28), in der die Nebenbedingung $\|s\| \le \Delta_k$ stark aktiv ist (d.h. nach Vergrößern von Δ_k wäre s^k nicht mehr optimal), so gibt es $\gamma_k > 0$, so dass s^k das Problem (2.27) löst.

Im folgenden ist es günstig, l_j folgendermaßen umzuschreiben:

$$\begin{aligned} l_j(x) &= f^j + {g^j}^T(x - y^j) = f(x^k) + {g^j}^T(x - x^k) - (f(x^k) - f^j + {g^j}^T(y^j - x^k)) \\ &= f(x^k) + {g^j}^T(x - x^k) - \alpha^k_j \qquad (2.29) \end{aligned}$$

mit

$$\alpha^k_j = f(x^k) - f^j - {g^j}^T(x^k - y^j) = f(x^k) - l_j(x^k). \tag{2.30}$$

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

$$f^{\mathrm{se}}_{J_k}(x) = \max_{j \in J_k} l_j(x) = f(x^k) + \max_{j \in J_k}({g^j}^T(x - x^k) - \alpha^k_j) =: f(x^k) + \bar{f}^{\mathrm{se}}_{J_k}(x). \tag{2.31}$$

In dem Problem (2.27) lassen wir den konstanten Offset $f(x^k)$ weg und erhalten

$$\min_s \ \bar{f}^{\text{se}}_{J_k}(x^k + s) + \frac{1}{2\gamma_k}\|s\|^2. \tag{2.32}$$

Gemäß Lemma 2.5.1 können wir dieses Problem als QP schreiben:

$$\min_{s\in\mathbb{R}^n,\xi\in\mathbb{R}} \xi + \frac{1}{2\gamma_k}\|s\|^2 \quad \text{u.d.N.} \quad {g^j}^T s - \alpha_j^k - \xi \le 0, \ \ j \in J_k. \tag{2.33}$$

Das Paar (s^k, ξ^k) ist genau dann Lösung von (2.33), wenn s^k Lösung von (2.32) ist und $\xi^k = \bar{f}^{\text{se}}_{J_k}(x^k + s^k)$ gilt.

Die Lösung von (2.33) erfüllt folgende Karush-Kuhn-Tucker-Bedingungen:

Lemma 2.7.1 (s^k, ξ^k) *ist genau dann Lösung von* (2.33)*, wenn die KKT-Bedingungen gelten, d.h. wenn es* $\lambda_j^k \in \mathbb{R}$, $j \in J_k$, *gibt mit*

$$\begin{gathered} \frac{1}{\gamma_k}s^k + \sum_{j\in J_k} \lambda_j^k g^j = 0, \\ \sum_{j\in J_k} \lambda_j^k = 1, \\ {g^j}^T s^k - \alpha_j^k - \xi^k \le 0, \ \ \lambda_j^k \ge 0, \ \ \lambda_j^k({g^j}^T s^k - \alpha_j^k - \xi^k) = 0, \ \ j \in J_k. \end{gathered} \tag{2.34}$$

Beweis: Für das konvexe QP (2.33) sind die KKT-Bedingungen notwendig und hinreichend. Diese KKT-Bedingungen sind in (2.34) angegeben. □

Ist der Schritt s^k berechnet, so prüfen wir (wie auch bei Trust-Region Verfahren der Nichtlinearen Optimierung üblich), ob die tatsächliche Zielfunktionsabnahme

$$\text{ared}_k(s^k) = f(x^k) - f(x^k + s^k) \tag{2.35}$$

hinreichend groß ist im Vergleich zur Modellabnahme

$$\text{pred}_k(s^k) = f^{\text{se}}_{J_k}(x^k) - f^{\text{se}}_{J_k}(x^k + s^k) = f(x^k) - f^{\text{se}}_{J_k}(x^k + s^k) = -\bar{f}^{\text{se}}_{J_k}(x^k + s^k) = -\xi^k. \tag{2.36}$$

Hierzu verwenden wir die Bedingung

$$\text{ared}_k(s^k) \ge \eta \,\text{pred}_k(s^k) = -\eta\xi^k. \tag{2.37}$$

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.

Wir gehen nicht detailliert auf die Steuerung von γ_k ein, denn es gibt hier viele Varianten:

- In [GK02] wird stets $\gamma_k = 1$ gewählt. Dies ist nicht sonderlich effizient, vereinfacht aber die Konvergenzanalyse.
- In [SZ92] wird in einer inneren Iteration der Wert von γ_k (in gewisser Weise also der Trust-Region-Radius) geeignet justiert. Dies hat folgenden Hintergrund: Das Fehlschlagen von (2.37) kann zwei Gründe haben, nämlich

  a) der Trust-Region-Radius (also γ_k) ist zu groß, oder

  b) das Schnittebenenmodell ist zu schlecht.

  Im Fall a) macht es Sinn, ein kleineres γ_k zu probieren. Im Fall b) ist es hingegen sinnvoll, einen Nullschritt zu machen. Man bleibt dann im Punkt x^k, verbessert aber das Modell im Punkt $y^{k+1} = x^k + s^k$ durch Aufnehmen des Punktes y^{k+1} in das Bündel.

  Wenn der Test in (2.37) sehr positiv ausfällt, kann man auch noch einmal ein größeres γ_k probieren, um eventuell einen noch besseren Schritt s^k zu erhalten.
- Wir fordern im folgenden nur, dass $\gamma_k \in [\gamma^-, \gamma^+] \subset (0, \infty)$ gewählt wird und lassen damit viele Freiheiten bei der Wahl einer Strategie zur adaptiven Bestimmung von γ_k zu.

Es ergibt sich das folgende Verfahren:

Algorithmus 5 (Bundle-Verfahren)

0. Wähle $x^0 \in \mathbb{R}^n$, $\gamma^+ \geq \gamma^- > 0$, $\eta \in (0,1)$ 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 $k = 0, 1, 2, \ldots$:

1. Wähle $\gamma_k \in [\gamma^-, \gamma^+]$ und berechne ein KKT-Tupel (s^k, ξ^k, λ^k) des Problems (2.33).
2. Berechne $v^k = -\dfrac{1}{\gamma_k} s^k$ 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:
$$y^{k+1} = x^k + s^k, \;\; x^{k+1} = y^{k+1}, \;\; t_k = 1, \;\; J_{k+1} = \left\{ j \in J_k \;:\; \lambda_j^k > 0 \right\} \cup \{k+1\}.$$
5. Gilt
$$f(x^k + s^k) - f(x^k) > \eta \xi^k,$$
so führe einen *Nullschritt* durch:
$$y^{k+1} = x^k + s^k, \;\; x^{k+1} = x^k, \;\; t_k = 0, \;\; J_{k+1} = \left\{ j \in J_k \;:\; \lambda_j^k > 0 \text{ oder } y^j = x^k \right\} \cup \{k+1\}.$$

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

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

Einige Bemerkungen:

- Die Abbruchbedingung wird erst mit Lemma 2.7.4 klar werden. Sie sichert ein gewisse Form von ε-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.

2.7.2 Eine duale Interpretation des Bundle-Verfahrens

Wir können das Bundle-Verfahren auch als Modifikation von Algorithmus 4 interpretieren. In Algorithmus 4 haben wir

$$x_{k+1} - x_k = -\sigma_k P_{\partial_\varepsilon f(x^k)}(0).$$

Wir werden in diesem Abschnitt sehen, dass beim Bundle-Verfahren sehr ähnlich gilt

$$x_{k+1} - x_k = s^k = -\gamma_k v_k = -\gamma_k P_{G^k_{\varepsilon_k}}(0), \tag{2.38}$$

mit

$$\varepsilon_k = \sum_{j \in J_k} \lambda_j^k \alpha_j^k$$

wie in Schritt 2 von Algorithmus 5 und

$$G_\varepsilon^k = \left\{ \sum_{j \in J_k} \lambda_j g^j \; : \; \sum_{j \in J_k} \lambda_j \alpha_j^k \le \varepsilon, \; \sum_{j \in J_k} \lambda_j = 1, \; \lambda_j \ge 0, \; j \in J_k \right\}. \tag{2.39}$$

Der wichtige Punkt dabei ist, dass

$$G_\varepsilon^k \subset \partial_\varepsilon f(x^k) \tag{2.40}$$

gilt, $\partial_\varepsilon f(x^k)$ also durch die innere Approximation G_ε^k ersetzt wird. Damit kann das Bundle-Verfahren als implementierbare Modifikation von Algorithmus 4 angesehen werden.

Wir zeigen nun zunächst (2.40) und wenden uns dann (2.38) zu.

Lemma 2.7.2 *Sei $f : \mathbb{R}^n \to \mathbb{R}$ konvex. Weiter seien $x^k \in \mathbb{R}^n$, $y^j \in \mathbb{R}^n$, $g^j \in \partial f(y^j)$, $j \in J_k$, $\varepsilon \ge 0$ und G_ε^k definiert gemäß* (2.39). *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 α_j^k und wegen $g^j \in \partial f(y^j)$ gilt für alle $x \in \mathbb{R}^n$

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

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

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

$$f(x) - f(x^k) \ge {g^j}^T (x - x^k) - \alpha_j^k$$

mit λ_j und Aufaddieren liefert

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

Damit ist

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

gezeigt. $\square$

Wir wollen nun noch (2.38) nachweisen, also dass der Schritt des Bundle-Verfahrens gegeben ist durch

$$s^k = -\gamma_k v_k = -\gamma_k P_{G^k_{\varepsilon_k}}(0),$$

Hierzu leiten wir für das Teilproblem (2.33) ein duales Problem her.

Lemma 2.7.3 a) (s^k, ξ^k) *ist genau dann Lösung von* (2.33)*, wenn die KKT-Bedingungen gelten, d.h. wenn es* $\lambda_j^k \in \mathbb{R}$*,* $j \in J_k$*, gibt mit*

$$\begin{aligned} \frac{1}{\gamma_k} s^k + \sum_{j \in J_k} \lambda_j^k g^j &= 0, \\ \sum_{j \in J_k} \lambda_j^k &= 1, \\ {g^j}^T s^k - \alpha_j^k - \xi^k \le 0, \quad \lambda_j^k \ge 0, \quad \lambda_j^k ({g^j}^T s^k - \alpha_j^k - \xi^k) &= 0, \quad j \in J_k. \end{aligned} \tag{2.34}$$

b) *Der Vektor* λ^k *ist genau dann Lösung des Problems*

$$(2.41)\qquad \begin{aligned} &\min_{\lambda} \quad \frac{\gamma_k}{2}\left\|\sum_{j\in J_k}\lambda_j g^j\right\|^2 + \sum_{j\in J_k}\lambda_j\alpha_j^k \\ &\text{u.d.N.} \quad \sum_{j\in J_k}\lambda_j = 1,\ \ \lambda_j \geq 0,\ \ j\in J_k, \end{aligned}$$

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* (2.41)*, wenn* (2.34) *für*

$$(2.42)\qquad s^k = -\gamma_k \sum_{j\in J_k}\lambda_j^k g^j,$$

erfüllt ist und zusätzlich gilt:

$$(2.43)\qquad \mu_j^k = -{g^j}^T s^k + \alpha_j^k + \xi^k,\ \ j\in J_k.$$

gilt.

c) *Sei* λ^k *eine Lösung von* (2.41)*. Dann ist* λ^k *auch Lösung des folgenden Problems:*

$$(2.44)\qquad \begin{aligned} &\min_{\lambda} \quad \frac{\gamma_k}{2}\left\|\sum_{j\in J_k}\lambda_j g^j\right\|^2 \\ &\text{u.d.N.} \quad \sum_{j\in J_k}\lambda_j = 1,\ \ \lambda_j \geq 0,\ \ j\in J_k,\ \ \sum_{j\in J_k}\lambda_j\alpha_j^k \leq \varepsilon_k \end{aligned}$$

mit

$$(2.45)\qquad \varepsilon_k = \sum_{j\in J_k}\lambda_j^k\alpha_j^k.$$

Beweis: zu a): Das wissen wir aus Lemma 2.7.1.

zu b):

Die KKT-Bedingungen sind für das konvexe QP (2.41) notwendig und hinreichend und lauten:

$$(2.46)\qquad \begin{aligned} \gamma_k {g^j}^T \sum_{i\in J_k}\lambda_i^k g^i + \alpha_j^k + \xi^k - \mu_j^k = 0,\ \ j\in J_k \\ \sum_{j\in J_k}\lambda_j^k = 1, \\ \lambda_j^k \geq 0,\ \ \mu_j^k \geq 0,\ \ \mu_j^k\lambda_j^k = 0,\ \ j\in J_k. \end{aligned}$$

Gelte nun (2.34). Die erste Zeile in (2.34) liefert dann (2.42). Definieren wir nun μ_j^k gemäß (2.43), und setzen (2.42) ein, so ergibt sich die erste Gleichung in (2.46). Umschreiben der dritten Gleichung von (2.34) auf μ_j^k ergibt die dritte Gleichung in (2.46). Die zweite Gleichung in (2.34) und in (2.46) sind identisch.

Gelte nun umgekehrt (2.46). Definieren wir s^k gemäß (2.42), so folgt die erste Gleichung in (2.34). Einsetzen von (2.42) in die erste Gleichung von (2.46) ergibt (2.43) und Einsetzen von (2.43) in die dritte Zeile von (2.46) liefert die dritte Zeile von (2.34). Die zweiten Zeilen sind wiederum identisch.

zu c):

Für das konvexe QP (2.44) sind die folgenden KKT-Bedingungen notwendig und hinreichend:

$$
\begin{aligned}
\gamma_k g^{j^T} \sum_{i\in J_k} \lambda_i^k g^i + \xi^k - \mu_j^k + \tau^k \alpha_j^k = 0, \;\; j \in J_k \\
\sum_{j\in J_k} \lambda_j^k = 1, \\
\lambda_j^k \geq 0, \;\; \mu_j^k \geq 0, \;\; \mu_j^k \lambda_j^k = 0, \;\; j \in J_k, \\
\sum_{j\in J_k} \lambda_j^k \alpha_j^k \leq \varepsilon_k, \;\; \tau^k \geq 0, \;\; \tau^k \left(\sum_{j\in J_k} \lambda_j^k \alpha_j^k - \varepsilon_k \right) = 0.
\end{aligned}
\tag{2.47}
$$

Ist nun λ^k Lösung von (2.41), so gelten die Bedingungen (2.46). Definieren wir ε_k gemäß (2.45) und setzen wir $\tau^k = 1$, so folgt unmittelbar (2.47). $\square$

Wie soeben gezeigt, sind also die Probleme (2.33) und (2.41) äquivalent. Aus einer Lösung von (2.41) läßt sich s^k aus (2.42) zurückgewinnen. Der Wert ξ^k kann ebenfalls direkt berechnet werden. Hierzu benutzen wir die Komplementaritätsbedingung in (2.34), (2.42) und (2.45):

$$
\begin{aligned}
\xi^k = \xi^k \sum_{j\in J_k} \lambda_j^k &= \sum_{j\in J_k} \lambda_j^k \xi^k = \sum_{j\in J_k} \lambda_j^k (g^{j^T} s^k - \alpha_j^k) \\
&= \left(\sum_{j\in J_k} \lambda_j^k g^j \right)^T s^k - \sum_{j\in J_k} \lambda_j^k \alpha_j^k = -\frac{1}{\gamma_k} \|s^k\|^2 - \varepsilon_k.
\end{aligned}
$$

Damit ergeben sich folgende Formeln:

$$
s^k = -\gamma_k v^k, \;\; \xi^k = -\frac{1}{\gamma_k} \|s^k\|^2 - \varepsilon_k = -\gamma_k \|v^k\|^2 - \varepsilon_k \tag{2.48}
$$

$$
\text{mit} \quad v^k = \sum_{j\in J_k} \lambda_j^k g^j, \quad \varepsilon_k = \sum_{j\in J_k} \lambda_j^k \alpha_j^k. \tag{2.49}
$$

Wir kommen nun zu einem wichtigen Punkt: Gemäß Lemma 2.7.3 ist λ^k Lösung von (2.44) und daher gilt

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

wobei

$$G_\varepsilon^k = \left\{ \sum_{j\in J_k} \lambda_j g^j \;:\; \sum_{j\in J_k} \lambda_j \alpha_j^k \le \varepsilon,\; \sum_{j\in J_k} \lambda_j = 1,\; \lambda_j \ge 0,\; j \in J_k \right\}. \tag{2.39}$$

Somit ist auch (2.38) nachgewiesen.

Wir können daher Algorithmus 5 auch so formulieren:

Algorithmus 6 (Bundle-Verfahren 5 in dualer Formulierung)

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

Für $k = 0, 1, 2, \ldots$:

1. Wähle $\gamma_k \in [\gamma^-, \gamma^+]$ und berechne λ^k durch Lösen des Problems (2.41).
2. Berechne
$$v^k = \sum_{j\in J_k} \lambda_j^k g^j, \quad s^k = -\gamma_k v^k, \quad \varepsilon_k = \sum_{j\in J_k} \lambda_j^k \alpha_j^k, \quad \xi^k = -\gamma_k \|v^k\|^2 - \varepsilon_k.$$
3. Prüfe auf Abbruch: Falls $\|v^k\| \le \varepsilon$ und $\varepsilon_k \le \varepsilon$, STOP.
4. Gilt
$$f(x^k + s^k) - f(x^k) \le \eta \xi^k,$$
so führe einen *wesentlichen Schritt* durch:
$$y^{k+1} = x^k + s^k,\; x^{k+1} = y^{k+1},\; t_k = 1,\; J_{k+1} = \left\{ j \in J_k \;:\; \lambda_j^k > 0 \right\} \cup \{k+1\}.$$
5. Gilt
$$f(x^k + s^k) - f(x^k) > \eta \xi^k,$$
so führe einen *Nullschritt* durch:
$$y^{k+1} = x^k + s^k,\; x^{k+1} = x^k,\; t_k = 0,\; J_{k+1} = \left\{ j \in J_k \;:\; \lambda_j^k > 0 \text{ oder } y^j = x^k \right\} \cup \{k+1\}.$$
6. Berechne $f^{k+1} = f(y^{k+1})$, $g^{k+1} \in \partial f(y^{k+1})$ und
$$\alpha_j^{k+1} = f(x^{k+1}) - f^j - {g^j}^T (x^{k+1} - y^j), \quad j \in J_{k+1}.$$

Wir sehen uns nun die Abbruchbedingung näher an.

Lemma 2.7.4 *In Algorithmus 5 (bzw. Algorithmus 6) sei für $\varepsilon > 0$ die Abbruchbedingung*

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

erfüllt. Dann ist x_k ε-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 (2.29)

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

Multiplizieren mit λ_j^k und Summieren ergibt:

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

Dies ergibt mit der Cauchy-Schwarz-Ungleichung

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

□

2.7.3 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:

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

Wir beginnen mit einem technischen Resultat:

Lemma 2.7.5 *Sei $f : \mathbb{R}^n \to \mathbb{R}$ konvex. Sei $\varepsilon = 0$ und die Folgen (x^k), (s^k) usw. seien von Algorithmus 5 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) $\lim_{k\to\infty} f(x^{k+1}) - f(x^k) = 0$.

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

c) *Erzeugt der Algorithmus unendlich viele wesentliche Schritte, d.h. gilt* $|\mathcal{K}| = \infty$, *so folgt*

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

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

$$f(x^{k+1}) - f(x^k) \le \eta t_k \xi^k$$

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

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

wobei wir (2.48) benutzt haben.

zu c):

Nach b) gilt

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

daher ist $(\gamma^-\|v^k\|^2 + \varepsilon_k)_{\mathcal{K}}$ eine Nullfolge und somit (wegen $\gamma^- > 0$ und $\varepsilon_k \ge 0$) 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 2.7.6 *Algorithmus 5 mit* $\varepsilon = 0$ *erzeuge unendlich viele wesentliche Schritte. Dann ist jeder Häufungspunkt von* (x^k) *ein (globales) Minimum von* f.

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

$$\{x^k \,:\, k \ge 0\} = \{x^0\} \cup \{x^{k+1} \,:\, k \in \mathcal{K}\} = \{x^k \,:\, k \in \mathcal{K}\}$$

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

$$(x^k)_{\mathcal{K}'} \to \bar{x}.$$

Gemäß Lemma 2.7.2 gilt $v^k \in \partial_{\varepsilon_k} f(x^k)$ und aus Lemma 2.7.5 c) folgt

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

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

$$f(x) \geq f(x^k) + v^{k^T}(x - x^k) - \varepsilon_k \stackrel{\mathcal{K}' \ni k \to \infty}{\longrightarrow} f(\bar{x}) + 0^T(x - \bar{x}) + 0 = f(\bar{x}).$$

Damit ist $\bar{x}$ globales Minimum von f. $\Box$

Wir untersuchen nun den Fall, dass nur endlich viele wesentliche Schritte auftreten. Wir betrachten zunächst den einfacheren Fall, dass im Falle von Nullschritten das Bundle nicht bereinigt wird, also bei einem *Nullschritt*:

$$y^{k+1} = x^k + s^k, \;\; x^{k+1} = x^k, \;\; t_k = 0, \;\; J_{k+1} = J_k \cup \{k+1\}\,.$$

Lemma 2.7.7 *Algorithmus 5 mit dem eben genannten Update von J_{k+1} bei Nullschritten und $\varepsilon = 0$ erzeuge eine unendliche Folge (x^k). 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. Zudem gilt auch $v^k \to 0$ und $\varepsilon_k \to 0$ für $k \to \infty$.*

Beweis: Annahme, ab Iteration l finden nur Nullschritte statt. Sei x^l nicht optimal. Dann gibt es $\bar{x}$ mit $f^{se}_{J_k}(\bar{x}) \leq f(\bar{x}) < f(x^l)$ für alle $k \geq l$. Es gibt also $\tilde{x}$ auf der Strecke $[\bar{x}, x^l]$ und $\delta > 0$ mit

$$f(\tilde{x}) + \frac{1}{2\gamma_k}\|\tilde{x} - x^l\|^2 < f(x^l) - \delta \quad \forall\, k \geq l$$

und somit

$$f^{se}_{J_k}(\tilde{x}) + \frac{1}{2\gamma_k}\|\tilde{x} - x^l\|^2 \leq f(\tilde{x}) + \frac{1}{2\gamma_k}\|\tilde{x} - x^l\|^2 < f(x^l) - \delta \quad \forall\, k \geq l.$$

Nun gilt $f^{se}_{J_k}(y^{k+1}) + \frac{1}{2\gamma_k}\|y^{k+1} - x^l\|^2 \leq f(x^l)$ und wegen $f^{se}_{J_k}(x^l + s) + \frac{1}{2\gamma_k}\|s\|^2 \geq f(x^l) + (g^l)^T s + \frac{1}{2\gamma_+}\|s\|^2$ liegen alle y^k, $k \geq l$ in einer kompakten Niveaumenge. Es gibt also $K \subset \{k : \; k \geq l\}$ mit

$$(y^{k+1})_{k+1 \in K} \to x^*$$

für ein x^*.

Nun gilt

$$f^{se}_{J_k}(y^{k+1}) + \frac{1}{2\gamma_k}\|y^{k+1} - x^l\|^2 \leq f^{se}_{J_k}(\tilde{x}) + \frac{1}{2\gamma_k}\|\tilde{x} - x^l\|^2 \leq f(\tilde{x}) + \frac{1}{2\gamma_k}\|\tilde{x} - x^l\|^2 < f(x^l) - \delta$$

und somit

$$\mathrm{pred}_k(s^k) = f_{J_k}^{se}(x^l) - f_{J_k}^{se}(y^{k+1}) = f(x^l) - f_{J_k}^{se}(y^{k+1}) \geq \delta + \frac{1}{2\gamma_k}\|y^{k+1} - x^l\|^2 \geq \delta \quad \forall\, k \geq l.$$

Seien nun $k+1 \geq i \geq l$ mit $k+1, i \in K$ beliebig. Dann gilt

$$f(x^*) \overset{k+1\in K\to\infty}{\longleftarrow} f(y^{k+1}) \geq f_{J_k}^{se}(y^{k+1}) \geq f(y^i) + (g^i)^T(y^{k+1} - y^i) \overset{k+1,i\in K\to\infty}{\longrightarrow} f(x^*).$$

Dies zeigt $f(y^{k+1}) - f_{J_k}^{se}(y^{k+1}) \overset{k+1\in K\to\infty}{\longrightarrow} 0$ und daher

$$\mathrm{ared}_k(s^k) - \mathrm{pred}_k(s^k) = f_{J_k}^{se}(y^{k+1}) - f(y^{k+1}) \overset{k+1\in K\to\infty}{\longrightarrow} 0.$$

Da aber $\mathrm{pred}_k(s^k) \geq \delta > 0$ gäbe es also $k \geq l$ mit

$$\mathrm{ared}_k(s^k) \geq \eta \mathrm{pred}_k(s^k)$$

und Schritt k wäre kein Nullschritt. Widerspruch!

Damit ist x^l globales Minimum von f.

Wir zeigen noch $v^k \to 0$, $\varepsilon_k \to 0$. Wegen $v^k = -\gamma_k s^k$, $\xi^k = -\mathrm{pred}_k(s^k)$ und $\varepsilon_k = -\gamma_k\|v^k\|^2 - \xi^k$ reicht es zu zeigen, dass $y^k \to x^l$ für $k \to \infty$.

Falls $y^k \not\to x^l$ für $k \to \infty$, dann gäbe es eine Teilfolge $(y^k)_{k\in K} \to x^* \neq x^l$. Wie eben gilt ($x^l$ ist globales Minimum!)

$$f(y^{k+1}), f_{J_k}^{se}(y^{k+1}) \overset{k+1\in K\to\infty}{\longrightarrow} f(x^*) \geq f(x^l)$$

und

$$f_{J_k}^{se}(y^{k+1}) + \frac{1}{2\gamma_k}\|y^{k+1} - x^l\| \leq f_{J_k}^{se}(x^l) = f(x^l).$$

Grenzübergang $k+1 \in K \to \infty$ liefert $y^{k+1} \overset{k+1\in K\to\infty}{\longrightarrow} x^l$, also gilt doch $x^* = x^l$ und $(y^k) \to x^l$. $\square$

Der ursprüngliche Update von J_k ist komplizierter zu analysieren: Hierzu treffen wir, um uns das Leben etwas zu erleichtern, die folgende Annahme:

$$\text{(2.50)} \qquad \begin{array}{l} \exists\, m > 0: \ \ \forall\, k \geq m: \\ \{k-m, k-m+1, \dots, k-1\} \cap \mathcal{K} = \emptyset \implies \gamma_k = \gamma^-. \end{array}$$

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

Lemma 2.7.8 *Algorithmus 5 mit $\varepsilon = 0$ erzeuge eine unendliche Folge (x^k) und es gelte* (2.50). *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
$$\alpha_j^{k+1} = \alpha_j^k \quad \forall\, j \in J_k^+,\ k \geq l.$$
Ohne Einschränkung können wir wegen (2.50) annehmen, dass gilt:
$$\gamma_k = \gamma^- \quad \forall\, k \geq l.$$
Wir setzen nun

(2.51)
$$\theta_k := \sum_{j \in J_k^+} \lambda_j^k \alpha_j^{k+1} = \sum_{j \in J_k^+} \lambda_j^k \alpha_j^k = \varepsilon_k \quad \forall\, k \geq l.$$

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 (2.41).

Im folgenden sei $k > l$ beliebig.

Wir wählen nun zu beliebigem $\mu \in [0,1]$ den Vektor $\lambda^{k\mu}$ in folgender Weise:
$$\lambda_k^{k\mu} = \mu, \quad \lambda_j^{k\mu} = (1-\mu)\lambda_j^{k-1},\ \ j \in J_{k-1}^+, \quad \lambda_j^{k\mu} = 0,\ \ j \in J_k \setminus (J_{k-1}^+ \cup \{k\}).$$
Hierbei sei λ^{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 (2.41) 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 θ_{k-1} wie in (2.51) 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 μ^k das Minimum von q_k auf $[0,1]$. Weiter sei $\nu_k = q_k(\mu_k)$. Dann ergibt sich

$$\begin{aligned}\nu_k = q_k(\mu_k) \le q_k(0) &= \frac{\gamma^-}{2}\|v^{k-1}\|^2 + \theta_{k-1} = \frac{\gamma^-}{2}\|v^{k-1}\|^2 + \varepsilon_{k-1}\\ &= Q_{k-1}(\lambda^{k-1}) \le Q_{k-1}(\lambda^{k-1,\mu_{k-1}}) = q_{k-1}(\mu_{k-1}) = \nu_{k-1}.\end{aligned} \tag{2.52}$$

Da q_k quadratisch ist, ergibt sich

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

Wir schätzen nun $q_k'(0)$ ab:

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

Wir benutzen $x^k = x^{k-1}$, $y^k = x^{k-1} + s^k$, $\mathrm{ared}_{k-1}(s^{k-1}) = f(x^{k-1}) - f^k < -\eta\xi^{k-1}$ (sonst wäre $k-1 \in \mathcal{K}$) sowie (2.48) und erhalten

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

Daraus folgt

$$q_k'(0) = -\gamma^-\|v^{k-1}\|^2 + \gamma^- {g^k}^T v^{k-1} + \alpha_k^k - \theta_{k-1} \le -(1-\eta)\big(\theta_{k-1} + \gamma^-\|v^{k-1}\|^2\big).$$

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

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

Daraus wiederum folgt mit (2.52)

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

Somit ist die Folge $(y^k)_{k>l}$ beschränkt und daher auch die Folge $(g^k)_{k>l}$, siehe Satz 2.3.4 und Lemma 2.4.6. Insgesamt gibt es daher eine Konstante $C > 0$ mit

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

Nun folgt

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

Für alle $\mu \ge 0$ ergibt sich

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

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

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

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

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

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

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

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

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

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

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

Wir können nun fortfahren wie im zweiten Teil des Beweises von Lemma 2.7.6, 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. $\Box$

Nehmen wir Lemma 2.7.6 und Lemma 2.7.8 (oder Lemma 2.7.7 für den vereinfachten J_k-update bei Nullschritten) zusammen, dann erhalten wir den folgenden Konvergenzsatz:

Satz 2.7.9 *Algorithmus 5 (oder mit dem vereinfachten J_k-update bei Nullschritten) mit $\varepsilon = 0$ erzeuge die Folge (x^k) und es gelte* (2.50). *Dann ist jeder Häufungspunkt von (x^k) ein globales Minimum von f.*

2.8 Anwendungen

2.8.1 Eigenwertoptimierung

In zahlreichen Anwendungen tritt das Problem auf, den größten Eigenwert einer symmetrischen Matrix zu minimieren. Wir betrachten also die Funktion

$$\lambda_{\max} : A \in S_n \mapsto \max_{\|v\|=1} v^T Av,$$

wobei $S_n = \{A \in \mathbb{R}^{n,n} \;:\; A = A^T\}$ die Menge aller symmetrischen $n \times n$-Matrizen bezeichne. Beachte hierbei, dass bei symmetrischen Matrizen die Eigenvektoren eine Orthogonalbasis bilden und daher tatsächlich $\lambda_{\max}(A) = \max_{\|v\|=1} v^T Av$ ist.

Wir versehen S_n mit dem Skalarprodukt $A \bullet B := \sum_{i,j=1}^n A_{ij}B_{ij}$, wobei $A = (A_{ij})$, $B = (B_{ij})$ (also dem euklidischen Skalarprodukt, wenn wir Matrizen als n^2-Vektoren auffassen).

Man kann folgendes zeigen:

a) Die Funktion $\lambda_{\max}$ is als Maximum linearer Funktionen konvex.

b) Für alle $v \in \mathbb{R}^n$ und $A \in S_n$ gilt:

$$(vv^T) \bullet A = v^T Av \leq \lambda_{\max}(A)\|v\|^2.$$

Zudem gilt Gleichheit in der rechten Ungleichung, falls v ein Eigenvektor von A zu $\lambda_{\max}(A)$ ist.

c) Sei q, $\|q\| = 1$, ein Eigenvektor zum maximalen Eigenwert von $A \in S_n$.
Dann ist qq^T ein Subgradient von $\lambda_{\max}$ in A, also:

$$\lambda_{\max}(B) - \lambda_{\max}(A) \geq (qq^T) \bullet (B - A) \quad \forall\, B \in S_n.$$

d) Allgemein gilt

$$\partial\lambda_{\max}(A) = \{W \in S_n \;:\; W \text{pos. semidefinit},\;\; W \bullet A = \lambda_{\max}(A),\;\; \operatorname{spur}(W) = 1\}.$$

Für a)–c) siehe Blatt 2. d) zur Übung.

Also ist für eine konvexe abgeschlossene Menge $\mathcal{Z} \subset S_n$ das Problem

$$\min \lambda_{\max}(X) \;\text{ s.t. }\; X \in \mathcal{Z}$$

ein konvexes nichtglattes Optimierungsproblem.

2.8.2 SDP-Relaxierungen kombinatorischer Optimierungsprobleme

Semidefinite Programme

Sei wie eben $S_n \subset \mathbb{R}^{n,n}$ die Menge der symmetrischen $n \times n$-Matrizen und S_n^+ die Menge der (symmetrischen) positiv semidefiniten Matrizen. Wir schreiben $A \succeq 0$ genau dann, wenn $A \in S_n^+$.

Ein semidefinites Programm hat nun die Form

(P) $$\max_X C \bullet X \;\text{ s.t. }\; A_i \bullet X = b_i, \;\; i = 1, \ldots, m, \quad X \succeq 0$$

mit $A_i, C \in S_n$. Das dazu duale Problem ist definiert durch

(D) $$\min_{Z,y} b^T y \;\text{ s.t. }\; Z + C - \sum_{i=1}^m y_i A_i = 0, \quad Z \succeq 0.$$

Man kann (D) offensichtlich auch schreiben als

(D) $$\min_y b^T y \;\text{ s.t. }\; \lambda_{max}\left(C - \sum_{i=1}^m y_i A_i\right) \leq 0.$$

Besitzen (P) und (D) strikt innere Punkte, dann haben (P) und (D) optimale Lösungen mit gleichem Zielfunktionswert.

Umwandlung in ein unrestringiertes nichtglattes Problem

Wir setzen
$$\mathcal{A}(v) := \sum_{i=1}^m v_i A_i$$
und nehmen nun an, dass es ein $v \in \mathbb{R}^m$ gibt mit
$$\mathcal{A}(v) := \sum_{i=1}^m v_i A_i = I.$$
Dann können wir das duale Problem

(D) $$\min_y b^T y \;\text{ s.t. }\; \lambda_{max}\left(C - \mathcal{A}(y)\right) \leq 0$$

in ein äquivalentes nichtglattes Problem umwandeln.

Sei hierzu $b \neq 0$. Wir nehmen die Nebenbedingung über einen Lagrange-Multiplikator $\mu \in \mathbb{R}$ in die Zielfunktion auf, betrachten also

(D′) $$\min_y b^T y + \mu\, \lambda_{max}\left(C - \mathcal{A}(y)\right).$$

Mit geeigneter Wahl von μ ist dann (D′) tatsächlich äquivalent zu (D):

Satz 2.8.1 *Sei $v \in \mathbb{R}^m$ mit*

$$\mathcal{A}(v) = \sum_{i=1}^{m} v_i A_i = I.$$

(D) besitze eine optimale Lösung.

Für $\mu := b^T v$ sind dann (D) und (D′) äquivalent in folgendem Sinne:

a) Ist $\bar{y}$ Lösung von (D), dann ist $\bar{y}$ auch Lösung von (D′).

b) Ist $\bar{y}$ Lösung von (D′), dann $\tilde{y} = \bar{y} + \lambda_{max}(C - \mathcal{A}(\bar{y}))\, v$ Lösung von (D).

Beweis: Vorüberlegung: Die Punkte y und $y(t) := y + tv$, $t \in \mathbb{R}$ haben in (D′) durch die Wahl $\mu = b^T v$ denselben Zielfunktionswert:

$$\begin{aligned} &b^T(y + tv) + \mu\, \lambda_{max}(C - \mathcal{A}(y + tv)) = b^T y + t\mu + \mu\, \lambda_{max}(C - \mathcal{A}(y) - tI) \\ &= b^T y + t\mu + \mu\, \lambda_{max}(C - \mathcal{A}(y)) - t\mu = b^T y + \mu\, \lambda_{max}(C - \mathcal{A}(y)). \end{aligned}$$

Wählt man $t = \lambda_{max}(C - \mathcal{A}(y))$, dann gilt zudem $\lambda_{max}(C - \mathcal{A}(y(t))) = 0$.

Da (D) eine Lösung hat, gilt zudem $\mu \geq 0$. Andernfalls sind die Punkte $\bar{y}(t) = \bar{y} + tv$ für alle $t \geq 0$ zulässig, da

$$\lambda_{max}(C - \mathcal{A}(\bar{y} + tv)) = \lambda_{max}(C - \mathcal{A}(\bar{y}) - tI) = \lambda_{max}(C - \mathcal{A}(\bar{y})) - t \leq 0.$$

Für den Zielfunktionswert gilt

$$b^T y = b^T \bar{y} + t\mu$$

und er ist für alle $t \geq 0$ nur dann nach unten beschränkt, wenn $\mu \geq 0$ ist.

zu a): Sei $\bar{y}$ optimal für (D).

Die Zielfunktionswerte von (D) bzw. (D′) sind für $\bar{y}$ jeweils gleich $b^T \bar{y}$ bzw.

$$b^T \bar{y} + \mu \lambda_{max}(C - \mathcal{A}(\bar{y})) \leq b^T \bar{y}.$$

Annahme, es gäbe ein $tildey$ mit

$$b^T \tilde{y} + \mu\, \lambda_{max}(C - \mathcal{A}(\tilde{y})) < b^T \bar{y} + \mu \lambda_{max}(C - \mathcal{A}(\bar{y})) \leq b^T \bar{y}.$$

Setze nun $t = \lambda_{max}(C - \mathcal{A}(\tilde{y}))$ und betrachte $y = \tilde{y} + tv$. Dann haben y und $\tilde{y}$ in (D′) denselben Zielfunktionswert und es gilt $\lambda_{max}(C - \mathcal{A}(y)) = 0$. y ist also zulässig für (D) und es gilt

$$b^T y = b^T \tilde{y} + \mu\, \lambda_{max}(C - \mathcal{A}(\tilde{y})) < b^T \bar{y}.$$

Dies ist ein Widerspruch zur Optimalität von $\bar{y}$.

zu b): Ist $\bar{y}$ Lösung von (D′), dann ist nach der Vorüberlegung $\tilde{y} = \bar{y} + \lambda_{max}(C - \mathcal{A}(\bar{y}))\, v$ ebenfalls optimal und es gilt $\lambda_{max}(C - \mathcal{A}(\tilde{y})) = 0$.

Annahme, es gäbe einen zulässigen Punkt y für (D) mit $b^T y < b^T \tilde{y}$. Dann wäre wegen $\mu \geq 0$ der Punkt y muss $\mu < 0$ gelten, sonst besser als $\tilde{y}$ in (D′). Dies ist ein Widerspruch zur Optimalität von $\tilde{y}$. $\square$

SDP-Relaxierung des Max-Cut-Problems

Sei $G = (V, E, w)$ ein ungerichteter, gewichteter Graph mit Knotenmenge $V = \{1, \dots, n\}$, Kantenmenge E und nichtnegativen Gewichten $w = (w_{ij})_{(i,j)\in E}$.

Wir betrachten das *Max-Cut-Problem*: Finde eine Partition $V = V_1 \cup V_2$ von V, so dass

$$\sum_{(i,j)\in E,\ (i,j)\in(V_1\times V_2)\cup(V_2\times V_1)} w_{ij}$$

maximal wird.

Definieren wir nun $C \in S_n$, so dass gilt

$$x^T Cx = \sum_{(i,j)\in E} w_{ij}\frac{1}{2}(x_i - x_j)^2, \tag{2.53}$$

dann ist das Max-Cut-Problem offensichtlich äquivalent zum diskreten quadratischen Programm

$$\max x^T Cx \ \text{ s.t. } \ x \in \{-1, 1\}^n.$$

Betrachte nun allgemein ein Problem der Form

$$\max x^T Cx \ \text{ s.t. } \ x \in \{-1, 1\}^n \tag{MC}$$

mit $C \in S_n$.

SDP-Relaxierung von (MC): Es gilt

$$x^T Cx = C \bullet (xx^T).$$

Nun ist für alle $x \in \{-1, 1\}^n$ die Matrix xx^T positiv semidefinit mit lauter Einsen in der Diagonale. Wir relaxieren nun xx^T zu $X \succeq 0$, $\operatorname{diag}(X) = e$, wobei $e = (1, \dots, 1)^T$.

Dies führt auf die SDP-Relaxierung

$$\max C \bullet X \ \text{ s.t. } \ \operatorname{diag}(X) = e, \quad X \succeq 0. \tag{MCR}$$

Das duale Problem hierzu ist

$$\min e^T u \ \text{ s.t. } \ Z + C - \operatorname{diag}(u) = 0, \quad Z \succeq 0 \tag{MCR-D}$$

oder äquivalent

$$\min e^T u \ \text{ s.t. } \ \lambda_{max}(C - \operatorname{diag}(u)) \le 0. \tag{MCR-D}$$

Das berühmte Ergebnis von Goemans und Williamson [GW95] besagt, dass für einen ungerichteten Graphen die SDP-Relaxation (MCR) des Max-Cut-Problems (MC), (2.53) der Optimalwert des Max-Cut-Problems mindestens das 0,878-fache des Optimalwerts der Relaxierung (MCR) ist.

Umwandlung in ein nichtglattes Problem

Wir wandeln nun das duale Problem

$$\text{(MCR-D)} \qquad \min e^T u \ \text{ s.t. } \ \lambda_{max}(C - \operatorname{diag}(u)) \leq 0.$$

in ein äquivalentes nichtglattes Problem um. Zunächst gilt $\operatorname{diag}(e) = I$ für $e = (1, \ldots, 1)^T$ und damit ist Satz 2.8.1 mit $\mu = e^T e = n$ anwendbar. Damit ist (MCR-D) im Sinne von Satz 2.8.1 äquivalent zu

$$\text{(MCR-D}'\text{)} \qquad \min_u e^T u + n\, \lambda_{max} \left(C - \operatorname{diag}(u)\right).$$

Da die Matrix C oft groß, aber dünn besetzt ist, erweist sich die Anwednung von geeigneten Varianten der Bundle-Methode auf (MCR-D′) im Vergleich zu Innere-Punkte-Verfahren für das SDP-Problem (MCR) als deutlich schneller, siehe z.B. Helmberg und Rendl [HR00].

Viele andere kombinatorische Optimierungsprobleme (z.B. Quadratic Assignment Problem, Graph Coloring Problem, Maximal Clique Problem, ...) erlauben ebenfalls gute SDP-Relaxationen.

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

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

3.1 Beispiele

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.

3.2 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

$$F(x) = 0, \tag{3.1}$$

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

Algorithmus 7 (Newton-artiges Verfahren)

0. Wähle $x^0 \in \mathbb{R}^n$.

Für $k = 0, 1, 2, \ldots$:

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

Bemerkung 3.2.2 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 7 erzeugt. Wir erhalten dann

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

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

Satz 3.2.3 *Sei x^* eine Lösung von* (3.1)*. Algorithmus* 7 *erzeuge die Folgen (M_k), (s^k) und (x^k), $x^k \to x^*$. Dann gilt:*

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

$$\|M_k^{-1}\big(F(x^k) - F(x^*) - M_k(x^k - x^*)\big)\| = o(\|x^k - x^*\|) \quad \textit{für } k \to \infty. \tag{3.3}$$

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

$$\|M_k^{-1}\big(F(x^k) - F(x^*) - M_k(x^k - x^*)\big)\| = O(\|x^k - x^*\|^2) \quad \textit{für } k \to \infty. \tag{3.4}$$

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. □

In Algorithmus 7 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 3.2.4 In Algorithmus 7 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 3.2.5 Später werden wir für M_F verallgemeinerte Differentiale von F verwenden. Dies sind mengenwertige Abbildungen, was unsere Wahl eines mengenwertigen M_F begründet.

Satz 3.2.6 *Sei $x^* \in U$ eine Lösung von* (3.1) *und es gelte Voraussetzung 3.2.4. Weiter gebe es $\eta > 0$, so dass gilt:*

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

a) *Gilt*

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

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*

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

so ist die Konvergenzrate q-quadratisch.

Beweis:

zu a):

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

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

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)

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

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

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

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 $x^k \to x^*$ und (3.9) liefert q-superlineare Konvergenzrate.

zu b):

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

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

Regularitätsbedingung:

(3.11)

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

Approximationsbedingung:

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

Für q-quadratische Konvergenz benötigen wir die stärkere

Quadratische Approximationsbedingung:

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

Definition 3.2.7 Erfüllt M_F die Approximationsbedingung (3.12), so nennen wir M_F *punktbasierte Approximation* von F bei x^*. Gilt sogar (3.13), dann nennen wir M_F *punktbasierte Approximation der Ordnung* 1 von F bei x^*.

Lemma 3.2.8 *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 $M \in M_F(x^* + s)$:

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

so dass (3.6) nun aus (3.12) und (3.7) aus (3.13) folgt. □

Als unmittelbare Konsequenz erhalten wir:

Korollar 3.2.9 *Die Aussagen von Satz 3.2.6 gelten auch, wenn wir die Bedingungen* (3.5), (3.6) *und* (3.7) *durch die Bedingungen* (3.11), (3.12) *und* (3.13) *ersetzen.*

3.2.1 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 3.2.9.

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

3.3 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 $F : U \subset \mathbb{R}^n \to \mathbb{R}^m$.

3.3.1 Einige wichtige Hilfsmittel

Satz von Rademacher

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

Satz 3.3.1 (Rademacher) *Sei $F : U \subset \mathbb{R}^n \to \mathbb{R}^m$ lokal Lipschitz-stetig auf der offenen, nichtleeren 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: $\mu(U \setminus U_d) = 0$.*

Konvexe Hülle, Satz von Carathéodory

Wir erinnern daran, dass die *konvexe Hülle* conv A einer Menge $A \subset \mathbb{R}^n$ die kleinste konvexe Menge ist, die A enthält:

$$\operatorname{conv} A = \bigcap_{\substack{C \text{ konvex} \\ A \subset C}} C = \left\{ \sum_{i=1}^{l} \lambda_i x_i \;:\; l \geq 1,\;\; x_i \in A,\;\; \lambda_i \geq 0,\;\; \sum_{i=1}^{l} \lambda_i = 1 \right\}.$$

Der folgende Satz besagt, dass wir die soeben angegebene Darstellung der konvexen Hülle auf Linearkombinationen der Länge $\leq n+1$ beschränken können:

Satz 3.3.2 (Carathéodory) *Sei $A \subset \mathbb{R}^n$ gegeben. Dann gilt:*

$$\operatorname{conv} A = \left\{ \sum_{i=1}^{n+1} \lambda_i x_i \;:\; x_i \in A,\;\; \lambda_i \geq 0,\;\; \sum_{i=1}^{n+1} \lambda_i = 1 \right\}.$$

Beweis:

Die Inklusion $\supset$ in der ersten Mengengleichheit ist klar. Zum Nachweis von $\subset$ nutzen wir, dass jedes $x \in \operatorname{conv} A$ als endliche Konvexkombination $x = \sum_{i=1}^{l} \lambda_i x_i$ von Punkten $x_i \in A$ geschrieben werden kann (o.E. seien alle λ_i positiv). Ist $l \leq n+1$, so ist nichts zu tun. Sonst sind die Vektoren $y_i = \lambda_i \binom{x_i}{1}$ linear abhängig und es gibt daher $\mu \in \mathbb{R}^l$ mit $\mu \neq 0$ und $\sum_{i=1}^{l} \mu_i y_i = 0$. Insbesondere ist $\sum_{i=1}^{l} \mu_i \lambda_i = 0$. Wähle j mit $\mu_j = \min \mu_i < 0$ und $\lambda_i^* = (1 - \mu_i/\mu_j)\lambda_i$. Dann gilt $\mu_i/\mu_j \leq 1$ und daher wegen $\lambda_i \geq 0$, so haben wir

$$\lambda_i^* = (1 - \mu_i/\mu_j)\lambda_i \geq 0.$$

Darüber hinaus ist $\lambda_j^* = 0$. Schließlich gilt

$$\sum_{i=1}^{l} \lambda_i^* = \sum_{i=1}^{l} \lambda_i - \frac{1}{\mu_j} \sum_{i=1}^{l} \mu_i \lambda_i = 1 - 0 = 1,$$

$$\sum_{i \neq j} \lambda_i^* x_i = \sum_{i=1}^{l} \lambda_i^* x_i = \sum_{i=1}^{l} \lambda_i x_i - \frac{1}{\mu_j} \sum_{i=1}^{l} \mu_i \lambda_i x_i = x - 0 = x.$$

Wir haben somit x als Konvexkombination von $l-1$ Punkten dargestellt. Diese Reduktion lässt sich fortsetzen bis gilt $l \leq n+1$.

Die Darstellung mit genau $n+1$ Summanden lässt sich erzielen durch Aufsplitten eines Summanden $\lambda_i x_i$ in mehrere. $\square$

Aus dem Satz von Carathéodory ergibt sich:

Satz 3.3.3 *Ist $A \subset \mathbb{R}^n$ kompakt, so ist* conv A *ebenfalls kompakt.*

Beweis:

Die Menge $K = \{(\lambda_1, \ldots, \lambda_{n+1}, x_1, \ldots, x_{n+1}) \,:\, \lambda_i \geq 0, \;\; \sum \lambda_i = 1, \;\; x_i \in A\}$ ist abgeschlossen und beschränkt, also kompakt. Weiter ist nach dem Satz von Carathéodory die Menge conv A das Bild von K unter der stetigen Abbildung $(\lambda_1, \ldots, \lambda_{n+1}, x_1, \ldots, x_{n+1}) \mapsto \sum \lambda_i x_i$ und daher ebenfalls kompakt. $\square$

3.3.2 Einige Verallgemeinerte Differentiale

Die folgenden verallgemeinerten Differentiale sind von fundamentaler Bedeutung:

Definition 3.3.4 Sei $F : U \subset \mathbb{R}^n \to \mathbb{R}^m$ 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

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

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 $\partial_B F(x)$:

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

Beispiel 3.3.1 Wir betrachten $F : \mathbb{R} \to \mathbb{R}$, $F(x) = |x|$.
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 $x \neq 0$:

$$x_k \to x \quad \Longrightarrow \quad F'(x_k) \to F'(x) \quad \Longrightarrow \quad \partial_B F(x) = \{F'(x)\} = \begin{cases} \{-1\} & \text{für } x < 0, \\ \{1\} & \text{für } x > 0. \end{cases}$$

Weiter gilt für $(x_k) \subset U_d$, $x_k \to 0$,

$$F'(x_k) = \begin{cases} -1 & \text{für } x_k < 0, \\ 1 & \text{für } x_k > 0. \end{cases}.$$

Im Falle $F'(x_k) \to M$ gilt $M = -1$ oder $M = 1$ und beide Limiten lassen sich offensichtlich erzeugen (durch $x_k \to 0^-$ bzw. $x_k \to 0^+$). Somit gilt $\partial_B F(0) = \{-1, 1\}$.

Damit folgt:

$$\partial^{cl} F(x) = \text{conv}\, \partial_B F(x) = \begin{cases} \{-1\} & \text{für } x < 0, \\ [-1, 1] & \text{für } x = 0, \\ \{1\} & \text{für } x > 0. \end{cases}$$

Beispiel 3.3.2 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 ϕ_{FB} nach dem Mittelwertsatz Lipschitz-stetig mit $L = 1 + \sqrt{2}$, denn entlang jeder Geraden ist ϕ_{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 $(a_k, b_k) \to (0, 0)$ genau alle Vektoren der Form $(1, 1) - v^T$, $\|v\| = 1$, sind. Somit gilt

$$\partial_B \phi_{FB}(0, 0) = \left\{(1, 1) - v^T \,:\, \|v\| = 1\right\}.$$

Beispiel 3.3.3 Ist $F : U \subset \mathbb{R}^n \to \mathbb{R}^m$ stetige Auswahl der C^1-Funktionen $F^r : U \subset \mathbb{R}^n \to \mathbb{R}^m$, $r = 1, \ldots, l$, ist also F stetig mit

$$F(x) \in \left\{F^1(x), \ldots, F^r(x)\right\} \quad \forall\, x \in \mathbb{R},$$

so gilt

$$\partial_B F(x) = \left\{F^{r\prime}(x) \,:\, r \in I_e(x)\right\},$$

wobei $I_e(x)$ die Menge der in x *essentiell aktiven Indizes* ist:

$$I_e(x) = \big\{r \; : \; x \in \mathrm{cl}\big(\mathrm{int}\,\{y \in U \; : \; F^r(y) = F(y)\}\big\}\big).$$

Der Nachweis erfordert einige Arbeit, siehe Scholtes [Sch94] und wird hier nicht geführt.

Lemma 3.3.5 *Ist $F : U \subset \mathbb{R}^n \to \mathbb{R}^m$ stetig differenzierbar in einer Umgebung von $x \in U$, so gilt $\partial^{cl} F(x) = \partial_B F(x) = \{F'(x)\}$.*

Beweis:

Für $x_k \to x$ existiert $F'(x_k)$ nahe bei x und es gilt $F'(x_k) \to F'(x)$ (Stetigkeit). Daher $\partial_B F(x) = \{F'(x)\}$. Offensichtlich ist $\partial^{cl} F(x) = \partial_B F(x)$. $\square$

Das Clarksche Differential verallgemeinert das konvexe Subdifferential:

Satz 3.3.6 *Sei $f : U \to \mathbb{R}$ 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 angegegeben Differentiale.

Definition 3.3.7 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 3.3.8 *Sei $F : U \subset \mathbb{R}^n \to \mathbb{R}^m$ 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 $M \in \partial^{cl} F(x^*) \supset \partial_B F(x^*)$.

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 δ-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:

$$\|F'(x)s\| = \lim_{t\to 0} \left\| \frac{F(x+ts) - F(x)}{t} \right\| \leq \lim_{t\to 0} \frac{L\|x+ts-x\|}{|t|} = L\|s\|.$$

Somit gilt

$$\|F'(x)\| \le L \quad \forall\, x \in U_d \cap B_\delta(x^*). \tag{3.14}$$

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 $x_k \to x$ und $F'(x_k) \to M$, also

$$\|M\| = \lim_{k\to\infty} \|F'(x_k)\| \le L. \tag{3.15}$$

• $\partial_B F(x^*) \ne \emptyset$:

Da U_d dicht in U ist, gibt es $(x_k) \subset U_d \cap B_\delta(x^*)$ mit $x_k \to x^*$. Wegen (3.14) gibt es eine Teilfolge $(x_{k'})$ mit $F(x_{k'}) \to M$ 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 $M_k \to M$. Zu M_k finden wir nach Definiton 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 $U_d \ni x_k \to x^*$ und

$$\|F'(x_k) - M\| \le \|F'(x_k) - M_k\| + \|M_k - M\| \to 0 \quad \text{für } k \to \infty.$$

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 $B_\delta(x^*) \ni y_k \to x^*$ mit $M_k \in \partial_B F(y_k)$, $\|M_k - M\| \ge \varepsilon$ 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 $F'(x_{k'}) \to M$. Nun gilt

$$\|x_{k'} - x^*\| \le \|x_{k'} - y_{k'}\| + \|y_{k'} - x^*\| \to 0 \quad \text{für } k' \to \infty,$$

also $M \in \partial_B F(x^*)$. Weiter gilt

$$\|M_{k'} - M\| \le \|M_{k'} - F'(x_{k'})\| + \|F'(x_{k'}) - M\| \to 0 \quad \text{für } k' \to \infty.$$

Dies widerspricht $\|M_k - M\| \ge \varepsilon$.

Nun zu $\partial^{cl} F$:

Wegen $\partial^{cl} F(x^*) = \operatorname{conv} \partial_B F(x^*) \supset \partial_B F(x^*)$ ist $\partial^{cl} F(x^*)$ nichtleer und nach Satz 3.3.3 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

$\lambda_i \geq 0$, $\sum_i \lambda_i = 1$, $i = 1, \ldots, l$, 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:

$$\|M - M^*\| \leq \sum_i \lambda_i \|M_i - M_i^*\| < \varepsilon.$$

Die Aussagen über $\partial_C F$ folgen unmittelbar aus denen über $\partial^{cl} F$.

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 3.3.9 (Kettenregel) *Seien $F : U \subset \mathbb{R}^n \to \mathbb{R}^m$ und $G : V \subset \mathbb{R}^m \to \mathbb{R}^l$ lokal Lipschitz-stetig und es gelte $F(U) \subset V$. Dann ist $G \circ F : U \to \mathbb{R}^l$ lokal Lipschitz-stetig und es gilt für alle $x \in U$ und $v \in \mathbb{R}^n$:*

$$\partial^{cl}(G \circ F)(x)v \subset \text{conv}\left(\partial^{cl} G(F(x)) \partial^{cl} F(x) v\right) = \text{conv}\left(\partial^{cl} G(F(x)) \partial^{cl} F(x)\right) v. \tag{3.16}$$

Ist G stetig differenzierbar in einer Umgebung von $F(x)$, so gilt

$$\partial^{cl}(G \circ F)(x)v = G'(F(x)) \partial^{cl} F(x) v. \tag{3.17}$$

Ist F stetig differenzierbar in einer Umgebung von x, so gilt

$$\partial^{cl}(G \circ F)(x)v \subset \partial^{cl} G(F(x)) F'(x) v. \tag{3.18}$$

Im Falle $l = 1$ kann der rechts stehende Vektor v in (3.16)–(3.18) *entfallen.*

Beweis: Siehe z.B. Clarke [Cla83], §2.6. □

3.4 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 3.4.1 Die Funktion $F : U \to \mathbb{R}^m$, $U \subset \mathbb{R}^n$ offen, heißt *semiglatt* in $x \in U$, falls sie Lipschitz-stetig nahe x und richtungsdifferenzierbar in x ist und falls gilt:

$$\max_{M \in \partial^{cl} F(x+s)} \|F(x+s) - F(x) - Ms\| = o(\|s\|) \quad \text{für } s \to 0. \tag{3.19}$$

Die Bedingung (3.19) ist nichts anderes als die Forderung, dass $\partial^{cl} F$ eine punktbasierte Approximation von F in x ist.

Mit Blick auf q-quadratische Konvergenz ergibt sich daraus unmittelbar auch das folgende Konzept:

Definition 3.4.2 Die Funktion $F : U \to \mathbb{R}^m$, $U \subset \mathbb{R}^n$ offen, heißt *stark semiglatt* (oder semiglatt der Ordnung 1) in $x \in U$, falls sie Lipschitz-stetig nahe x und richtungsdifferenzierbar in x ist und falls gilt:

$$\max_{M \in \partial^{cl} F(x+s)} \|F(x+s) - F(x) - Ms\| = O(\|s\|^2) \quad \text{für } s \to 0. \tag{3.20}$$

Die Definition 3.4.1 von Semiglattheit wird zunehmend beliebter, weil sie sehr praxisnah ist. Historisch gesehen wurde Semiglattheit zunächst in der folgenden abstrakten, aber äquivalenten, Form definiert:

Definition 3.4.3 (Mifflin [Mif77], Qi [Qi93], Qi und Sun [QS93]) Die Funktion $F : U \to \mathbb{R}^m$, $U \subset \mathbb{R}^n$ offen, heißt *semiglatt* 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:

$$\ell(s) = \lim_{\substack{M \in \partial^{cl} F(x+ts') \\ s' \to s,\ t \to 0^+}} Ms'. \tag{3.21}$$

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

$$\lim_{k \to \infty} t_k = 0, \quad \lim_{k \to \infty} s^k = s \quad \text{und} \quad M_k \in \partial^{cl} F(x + t_k s^k)$$

gilt:

$$\lim_{k \to \infty} M_k s^k = \ell(s).$$

Wir stellen zunächst fest, dass C^1-Funktionen semiglatt sind.

Satz 3.4.4 *Die Funktion $F : U \to \mathbb{R}^m$, $U \subset \mathbb{R}^n$ offen, sei stetig differenzierbar in einer Umgebung von $x \in U$. Dann ist F semiglatt im Punkt x. Ist F' Lipschitz-stetig nahe x, so ist F stark semiglatt 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 3.3.5. Daraus folgt

$$\max_{M \in \partial^{cl} F(x+s)} \|F(x+s) - F(x) - Ms\| = \|F(x+s) - F(x) - F'(x+s)s\|$$
$$\leq \int_0^1 \|(F'(x+ts) - F'(x+s))s\| dt = o(\|s\|).$$

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 3.4.5 *Die Funktion $F : U \to \mathbb{R}^m$, $U \subset \mathbb{R}^n$ offen, ist semiglatt in $x \in U$ genau dann, wenn alle Komponentenfunktionen F_i semiglatt 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:

$$\max_{v \in \partial^{cl} F_i(x+s)} |F_i(x+s) - F_i(x) - vs| = \max_{M \in \partial^{cl} F(x+s)} |F_i(x+s) - F_i(x) - e_i^T Ms|$$
$$= \max_{M \in \partial^{cl} F(x+s)} |e_i^T (F(x+s) - F(x) - Ms)| \leq \max_{M \in \partial^{cl} F(x+s)} \|F(x+s) - F(x) - Ms\|$$
$$= o(\|s\|).$$

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

$$\max_{M \in \partial^{cl} F(x+s)} \|F(x+s) - F(x) - Ms\| \leq \sum_{i=1}^n \max_{M \in \partial^{cl} F(x+s)} |F_i(x+s) - F_i(x) - e_i^T Ms|$$
$$\overset{\text{s.o.}}{=} \sum_{i=1}^n \max_{v \in \partial^{cl} F_i(x+s)} |F_i(x+s) - F_i(x) - vs| = \sum_{i=1}^n o(\|s\|) = o(\|s\|).$$

□

Wir stellen nun noch einige Aussagen bereit, die hilfreich sind, um die Semiglattheit von Funktionen nachzuweisen:

Satz 3.4.6

a) *Stückweise C^1-Funktionen sind semiglatt.*

b) *Konvexe Funktionen sind semiglatt.*

c) *Summen, Produkte und Kompositionen semiglatter Funktionen sind semiglatt.*

Folgerungen:

1. Die Fischer-Burmeister-Funktion und die min-NCP-Funktion sind semiglatt, da konvex.

2. Betrachte für stetig differenzierbares $F : \mathbb{R}^n \to \mathbb{R}^n$ das NCP

   $$x \geq 0, \quad F(x) \geq 0, \quad x_i F_i(x) = 0, \;\; i = 1, \ldots, n.$$

   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 ϕ semiglatt.

3. Reformulierungen von KKT-Systemen mit Hilfe von obigen NCP-Funktionen für Optimierungsprobleme mit zweimal stetig differenzierbaren Nebenbedingungen und Zielfunktion sind semiglatt.

3.5 Semiglatte Newton-Verfahren

Zur Lösung des Gleichungssystems

$$F(x) = 0, \tag{3.22}$$

wobei $F : U \to \mathbb{R}^n$, $U \subset \mathbb{R}^n$, semiglatt in der Lösung $x^* \in U$ sei, betrachten wir nun unser allgemeines nichtglattes Newton-Verfahren 7 mit der punktbasierten Wahl $M_k \in \partial^{cl} F(x^k)$, setzen also $M_F(x) = \partial^{cl} F(x)$:

Algorithmus 8 (Semiglattes Newton-Verfahren von Qi und Sun [Qi93,QS93])

0. Wähle $x^0 \in \mathbb{R}^n$.

Für $k = 0, 1, 2, \ldots$:

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 3.5.1 *Sei $F : U \subset \mathbb{R}^n \to \mathbb{R}^m$ 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 $M_F = \partial^{cl} F$. □

Neben der durch Lemma 3.5.1 bereitgestellten Approximationsbedingung fehlt uns nun nur noch der Nachweis der Regularitätsbedingung (3.11), um die lokale q-superlineare Konvergenz von Algorithmus 8 nachzuweisen.

Wir fordern hierzu:

Voraussetzung 3.5.2 $\partial^{cl} F$ ist in der Lösung x^* CD-regulär (CD für Clarke-Differential), d.h. alle $M \in \partial^{cl} F(x^*)$ sind invertierbar.

Lemma 3.5.3 *Die Funktion $F : U \subset \mathbb{R}^n \to \mathbb{R}^n$ sei Lipschitz-stetig in einer Umgebung von $x^* \in U$ und es gelte Voraussetzung 3.5.2. Dann gibt es $\eta > 0$ und $C > 0$ mit*

$$\text{(3.23)} \qquad M \;\; \textit{invertierbar}, \;\; \|M^{-1}\| \le C \quad \forall\, M \in \partial^{cl} F(x), \;\; \forall\, x \in B_\eta(x^*).$$

Beweis: Nach Satz 3.3.8 ist $\partial^{cl} F(x^*)$ nichtleer, kompakt und $\partial^{cl} F(x^*)$ ist oberhalbstetig in x^*. Die Funktion $M \in \partial^{cl} F(x^*) \to \|M^{-1}\|$ ist nach Voraussetzung 3.5.2 wohldefiniert und bekanntlich stetig, nimmt also auf dem Kompaktum $\partial^{cl} F(x^*)$ ein Maximum $\hat{C}$ an. Es gilt also

$$\|M^{-1}\| \le \hat{C} \quad \forall\, M \in \partial^{cl} F(x^*).$$

Für jede Störung $E \in \mathbb{R}^{n,n}$ mit $\|E\| \le \frac{1}{2\hat{C}} =: \varepsilon$ gilt nach dem Banach-Lemma also auch

$$\text{(3.24)} \qquad \|(M+E)^{-1}\| \le 2\hat{C} \quad \forall\, M + E \in \partial^{cl} F(x^*) + B_\varepsilon(0).$$

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 $C = 2\hat{C}$. □

Wir erhalten den folgenden Konvergenzsatz:

Satz 3.5.4 *Sei $x^* \in U$ eine Lösung von* (3.22)*, in der $F : U \subset \mathbb{R}^n \to \mathbb{R}^n$ semiglatt ist, und es gelte Voraussetzung 3.5.2. Dann gibt es $\delta > 0$, so dass Algorithmus 8 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 8 ist identisch zu Algorithmus 7 mit punktbasierter Wahl $M_k \in \partial^{cl} F(x_k)$, d.h. $M_F = \partial^{cl} F$. In den Lemmas 3.5.1 und 3.5.3 haben wir die (quadratische) Approximationsbedingung (3.12) bzw. (3.13) und die Regularitätsbedingung (3.11) nachgewiesen. Somit ist Korollar 3.2.9 anwendbar und liefert die Behauptung. □

Bemerkung 3.5.5 Die Aussagen dieses Abschnittes 3.5 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 3.5.2 stellt dann nur noch Anforderungen an die Elemente von $\partial_B F(x^*)$ und wird daher BD-Regularität genannt (BD für B-Differential).

Literaturverzeichnis

[GK02] C. Geiger, C. Kanzow, *Theorie und Numerik restringierter Optimierungsaufgaben*, Springer-Verlag, 2002.

[GW95] M.X. Goemans und D. P. Williamson, Improved Approximation Algorithms for Maximum Cut and Satisfiability Problems Using Semidefinite Programming. *Journal of the ACM* 42, 1995, pp. 1115-1145.

[HR00] C. Helmberg and F. Rendl, A spectral bundle method for semidefinite programming, *SIAM Journal on Optimization*, 10, 2000, pp. 673–696.

[HU93a] J.-B. Hiriart–Urruty und C. Lemaréchal, *Convex analysis and minimization algorithms I*, Grundlehren der Mathematischen Wissenschaften, 305, Springer-Verlag, Berlin, 1993.

[Mif77] R. Mifflin, *Semismooth and semiconvex functions in constrained optimization*, SIAM J. Control Optim. 15 (1977) 957–972.

[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.

[Cla83] F. H. Clarke, *Optimization and nonsmooth analysis*, John Wiley & Sons, Inc., New York, 1983.