Vorlesungszyklus
"Nichtdifferenzierbare konvexe Optimierung"

Vorlesung von Prof. Dr. K. Beer im WS 2001/2002

15. Oktober 2003

Inhaltsverzeichnis

1 Beispiele von NDOA **7**
- 1.1 Einleitende Bemerkungen . . . 7
- 1.2 T-Approximation . . . 9
- 1.3 Exakte Strafen . . . 10
- 1.4 n-Personen-Spiele . . . 10
- 1.5 Ungleichungssysteme . . . 11
- 1.6 Mehrzieloptimierung . . . 11
- 1.7 Duale Dekomposition . . . 13
- 1.8 Primale Dekomposition . . . 13
- 1.9 Abschließende Bemerkungen - Erweiterte konvexe Funktionen . . . 14

2 Subdifferentialrechnung konvexer Funktionen **17**
- 2.1 Existenz der Richtungsableitung . . . 17
- 2.2 Das Subdifferential einer konvexen Funktion . . . 24
  - 2.2.1 Begriff des Subgradienten . . . 24
  - 2.2.2 Eigenschaften des Subdifferentials . . . 27
- 2.3 Die Subdifferentialabbildung . . . 39
- 2.4 Die Subdifferentialrechnung . . . 46
- 2.5 Weitere Rechenbeispiele zur Subdifferentialrechnung . . . 58
  - 2.5.1 . . . 58
  - 2.5.2 . . . 59
  - 2.5.3 . . . 60
  - 2.5.4 . . . 61

3 Optimalitätskriterien **63**
- 3.1 Die allgemeine konvexe Aufgabe . . . 63
- 3.2 Aufgaben mit einer konvexen Restriktion . . . 65
- 3.3 Aufgaben mit endlich vielen konvexen Restriktionen . . . 68

4 Lösungsverfahren für die Aufgabe des freien Optimums **73**
- 4.1 Einige nützliche Beobachtungen (Motivation) . . . 73
  - 4.1.1 Rückführung auf glatte Aufgaben . . . 73
  - 4.1.2 Ein Antisubgradient muß keine Abstiegsrichtung sein . . . 74
  - 4.1.3 Eigenschaften der Antisubgradienten . . . 76

4.1.4 Kein geeignetes Stop-Kriterium 79
4.1.5 Klassische Methoden der Schrittweitenwahl versagen . . . 79
4.1.6 Möglichkeit der Konvergenz gegen einen nichtoptimalen Punkt 80
4.2 Subgradientenverfahren von N.Z. Shor 82
4.2.1 Das Verfahren SGV und seine Konvergenz bzgl. der Zielfunktionswerte 82
4.2.2 Konvergenz bzgl. der Argumente 86
4.2.3 Negativaussage zur Konvergenzgeschwindigkeit im algemeinen Fall 89
4.2.4 Schrittweitenwahl nach B.T. Poljak 91
4.3 Die gesamte Anzahl der Seiten 97

Literaturverzeichnis

[1] *N.Z. Shor*: Metody minimizacii nedifferenciruemych funkcij i ich prilosheni-ja, Verlag Naukova dumka, Kiev 1979 (in Russisch, englische Übersetzung 1985 im Springer-Verlag erschienen)

[2] *V.F. Demyanov, L.V. Vasil'ev*: Nedifferenciruemaja optimizacija, Verlag Nauka, Moskva 1981 (in Russisch)

[3] *F.H. Clarke*: Optimization and Nonsmooth Analysis, J.Wiley & Sons, 1983

[4] *K.C. Kiwiel*: Methods of Descent for Nondifferentiable Optimization, Springer-Verlag, Berlin 1985

[5] *V.F. Demyanov, A.M. Rubinov*: Osnovy negladkogo analiza i kvazidifferencial'noje iscislenije, Verlag Nauka, Moskva 1990 (in Russisch)

[6] *M.M. Mäkelä, P. Neittaanmäki*: Nonsmooth Optimization, Verlag World Scientific, Singapore 1992

[7] *Kiyotaka Shimizu, Yo Ishizuka, J.F. Bard*: Nondifferentiable and Two-Level Mathematical Programming, Verlag Kluwer Academic Publ., Boston 1997

Kapitel 1

Beispiele von NDOA

1.1 Einleitende Bemerkungen

Wir sagen, daß eine mathematische Optimierungsaufgabe im $\mathbb{R}^n$ gegeben ist, wenn gegeben sind

$$\begin{cases} S \subset \mathbb{R}^n \\ f : S \to \mathbb{R} \end{cases}$$

und gesucht ist ein $\bar{x} \in S$, so daß $f(\bar{x}) \leq f(x)$, $\forall x \in S$.

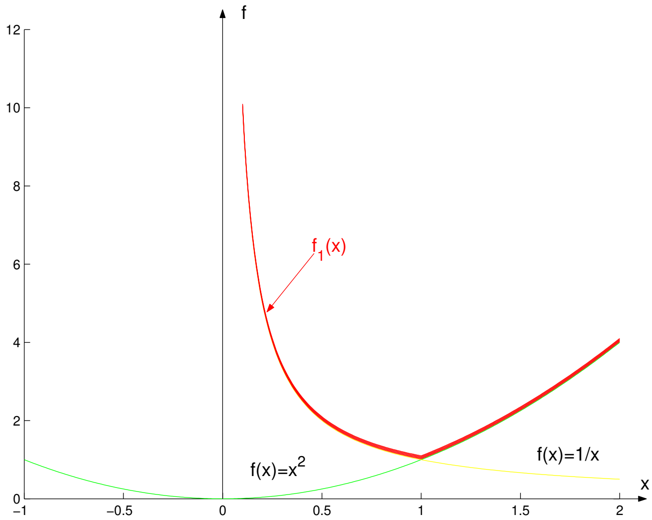
Uns interessiert der Fall, daß f konvex, aber im allgemeinen nicht in jedem Punkt von S differenzierbar ist. In der Regel werden wir den Fall zulassen, daß S auch über nichtdifferenzierbare Funktionen g_i definiert ist, z.B. nach $S = \{x \in \mathbb{R}^n : g_i(x) \leq 0, i = 1, 2, ..., m\}$. Wir beschränken uns aber auf den Fall der Kovexität von S. Aus Gründen wie sie z.B. aus den Beispielen der Abschnitte 1.7 und 1.8 folgen, wollen wir bewußt den Fall nicht von vornherein ausschließen, daß $S \cap \mathrm{dom} f \neq S$, d.h. die Möglichkeit, daß die Funktion f in gewissen Punkten $x \in S$ nicht definiert ist, soll erlaubt sein. Auch die Möglichkeit, daß f in Punkten $x \in S \setminus \mathrm{ri\ dom} f$ unstetig ist, werden wir unter Umständen zulassen.

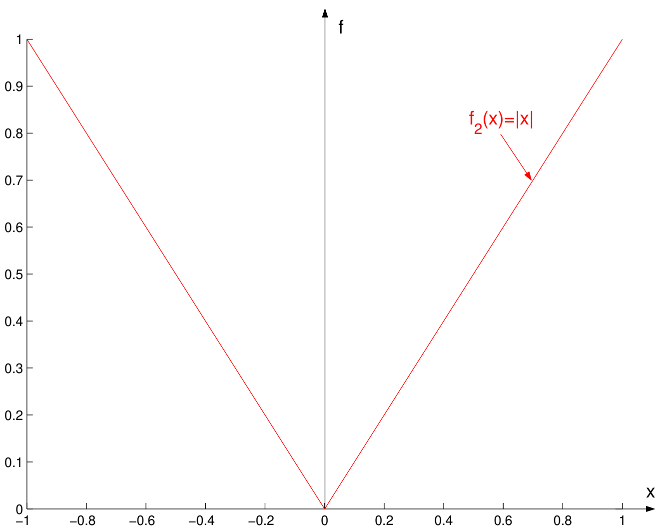
Nichtdifferenzierbare Funktionen entstehen meist in Verbindung mit einer Maximums - oder Minimumsbildung über einer Schar von (häufig glatten) Funktionen, also über eine Hüllenoperation (obere bzw. untere Einhüllende der Funktionenschar). Wie ein Blick auf die unten folgenden Beispiele zeigt, haben die entstehenden Einhüllenden häufig die Eigenschaft, auf Teilmengen der Dimension kleiner oder gleich $(n-1)$ von $\mathrm{dom} f$ nicht differenzierbar zu sein. Das Wort "nichtdifferenzierbar" soll also nicht mit "nirgendwo differenzierbar" interpretiert werden.

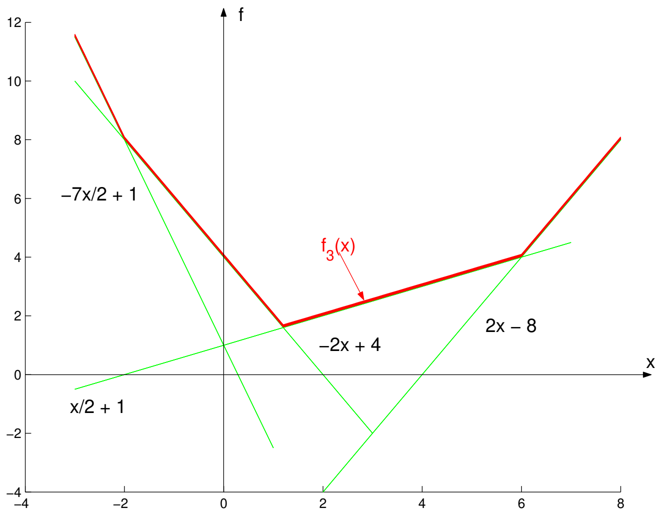
Betrachten wir zu Beginn einige einfache Beispiele $(n = 1)$ nichtdifferenzierbarer Funktionen.

Es sei

$$
\begin{aligned}
f_1(x) &= \max\{\frac{1}{x}, x^2\},\\
f_2(x) &= |x| = \max\{x, -x\},\\
f_3(x) &= \max\{-\frac{7}{2}x + 1, -2x + 4, \frac{1}{2}x + 1, 2x - 8\}.
\end{aligned}
$$







In weiteren sollen einige Beispiele von nichtdifferenzierbaren Optimierungsaufgaben (NDOA) betrachtet werden.

1.2 T-Approximation

Es sei $M \subset \mathbb{R}^n$, $f : M \to \mathbb{R}$ und die Funktion f solle über der Menge M bestmöglich im Sinne der Tschebyscheff-Norm durch eine Linearkombination von gegebenen Basisfunktionen g_i, $i = 1, 2, ..., m$, approximiert werden.

Wir setzen $h(x) = \max\{|f(y) - \sum_{i=1}^m x_i g_i(y)| : y \in M\}$. Die formulierte Zielstellung führt dann auf die Optimierungsaufgabe (OA)

$$(TA) \begin{cases} h(x) \to \min \\ x \in \mathbb{R}^m \end{cases}$$

Für jedes konkrete $y \in M$ sind

$$f(y) - \sum_{i=1}^m x_i g_i(y)$$

und

$$-f(y) + \sum_{i=1}^m x_i g_i(y)$$

affin-lineare Funktionen in x. Deshalb sind (vgl. Satz 2.25 der Grundvorlesung "Optimierung I") sowohl

$$h_1(x) = \max_{y \in M}\{f(y) - \sum_{i=1}^m x_i g_i(y)\}$$

als auch

$$h_2(x) = \max_{y \in M}\{-f(y) + \sum_{i=1}^{m} x_i g_i(y)\}$$

konvexe Funktionen und damit ist auch $h(x) = \max\{h_1(x), h_2(x)\}$ eine konvexe Funktion. Es sei $h(x) = h_i(x)$. Immer dann, wenn $h_1(x) = h_2(x)$ oder bei $h_1(x) \neq h_2(x)$ das $h_i(x)$ nicht eindeutig auf einem y realisiert wird, ist zu vermuten, daß x eine Nichtdifferenzierbarkeitsstelle von h ist (insbesondere wenn M eine diskrete Menge ist).

1.3 Exakte Strafen

In einigen Anwendungen der Optimierung möchte man sich von gewissen Nebenbedingungen befreien, indem man ihre Verletzung in der Zielfunktion mit einer Strafe belegt. In diesem Zugang ersetzt man die Ausgangsaufgabe

$$(OA) \begin{cases} f(x) \to \min \\ g_i(x) \leq 0, i = 1, 2, ..., m \\ x \in P \end{cases}$$

durch die Aufgabe

$$(E) \begin{cases} F(x) \to \min \\ x \in P, \end{cases}$$

wobei

$$\begin{cases} F(x) := f(x) + \sum_{i=1}^{m} r_i \max\{0, g_i(x)\} \\ r_i > 0, \text{ passend gewählte Konstante}, i = 1, 2, ..., m \end{cases}$$

Unter gewissen Regularitätsvoraussetzungen kann man zeigen, daß alle lokal-optimalen Lösungen von (OA) auch lokal-optimal in (E) sind. In gewissen Fällen gelingt es zu zeigen, daß die optimalen Lösungen von (OA) und (E) übereinstimmen(vgl. Satz ??5.1 in diesem Skript). Wenn f und g_i konvex sind, dann ist F ebenfalls konvex (vgl. Sätze 2.24 und 2.25 der Grundvorlesung "Optimierung I").

1.4 n-Personen-Spiele

In der nichtkooperativen Spieltheorie wird das sogenannte Nash-Gleichgewicht als Lösungskonzept diskutiert. Hier sind die Strategienmengen $S_i, i = 1, 2, ..., n$, der Spieler und ihre individuellen Gewinnfunktionen $H_i : S_1 \times S_2 \times ... \times S_n \to \mathbb{R}$, $i = 1, 2, ..., n$, gegeben und gesucht wird ein Punkt $\bar{x} = [\bar{x}_1, \bar{x}_2, ..., \bar{x}_n] \in \prod_{i=1}^{n} S_i$, so daß für $i = 1, 2, ..., n$ gilt

$$H_i(\bar{x}) \geq H_i(\bar{x}_1, ..., \bar{x}_{i-1}, x_i, \bar{x}_{i+1}, ..., \bar{x}_n), \forall x_i \in S_i.$$

In vielen Fällen läßt sich das Aufsuchen solch eines Punktes $\bar{x}$ auf das Lösen von Optimierungsaufgaben zurückführen.

Im Falle der sogenannten 2-Personen-Nullsummenspiele ($n = 2$, $H_2(x_1, x_2) = -H_1(x_1, x_2)$) läßt sich $\bar{x} = [\bar{x}_1, \bar{x}_2]$ (wenn seine Existenz gesichert ist) durch Lösen der beiden Optimierungsaufgaben

$$(P_1) \begin{cases} f_1(x_1) \to \max \\ x_1 \in S_1 \end{cases} \quad \text{und } (P_2) \begin{cases} f_2(x_2) \to \max \\ x_2 \in S_2 \end{cases}$$

berechnen. Hierbei bezeichnet

$$\begin{aligned} f_1(x_1) &= \min\{H_1(x_1, x_2) : x_2 \in S_2\} \\ f_2(x_2) &= \min\{H_2(x_1, x_2) : x_1 \in S_1\}. \end{aligned}$$

(P_1) und (P_2) sind NDOA.

Wenn die S_i konvexe Mengen sind und H_1 bzgl. x_2 konvex über S_2 und bzgl. x_1 konkav über S_1 ist, dann sind (P_1) und (P_2) konvexe Optimierungsaufgaben.

1.5 Ungleichungssysteme

Gegeben seien $S \subset \mathbb{R}^n$, $f_i : S \to \mathbb{R}$, $i = 1, 2, ..., m$.
Gesucht sei eine Lösung des Systems

$$(U) \begin{cases} f_i(x) \leq 0, i = 1, 2, ..., m \\ x \in S. \end{cases}$$

Wir setzen $f(x) = \max_{1 \leq i \leq m} \max\{f_i(x), 0\}$ und betrachten das Ersatzproblem

$$(E) \;\; f^* = \inf\{f(x) : x \in S\}.$$

(E) ist eine konvexe NDOA, wenn S eine konvexe Menge und die f_i auf S konvexe Funktionen sind. Wenn $f^* > 0$, dann ist (U) inkonsistent. Wenn $f^* = 0$ und x^* eine optimale Lösung von (E) ist, dann liefert x^* eine Lösung von (U).

1.6 Mehrzieloptimierung

Gegeben sei eine Menge $S \subset \mathbb{R}^n$ und m Zielfunktionen $f_i : S \to \mathbb{R}$, $i = 1, 2, ..., m$. Gesucht ist eine Lösung der Aufgabe

$$(MZ) \begin{cases} f_1(x) \to \min \\ f_2(x) \to \min \\ \vdots \\ f_m(x) \to \min \\ x \in S. \end{cases}$$

In der Theorie der Mehrzieloptimierung werden verschiedene Zugänge diskutiert, die näher klären, was man als Lösung von (MZ) ansehen könnte.

Wenn wir setzen

$$\begin{aligned} f_i^* &= \sup\{f_i(x) : x \in S\} \\ f_i^{**} &= \inf\{f_i(x) : x \in S\} \\ \bar{f}_i(x) &= \frac{f_i(x) - f_i^{**}}{f_i^* - f_i^{**}} \\ f(x) &= \max_{1 \le i \le m} \bar{f}_i(x), \end{aligned}$$

dann ist

$$(E1) \begin{cases} f(x) \to \min \\ x \in S \end{cases}$$

eine sinnvolle Ersatzaufgabe für (MZ). In ihr sind die f_i skaliert und bzgl. der Maßeinheit (Geld, Gewicht, Stückzahl, ...) untereinander vergleichbar.

Eine weitere weit akzeptierte Möglichkeit, für (MZ) ein Ersatzproblem zu definieren, benutzt den Zusammenhang zwischen den sogenannten effizienten Lösungen von (MZ) und den optimalen Lösungen von

$$(E2) \ \ \varphi(y) = \min\{\sum_{i=1}^{m} y_i f_i(x) : x \in S\}.$$

Hier seien $y_1 > 0$, $y_2 > 0$, ..., $y_m > 0$ gewisse (fixierte) Wichtungsfaktoren mit $\sum_{i=1}^m y_i = 1$.

Da die Wahl dieser Faktoren sich häufig problematisch gestaltet, kann man folgenden (pessimistischen) Standpunkt einnehmen: Wir geben uns eine Menge $Y \subset \{y \in \mathbb{R}^m : y \ge 0, \sum_{i=1}^m y_i = 1\}$ vor und lösen das Problem

$$(E3) \begin{cases} \varphi(y) \to \max \\ y \in Y, \end{cases}$$

genauer: Wir starten einen Dialog zwischen dem Entscheidungsträger (der Y wählt), und dem Computer (der $(E3)$ löst). Im Ergebnis dieses Dialogs kann ein Kompromiss in der Wahl der Gewichte y und der zugehörigen Lösung x^* von $(E2)$ gefunden werden.

Wir wollen noch eine Möglichkeit betrachten, für (MZ) ein Ersatzproblem einzuführen. Hier hat der Entscheidungsträger obere Schranken z_i auf die erlaubten Werte der i-ten Zielfunktion, $i = 1, 2, ..., m$, zu fixieren. Bei gewählten Gewichten y_i, $i = 1, 2, ..., m$, (z.B. nach Lösen von $(E3)$) wird die Aufgabe

$$(E4) \ \ \psi(z) = \min\{\sum_{i=1}^{m} y_i f_i(x) : x \in S, f_i(x) \le z_i, i = 1, 2, ..., m\}$$

gelöst. Wenn man hier die Schranken nicht exakt fixieren möchte, sondern aus einer Menge $Z \subset \mathbb{R}^m$ wählen will, so ist die Aufgabe (optimistischer Standpunkt)

$$(E5) \begin{cases} \psi(z) \to \min \\ z \in Z, \end{cases}$$

sinnvoll. Auf der Basis von $(E5)$ ist wiederum ein Dialog zwischen dem Z wählenden Entscheidungsträger und dem $(E5)$ lösenden Computer möglich.

Wenn die f_i konvexe Funktionen und S, Y, Z konvexe Mengen sind, dann sind $(E1)$, $(E3)$, sowie $(E5)$ konvexe NDOA.

1.7 Duale Dekomposition

Gegeben sei die Optimierungsaufgabe

$$(OA) \begin{cases} f(x) \to \min \\ g(x) \leq 0 \\ h(x) = 0 \\ x \in P \end{cases}$$

mit $P \subset \mathbb{R}^n$, $f : P \to \mathbb{R}$, $g : P \to \mathbb{R}^m$, $h : P \to \mathbb{R}^q$.

Ihr kann man die Lagrange-Duale

$$(D) \begin{cases} \varphi(u, v) \to \max \\ [u, v] \in Q \end{cases}$$

zuordnen, wobei

$$\begin{array}{l} \varphi(u, v) = \inf\{f(x) + < u, g(x) > + < v, h(x) >: x \in P\} \\ Q = \{[u, v] \in \mathbb{R}^m_+ \times \mathbb{R}^q : \varphi(u, v) > -\infty\}. \end{array}$$

In einigen Lösungsverfahren der Diskreten Optimierung sowie bei großen strukturierten Optimierungsaufgaben nutzt man häufig die Möglichkeit, durch Lösen von (D) wertvolle Information über (OA) zu gewinnen.

(D) ist stets eine konvexe NDOA.

1.8 Primale Dekomposition

Gegeben sei eine Optimierungsaufgabe, in der die Unbekannten in 2 Gruppen x und y aufgeteilt sind:

$$(OA) \begin{cases} f(x, y) \to \min \\ g(x, y) \leq 0 \\ h(x, y) = 0 \\ x \in X, y \in Y \end{cases}$$

wobei

$$\begin{array}{l} f : X \times Y \to \mathbb{R} \\ g : X \times Y \to \mathbb{R}^m \\ h : X \times Y \to \mathbb{R}^n \\ X \subset \mathbb{R}^p, Y \subset \mathbb{R}^q. \end{array}$$

In strukturierten Aufgaben (OA) kann es sinnvoll sein, (OA) in einem 2-Ebenen-Zugang einer Lösung zuzuführen.

Auf der unteren Ebene werden für fixierte y Teilaufgaben (T)

$$(T)\ \ \varphi(y) = \inf\{f(x,y) : g(x,y) \le 0, h(x,y) = 0, x \in X\}$$

bearbeitet, während auf der oberen Ebene die mit (OA) äquivalente Aufgabe

$$(E) \begin{cases} \varphi(y) \to \min \\ y \in Y \cap \mathrm{dom}\varphi \end{cases}$$

bearbeitet wird.

(E) ist eine NDOA, die eine konvexe Aufgabe ist, wenn X und Y konvexe Mengen und f, g über $X \times Y$ konvexe Funktionen sind und h affinlinear ist.

1.9 Abschließende Bemerkungen - Erweiterte konvexe Funktionen

Wie eingangs bereits hervorgehoben, wollen wir nur konvexe NDOA betrachten. Es soll aber erwähnt werden, daß alle nachfolgenden Abschnitte (Subdifferentialrechnung, Optimalitätskriterien, Lösungsverfahren) auch für Aufgaben mit Funktionen, die eine Oberklasse der Menge der konvexen Funktionen bilden, den sogenannten lokal Lipschitzstetigen Funktionen, betrachtet werden können.

Wir wollen hier für das weitere noch einige Begriffe aus der konvexen Analysis vereinbaren.

Definition. Es bezeichne $\bar{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$ die um das Element $+\infty$ erweiterte reelle Achse. Unter $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ werde verstanden, daß in einem konkreten Punkt $x \in \mathbb{R}^n$ der Funktionswert $f(x)$ entweder eine endliche reelle Zahl oder das Element $+\infty$ ist.

Für solche f bezeichne

$$\mathrm{dom} f = \{x \in \mathbb{R}^n : f(x) < +\infty\}$$

den sogenannten eigentlichen Definitionsbereich von f.

Bemerkung.

1. Wenn $\mathrm{dom} f \neq \varnothing$, dann sagen wir, daß f eine eigentliche Funktion ist.

2. Es sei hervorgehoben, daß die oben betrachteten Funktionen f in keinem Punkt des $\mathbb{R}^n$ den Wert $-\infty$ annehmen sollen.

   Das läßt aber immer noch die Möglichkeit offen, daß $\inf\{f(x) : x \in \mathbb{R}^n\} = -\infty$.

3. Der Fall, daß f nur auf einer Menge $S \subset \mathbb{R}^n$ eingeführt wurde, also z.B. $f : S \to \bar{\mathbb{R}}$, ist stets auf den oben betrachteten Fall rückführbar.

   So ist die Aufgabe

$$\begin{cases} f(x) \to \min \\ x \in S \end{cases}$$

mit $f : S \to \bar{\mathbb{R}}$, äquivalent mit der Aufgabe

$$\begin{cases} F(x) \to \min \\ x \in \mathbb{R}^n, \end{cases}$$

wenn wir setzen

$$F(x) = \begin{cases} f(x), x \in S \\ +\infty, x \notin S. \end{cases}$$

Für Funktionen $f : \mathbb{R}^n \to \mathbb{R}$ ist es üblich, von einer konvexen Funktion zu sprechen, wenn aus $x,\ y \in \mathbb{R}^n$ und $\lambda \in [0,1]$ stets folgt, daß

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y). \tag{1.9.1}$$

Wir wählen für $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ eine völlig analoge Definition.

Definition. Es sei $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ eine eigentliche Funktion. Wir sagen, daß f im $\mathbb{R}^n$ konvex ist, wenn aus $x,\ y \in \mathbb{R}^n$ und $\lambda \in (0,1)$ stets folgt, daß (1.9.1) gilt.

Bemerkung.

1. Wenn bei einem $\lambda \in (0,1)$ gilt $f(\lambda x + (1-\lambda)y) = +\infty$, dann muß nach (1.9.1) bei konvexem f wenigstens einer der Werte $f(x)$ und $f(y)$ gleich $+\infty$ sein.

2. Bei konvexem $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ kann in (1.9.1) nach unseren obigen Vereinbarungen nie ein unbestimmter Ausdruck der Art $0 \cdot \infty$ oder $+\infty - \infty$ entstehen.

3. Wenn in späteren Manipulationen mit konvexen $f, g : \mathbb{R}^n \to \bar{\mathbb{R}}$ ein Ausdruck der Art $0 \cdot \infty$ bzw. $\infty \cdot 0$ oder $+\infty - \infty$ entsteht, dann wollen wir per Definition die ersten beiden als 0 und den letzten als undefiniert interpretiert wissen.

4. Überzeugen Sie sich, daß nach unseren Vereinbarungen aus $f : \mathbb{R}^n \to \bar{\mathbb{R}}$, f - konvex, stets folgt, daß $\mathrm{dom} f$ eine konvexe Menge ist, d.h aus $x,\ y \in \mathrm{dom} f$ und $\lambda \in [0,1]$ folgt für solche f stets $(\lambda x + (1-\lambda)y) \in \mathrm{dom} f$.

Kapitel 2

Subdifferentialrechnung konvexer Funktionen

2.1 Existenz der Richtungsableitung

Definition. Es sei $f : \mathbb{R}^n \to \bar{\mathbb{R}}$. Sei $x \in \mathrm{dom} f$, sei $r \in \mathbb{R}^n$. Wenn

$$\lim_{\alpha \to +0} \frac{f(x + \alpha r) - f(x)}{\alpha}$$

existiert, dann sagen wir, daß f in x in Richtung r richtungsableitbar ist. Die Zahl

$$f'(x;r) := \lim_{\alpha \to +0} \frac{f(x + \alpha r) - f(x)}{\alpha}$$

heißt Richtungsableitung von f im Punkte x in Richtung r.

Bemerkung.

1. In der Bezeichnung $f'(x;r)$ schreiben wir ein Semikolon zwischen dem Punkt und der Richtung.
2. Was bedeutet "$\lim_{\alpha \to +0} \frac{f(x+\alpha r) - f(x)}{\alpha}$" existiert? In Anwendung auf unseren Fall könnte man es so ausdrücken:

   Es existiert eine (endliche) Zahl $f'(x;r) \in \mathbb{R}$, so daß zu beliebigem $\varepsilon > 0$ und beliebiger Folge $\{\alpha_k\}_{k=1}^{\infty}$ mit $\alpha_k \to +0$ bei $k \to \infty$ gilt: Es existiert ein k_0, so daß

   $$\left| \frac{f(x + \alpha_k r) - f(x)}{\alpha_k} - f'(x;r) \right| < \varepsilon,$$

   wenn nur $k \geq k_0$.
3. Wenn $x \in \mathrm{int\ dom} f$ und f in x differenzierbar ist, dann existiert $f'(x;r)$ für beliebiges $r \in \mathbb{R}^n \setminus \{0\}$ und es ist

   $$f'(x;r) = < \nabla f(x), r > .$$

Beweis. Da f in x differenzierbar ist, so gilt

$$\lim_{\|u\|\to 0} \frac{f(x+u)-f(x)-<\nabla f(x), r>}{\|u\|} = 0.$$

Mit $u := \alpha \cdot r$ erhalten wir

$$\lim_{\alpha\to +0} \frac{f(x+\alpha r)-f(x)-\alpha<\nabla f(x), r>}{\alpha\|r\|} = 0,$$

woraus sich ergibt

$$\lim_{\alpha\to +0} \frac{f(x+\alpha r)-f(x)}{\alpha} =< \nabla f(x), r> .$$

□

4. Welcher Zusammenhang besteht zwischen Richtungsableitung und partieller Ableitung? Wenn $f : \mathbb{R}^n \to \mathbb{R}$ in x die i-te partielle Ableitung $\frac{\partial f(x)}{\partial x_i}$ besitzt, d.h. der Grenzwert

$$\lim_{\alpha\to 0} \frac{f(x+\alpha e^i)-f(x)-\alpha\frac{\partial f(x)}{\partial x_i}}{|\alpha|} = 0 \tag{2.1.1}$$

existiert, dann existieren auch die beiden Richtungsableitungen $f'(x; e^i)$ und $f'(x; -e^i)$ und es gilt

$$f'(x; e^i) = \frac{\partial f(x)}{\partial x_i}, \;\; f'(x; -e^i) = -\frac{\partial f(x)}{\partial x_i}.$$

(e^i bezeichnet den i-ten Einheitsvektor $(0, ..., 0, 1, 0, ..., 0) \in \mathbb{R}^n$).

Beweis. Nach (2.1.1) existiert

$$\lim_{\alpha\to +0} \frac{f(x+\alpha e^i)-f(x)}{\alpha}$$

und ist gleich $\frac{\partial f(x)}{\partial x_i}$.
Außerdem existiert nach (2.1.1)

$$\lim_{\alpha\to -0} \frac{f(x+\alpha e^i)-f(x)-\alpha\frac{\partial f(x)}{\partial x_i}}{|\alpha|}$$

und ist gleich Null.
Mit der Substitution $\alpha := -\alpha$ existiert damit

$$\lim_{\alpha\to +0} \frac{f(x-\alpha e^i)-f(x)+\alpha\frac{\partial f(x)}{\partial x_i}}{|-\alpha|}$$

und ist gleich Null, d.h. es existiert

$$\lim_{\alpha \to +0} \frac{f(x + \alpha(-e^i)) - f(x)}{\alpha}$$

und ist gleich $-\frac{\partial f(x)}{\partial x_i}$. □

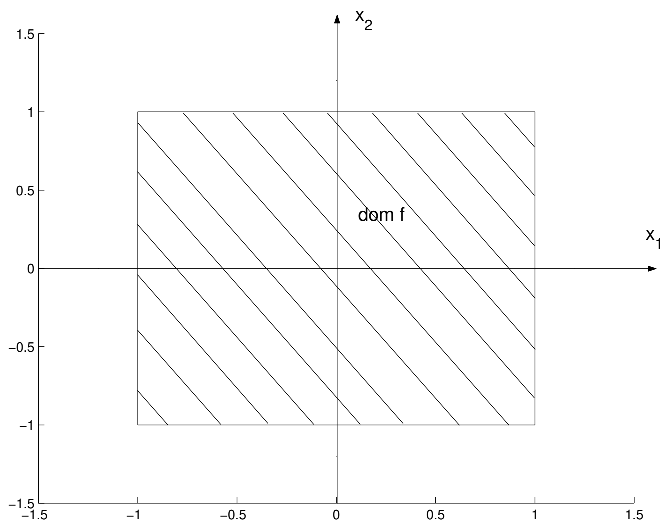
5. Wenn f in x differenzierbar ist, dann ergibt sich die Behauptung von 4. unmittelbar aus der Formel der 3.Bemerkung für $r = \pm e^i$, weil in diesem Fall $\bigtriangledown f(x) = [\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, ..., \frac{\partial f(x)}{\partial x_n}]$.

6. Wie steht es mit dem umgekehrten Zusammenhang zur 4.Bemerkung?

   (a) Wenn $f'(x; e^i)$ existiert, heißt das noch nicht, daß auch $f'(x; -e^i)$ existieren muß und schon gar nicht, daß $\frac{\partial f(x)}{\partial x_i}$ existieren muß.

   Erstes Beispiel: Sei $f : \mathbb{R}^2 \to \bar{\mathbb{R}}$, wobei

$$f(x) = \begin{cases} x_1^2 + x_2^2, -1 \leq x_1, x_2 \leq 1 \\ +\infty, \text{ sonst.} \end{cases}$$



   Sei $x^0 = [-1, -1]$.

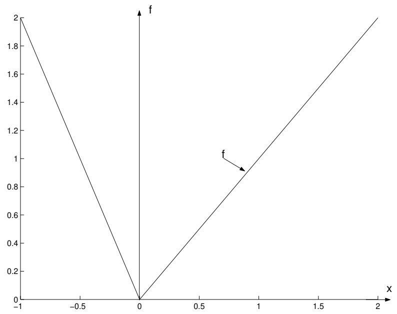
$$\begin{aligned} f'(x^0; e^1) &= \lim_{\alpha \to +0} \frac{f(x^0 + \alpha e^1) - f(x^0)}{\alpha} \\ &= \lim_{\alpha \to +0} \frac{(-1+\alpha)^2 + (-1)^2 - ((-1)^2 + (-1)^2)}{\alpha} \\ &= \lim_{\alpha \to +0} \frac{\alpha^2 - 2\alpha}{\alpha} = -2, \end{aligned}$$

d.h. $f'(x^0; e^1)$ existiert.

$$\begin{aligned} f'(x^0; -e^1) &= \lim_{\alpha \to +0} \frac{f(x^0 - \alpha e^1) - f(x^0)}{\alpha} \\ &= \lim_{\alpha \to +0} \frac{\infty - 2}{\alpha} \end{aligned}$$

existiert nicht.
Nach der 4.Bemerkung kann dann auch $\frac{\partial f(x)}{\partial x_1}$ nicht existieren.
Zweites Beispiel: Wir betrachten $f : \mathbb{R}^1 \to \mathbb{R}$, wobei

$$f(x) = \begin{cases} x, & x \geq 0 \\ -2x, & x < 0 \end{cases}$$



$$\begin{aligned} f'(0; e^1) &= \lim_{\alpha \to +0} \frac{f(0 + \alpha) - f(0)}{\alpha} = \lim_{\alpha \to +0} \frac{\alpha - 0}{\alpha} = 1 \\ f'(0; -e^1) &= \lim_{\alpha \to +0} \frac{f(0 - \alpha) - f(0)}{\alpha} = \lim_{\alpha \to +0} \frac{2\alpha - 0}{\alpha} = 2, \end{aligned}$$

d.h. $f'(0; e^1)$ und $f'(0; -e^1)$ existieren, aber $f'(0; e^1) \neq -f'(0; -e^1)$.
Deshalb kann nach 4.Bemerkung $\frac{\partial f(x)}{\partial x_1}$ nicht existieren.

(b) Wenn aber sowohl $f'(x; e^i)$ als auch $f'(x; -e^i)$ existieren und dabei gilt $f'(x; e^i) = -f'(x; -e^i)$, dann existiert auch $\frac{\partial f(x)}{\partial x_i}$ und es gilt $\frac{\partial f(x)}{\partial x_i} = f'(x; e^i)$.
Überprüfen Sie diese Behauptung, indem Sie die Definition von $\frac{\partial f(x)}{\partial x_i}$ ausnutzen.

7. Man soll nicht denken, daß aus der Existenz von $f'(x; r)$ $\forall r \in \mathbb{R}^n$ folgt, daß f in x differenzierbar ist (d.h. $\nabla f(x)$ existiert).

Beweis. Sei $f : \mathbb{R}^1 \to \mathbb{R}$ mit $f(x) = |x|$. Es existiert $f'(0; r)$ $\forall r$, aber $f'(0)$ existiert nicht.

$$\left(f'(0; r) = \begin{cases} r, & r > 0 \\ -r, & r < 0 \end{cases} \right).$$

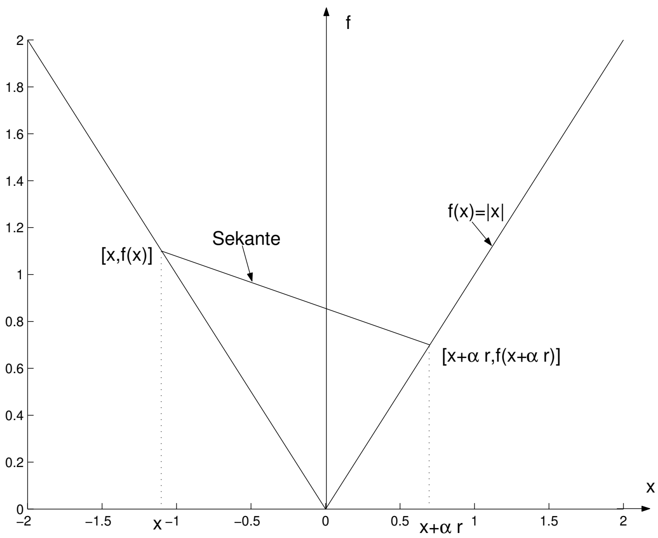
□

Zu den nützlichen Eigenschaften konvexer Funktionen gehört, daß sie richtungsableitbar sind. Zur Herleitung der exakten Aussage wollen wir nun übergehen.

Für $\bar{x} \in \mathrm{dom} f$ und $r \in \mathbb{R}^n$ setzen wir

$$\psi(\alpha) = \frac{f(\bar{x} + \alpha r) - f(\bar{x})}{\alpha}.$$

$\psi(\alpha)$ ist der Anstieg der Sekante durch die Punkte $[\bar{x}, f(\bar{x})]$ und $[\bar{x} + \alpha r, f(\bar{x} + \alpha r)]$, wenn $\|r\| = 1$.



Nach der Definition der Richtungsableitung haben wir: $f'(\bar{x}, r)$ existiert genau dann, wenn $\lim_{\alpha \to +0} \psi(\alpha)$ exisiert. Im Existenzfall ist

$$f'(\bar{x}, r) = \lim_{\alpha \to +0} \psi(\alpha),$$

unabhängig davon ob $\|r\| = 1$ gilt oder nicht.

Satz 2.1. *Sei $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ eine eigentliche konvexe Funktion. Sei $\bar{x} \in$ ri dom$f$, $r \in$ Lin dom$f$. Dann gilt*

1. *$\psi(\alpha)$ ist auf*
$$A = \{\alpha : \alpha > 0, (\bar{x} + \alpha r) \in dom f\}$$
monoton nichtfallend und nach unten beschränkt.

2. *$f'(\bar{x}, r)$ existiert und es gilt*
$$f'(\bar{x}, r) \le \psi(\alpha), \forall \alpha \in A,$$
$$f'(\bar{x}, r) = \inf_{\alpha > 0} \psi(\alpha).$$

Beweis. 1. Wir wollen zuerst einige Begriffe aus der Grundvorlesung "Optimierung I" wiederholen (gleich in Anwendung auf die hier verwendeten Mengen).

$$\text{aff dom}f = \{\lambda x + (1-\lambda)y : x, y \in \text{dom}f, \lambda \in (-\infty, +\infty)\}.$$

Nach Hilfssatz 2.4 der Grundvorlesung existiert eine Matrix $\mathfrak{A}$ vom Typ $[m, n]$ und ein Vektor $b \in \mathbb{R}^m$, so daß

$$\text{aff dom}f = \{x \in \mathbb{R}^n : \mathfrak{A}x = b\}.$$

Per Definition ist dann Lin dom$f = \{x \in \mathbb{R}^n : \mathfrak{A}x = 0\}$.

$$\text{ri dom}f = \{x \in \text{dom}f : \exists \varepsilon > 0, \text{ so daß aff dom}f \cap U_\varepsilon(x) \subset \text{dom}f\}.$$

2. Da $\bar{x} \in$ ri dom$f$ und $r \in$ Lin dom$f$, so ist $(\bar{x} + \alpha r) \in$ dom$f$, wenn nur $\alpha > 0$, α - genügend klein ist. Folglich ist $A \neq \emptyset$.

   Seien nunmehr $\alpha, \alpha' \in A$, wobei $0 < \alpha \leq \alpha'$. Dann ist $0 < \frac{\alpha}{\alpha'} \leq 1$ und aus der Konvexität von f folgt

$$\begin{aligned} f(\bar{x} + \alpha r) &= f((1 - \frac{\alpha}{\alpha'})\bar{x} + \frac{\alpha}{\alpha'}(\bar{x} + \alpha' r)) \\ &\leq (1 - \frac{\alpha}{\alpha'})f(\bar{x}) + \frac{\alpha}{\alpha'}f(\bar{x} + \alpha' r). \end{aligned}$$

   Daraus ergibt sich

$$\frac{f(\bar{x} + \alpha r)}{\alpha} \leq (\frac{1}{\alpha} - \frac{1}{\alpha'})f(\bar{x}) + \frac{1}{\alpha'}f(\bar{x} + \alpha' r),$$

   d.h.

$$\frac{f(\bar{x} + \alpha r) - f(\bar{x})}{\alpha} \leq \frac{f(\bar{x} + \alpha' r) - f(\bar{x})}{\alpha'},$$

   was

$$\psi(\alpha) \leq \psi(\alpha') \qquad (2.1.2)$$

   bedeutet.

   Somit ist ψ monoton nichtfallend auf A. Damit existiert $\lim_{\alpha \to +0} \psi(\alpha)$, d.h. $f'(\bar{x}; r)$ existiert, wenn nur $\psi(\alpha) \geq const$, $\forall \alpha > 0$.

   Nach (2.1.2) ist $\lim_{\alpha \to +0} \psi(\alpha) = \inf_{\alpha > 0} \psi(\alpha)$.

3. Es bleibt zu zeigen, daß $\inf_{\alpha > 0} \psi(\alpha) > -\infty$.

   Sei $\alpha \in A$, α - fixiert.

   Wir setzen $B := \{\beta : \beta < 0, \bar{x} + \beta r \in \text{dom}f\}$. Da mit $r \in$ Lin dom$f$ auch $(-r) \in$ Lin dom$f$, so ist (in Analogie zum Beginn von 1.) $B \neq \emptyset$.

   Wir wählen $\beta \in B$, β - fixiert.

Wir haben $0 < \frac{\alpha}{\alpha-\beta} \leq 1$, $(\bar{x}+\alpha r) \in \text{dom} f$, $(\bar{x}+\beta r) \in \text{dom} f$ und aufgrund der Konvexität von f ist

$$\begin{aligned} f(\bar{x}) &= f\left(\frac{\alpha}{\alpha-\beta}(\bar{x}+\beta r) - \frac{\beta}{\alpha-\beta}(\bar{x}+\alpha r)\right) \\ &\leq \frac{\alpha}{\alpha-\beta} f(\bar{x}+\beta r) - \frac{\beta}{\alpha-\beta} f(\bar{x}+\alpha r). \end{aligned}$$

Daraus erhalten wir nach Multiplikation mit $\frac{\alpha-\beta}{\alpha\beta} < 0$

$$\left(\frac{1}{\beta} - \frac{1}{\alpha}\right) f(\bar{x}) \geq \frac{f(\bar{x}+\beta r)}{\beta} - \frac{f(\bar{x}+\alpha r)}{\alpha},$$

d.h.

$$\psi(\beta) \leq \psi(\alpha). \tag{2.1.3}$$

Da $\alpha \in A$ beliebig, so gilt

$$\psi(\beta) \leq \inf_{\alpha>0} \psi(\alpha),$$

was zu beweisen war.

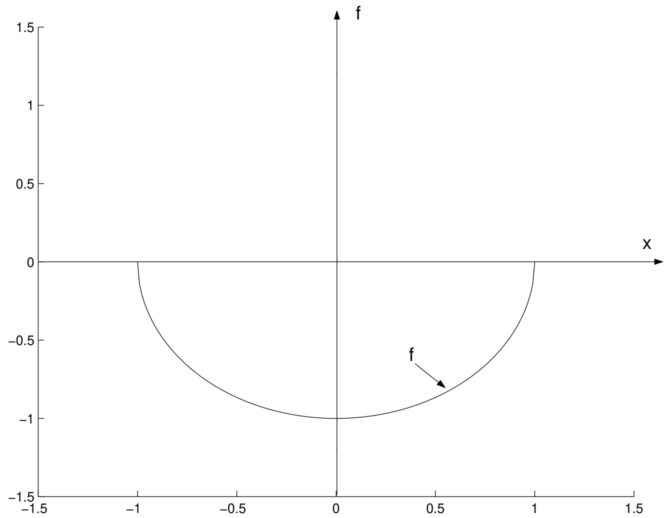
□

Bemerkung. 1. Im allgemeinen können wir die Forderung $\bar{x} \in \text{ri dom} f$ nicht durch $\bar{x} \in \text{dom} f$ ersetzen.

Beweis. Sei

$$f(x) = \begin{cases} -\sqrt{1-x^2}, -1 \leq x \leq 1 \\ +\infty, \text{ sonst.} \end{cases}$$

Hier ist $\text{dom} f = [-1, 1]$, $\text{ri dom} f = (-1, 1)$, $\text{Lin dom} f = \mathbb{R}$.



Sei $\bar{x} = +1$, $r = -1$. Dann ist $\bar{x} \in \mathrm{dom} f$, aber $\bar{x} \notin \mathrm{ri}\ \mathrm{dom} f$.

Wir haben

$$\begin{aligned}\psi(\alpha) &= \frac{f(\bar{x} + \alpha r) - f(\bar{x})}{\alpha} \\ &= \frac{-\sqrt{1-(1-\alpha)^2} - \sqrt{1-1^2}}{\alpha} \\ &= \frac{-\sqrt{2-\alpha}}{\sqrt{\alpha}} \to -\infty\end{aligned}$$

bei $\alpha \to +0$. Folglich ist $\inf_{\alpha>0}\{\psi(\alpha)\} = -\infty$, d.h. $f(\bar{x}; r)$ existiert nicht. $\square$

2. Wenn wir in obigen Beispiel $\bar{x} \notin \mathrm{dom} f$ wählen, dann ist $(\bar{x} + \alpha r) \notin \mathrm{dom} f$ für $|\alpha|$ genügend klein und $\psi(\alpha)$ ist nicht definiert, d.h. $f'(\bar{x}; r)$ ist nicht existent.

Folgerung 2.2. *Sei $f : \mathbb{R}^n \to \mathbb{R}$, f im $\mathbb{R}^n$ konvex. Dann besitzt f in jedem Punkt $\bar{x} \in \mathbb{R}^n$ in jeder Richtung $r \in \mathbb{R}^n$ die Richtungsableitung $f'(\bar{x}; r)$.*

Beweis. $\mathrm{dom} f = \mathbb{R}^n$ und Satz 2.1 ergeben die Behauptung. $\square$

2.2 Das Subdifferential einer konvexen Funktion

2.2.1 Begriff des Subgradienten

Aus dem Abschnitt über differentielle Eigenschaften konvexer Funktionen der Grundvorlesung "Optimierung I" wissen wir, daß bei einer konvexen, im Punkt $\bar{x}$ differenzierbaren Funktion f die Tangentialhyperebene eine Stützhyperebene ist. Genauer, nach Satz 2.28 der Grundvorlesung gilt

$$f(y) - f(\bar{x}) \geq < \nabla f(\bar{x}), y - \bar{x} >, \forall y \in \mathrm{dom} f. \tag{2.2.1}$$

Diese Eigenschaft des Gradienten konvexer Funktionen wird häufig in Beweisen benutzt. Wenn wir die Formel (2.2.1) zur Verallgemeinerung des Begriffes "Gradient" benutzen, dann lassen sich die angesprochenen Beweisteile weiterhin zeigen, auch wenn f in $\bar{x}$ nicht differenzierbar ist.

Definition. Sei $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ eine Funktion. Sei $\bar{x} \in \mathrm{dom} f$. Der Vektor $a \in \mathbb{R}^n$ heißt Subgradient von f im Punkt $\bar{x}$, wenn

$$f(y) - f(\bar{x}) \geq < a, y - \bar{x} >, \forall y \in \mathbb{R}^n. \tag{2.2.2}$$

Die Menge aller Subgradienten von f in $\bar{x}$ heißt Subdifferential von f in $\bar{x}$ und wird mit $\partial f(\bar{x})$ bezeichnet.

Bemerkung. 1. Wenn f in $\bar{x}$ differenzierbar und bezüglich $\mathbb{R}^n$ konvex ist, so gilt also nach Satz 2.28 der Grundvorlesung "Optimierung I", daß $\nabla f(\bar{x}) \in \partial f(\bar{x})$.

2. Wenn $\bar{x} \notin \mathrm{dom} f$, dann ist (2.2.2) für $y \in \mathrm{dom} f$ bei keinem a erfüllbar. Per Definition setzen wir für solche $\bar{x}$ $\partial f(\bar{x}) = \emptyset$.

3. Rechenbeispiel: Sei $f(x) = |x|, \forall x \in \mathbb{R}$. Wir zeigen, daß

$$\partial f(x) = \begin{cases} \{1\}, x > 0 \\ [-1,1] = \mathrm{conv}\{-1,1\}, x = 0 \\ \{-1\}, x < 0 \end{cases}$$

Beweis: (durch Anwenden der Definition).

(a) Fall: $\bar{x} = 0$.

$$a \in \partial f(\bar{x}) \Leftrightarrow f(x) - f(\bar{x}) \geq < a, x - \bar{x} >, \forall x$$

$$\Leftrightarrow \begin{cases} -x - 0 \geq a(x-0), \forall x < 0 \\ x - 0 \geq a(x-0), \forall x > 0 \end{cases}$$

$$\Leftrightarrow -1 = \frac{-x}{x} \leq a \leq \frac{x}{x} = 1, \forall x \neq 0 \Leftrightarrow a \in [-1,1]$$

(b) Fall: $\bar{x} > 0$.

$$a \in \partial f(\bar{x}) \Leftrightarrow \begin{cases} -x - \bar{x} \geq a(x - \bar{x}), \forall x \leq 0 \\ x - \bar{x} \geq a(x - \bar{x}), \forall x > 0 \end{cases}$$

$$\Leftrightarrow \begin{cases} -x - \bar{x} \geq a(x - \bar{x}), \forall x \text{ mit } x < -\bar{x} \\ -x - \bar{x} \geq a(x - \bar{x}), \forall x \text{ mit } 0 \geq x \geq -\bar{x} \\ x - \bar{x} \geq a(x - \bar{x}), \forall x \text{ mit } \bar{x} \geq x \geq 0 \\ x - \bar{x} \geq a(x - \bar{x}), \forall x \text{ mit } x > \bar{x} \end{cases}$$

$$\Leftrightarrow \begin{cases} 1 \geq a\frac{x-\bar{x}}{-x-\bar{x}}, \forall x \text{ mit } x < -\bar{x} < 0 \\ 1 \leq a\frac{x-\bar{x}}{-x-\bar{x}}, \forall x \text{ mit } 0 > x > -\bar{x} \\ 1 \leq a\frac{x-\bar{x}}{x-\bar{x}} = a, \forall x \text{ mit } \bar{x} \geq x \geq 0 \\ 1 \geq a\frac{x-\bar{x}}{x-\bar{x}} = a, \forall x \text{ mit } x > \bar{x} \end{cases}$$

$\Leftrightarrow a = 1$, weil die letzen zwei Ungleichungen zusammen nur bei $a = 1$ möglich sind und die ersten zwei Ungleichungen bei $a = 1$ wahr sind, da dort $\frac{x-\bar{x}}{-x-\bar{x}} < 0$ bzw. $\frac{x-\bar{x}}{-x-\bar{x}} > 1$.

(c) Fall: $\bar{x} < 0$.
Analog wie oben zeigt man, daß $a = -1$.

□

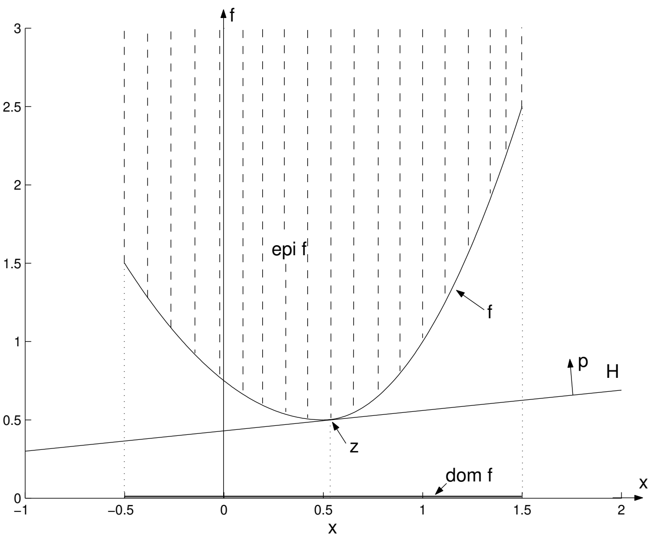
4. Geometrische Illustration des Subgradienten.

Sei $\bar{x} \in \mathrm{dom} f$, $f : \mathbb{R}^n \to \bar{\mathbb{R}}$.

Wir setzen

$$\mathrm{epi} f := \{[x, \alpha] \in \mathbb{R}^n \times \mathbb{R} : f(x) \leq \alpha\}.$$

Wenn f eine konvexe Funktion ist, dann ist epi$f$ eine konvexe Menge.



Wir betrachten die Hyperebene

$$H = \{z \in \mathbb{R}^{n+1} :< p, z >= \beta\}$$

des $\mathbb{R}^{n+1}$ mit $\beta = (f(\bar{x}) - < a, \bar{x} >) \in \mathbb{R}^1$ und

$$p = \begin{bmatrix} -a \\ +1 \end{bmatrix} \in \mathbb{R}^{n+1}.$$

Vermerke: da $\|p\| \neq 0$, so ist H wirklich eine Hyperebene.
Es gilt

$$a \in \partial f(\bar{x}) \Leftrightarrow H \text{ ist Stützhyperebene an epi} f \text{ in } \hat{z} = \begin{bmatrix} \bar{x} \\ f(\bar{x}) \end{bmatrix}.$$

Beweis. Nach der Definition einer Stützhyperebene (vgl. Abschnitt 2.2.3 der Grundvorlesung "Optimierung I") ist H Stützhyperebene an epi$f$ in $\hat{z}$

$$\Leftrightarrow \begin{cases} < p, z > \geq \beta, \forall z \in \text{epi} f \\ < p, \hat{z} >= \beta \end{cases}$$

Mit $z = \begin{bmatrix} x \\ \alpha \end{bmatrix} \in \text{epi} f \Leftrightarrow f(x) \leq \alpha$ erhalten wir, daß H Stützhyperebene ist.

$$\Leftrightarrow \begin{cases} < -a, x > +\alpha \geq \beta, \forall \begin{bmatrix} x \\ \alpha \end{bmatrix} \text{ mit } f(x) \leq \alpha \\ < -a, \bar{x} > +f(\bar{x}) = \beta \end{cases}$$

$$\Leftrightarrow < -a, x > +\alpha \geq f(\bar{x}) - < a, \bar{x} >, \forall \begin{bmatrix} x \\ \alpha \end{bmatrix} \text{ mit } f(x) \leq \alpha$$

$$\Leftrightarrow < -a, x > + f(x) \geq f(\bar{x}) - < a, \bar{x} >, \forall x \in \mathbb{R}^n$$
$$\Leftrightarrow f(x) - f(\bar{x}) \geq < a, x - \bar{x} >, \forall x \in \mathbb{R}^n$$

$\Leftrightarrow a$ ist Subgradient von f in $\bar{x}$. □

5. Wenn a Subgradient ist, dann ist $p = \begin{bmatrix} -a \\ 1 \end{bmatrix}$ Stellungsvektor der zugehörigen Stützhyperebene H. Folglich haben wir nach 4.
   (a) H ist nicht parallel zur $n+1$-ten Achse, d.h. H ist nichtvertikal (da die $n+1$-te Koordinate von p nicht Null ist).
   Folglich ist die Existenz eines Subgradienten in $\bar{x}$ äquivalent mit der Existenz einer nichtvertikalen Stützhyperebene.
   (b) Ist $p = \begin{bmatrix} b \\ \alpha \end{bmatrix}$ mit $b \in \mathbb{R}^n$, $\alpha \in \mathbb{R}$ und $\alpha > 0$ Stellungsvektor einer Stützhyperebene in $\bar{x}$ an epi$f$, so ist $(-\frac{1}{\alpha}b) \in \partial f(\bar{x})$.
6. Weiter unten (vgl. Satz 2.5) werden wir sehen, daß f Subgradienten in allen $x \in \mathrm{dom} f$ nur haben kann, wenn f konvex ist.
7. In der konvexen Analysis wird einer Funktion f über $f^*(a) = \sup\{< a, x > - f(x) : x \in \mathbb{R}^n\}$ die sogenannte konjugierte Funktion f^* zugeordnet. Es gilt: a ist Subgradient von f in $\bar{x}$ genau dann, wenn $f^*(a) = < a, \bar{x} > - f(\bar{x})$.

   Beweis. a ist Subgradient von f in $\bar{x}$

   $$\Leftrightarrow f(x) - < a, x > \geq f(\bar{x}) - < a, \bar{x} >, \forall x$$

   $\Leftrightarrow$ die Funktion $F(x) = < a, x > - f(x)$ hat in $\bar{x}$ ein globales Maximum, d.h.

   $$F(\bar{x}) = \max\{F(x) : x \in \mathbb{R}^n\} \Leftrightarrow$$
   $$< a, \bar{x} > - f(\bar{x}) = \max\{< a, x > - f(x) : x \in \mathbb{R}^n\} = f^*(a).$$

   □

2.2.2 Eigenschaften des Subdifferentials

Hilfssatz 2.3 (Satz über die dazwischen liegende Hyperebene). *Seien C_1, C_2 konvexe Teilmengen des $\mathbb{R}^n$. Seien $f_1 : C_1 \to \mathbb{R}^1$, $f_2 : C_2 \to \mathbb{R}^1$, seien f_1 konvex auf C_1, f_2 konkav auf C_2. Weiterhin sei $riC_1 \cap riC_2 \neq \emptyset$.*

Wenn unter diesen Voraussetzungen

$$f_1(x) \geq f_2(x), \forall x \in C_1 \cap C_2,$$

dann existiert eine affin-lineare Funktion $l(x) = < a, x > + \alpha$ (d.h. eine Hyperebene des $\mathbb{R}^{n+1}$ mit der Gleichung $z = l(x)$), so daß

$$f_1(x) \geq l(x), \forall x \in C_1,$$
$$f_2(x) \leq l(x), \forall x \in C_2.$$

Beweisgedanke. Setze

$$E_1 = \{[x, \beta] \in \mathbb{R}^n \times \mathbb{R}^1 : x \in C_1, f_1(x) < \beta\},$$
$$E_2 = \{[x, \beta] \in \mathbb{R}^n \times \mathbb{R}^1 : x \in C_2, f_2(x) > \beta\}.$$

Zeige, daß E_1 und E_2 beide konvexe, nichtleere Mengen sind und daß $E_1 \cap E_2 = \emptyset$.

Damit ist $\mathrm{ri}E_1 \cap \mathrm{ri}E_2 = \emptyset$ und nach Theorem 2.19 der Grundvorlesung "Optimierung I" sind E_1 und E_2 eigentlich trennbar.

Folglich existiert ein $\gamma \in \mathbb{R}$ und existiert ein $[p, \lambda] \in \mathbb{R}^n \times \mathbb{R}$, $[p, \lambda] \neq 0$, so daß

$$\begin{cases} < [p,\lambda], \begin{bmatrix} x \\ \beta \end{bmatrix} > \geq \gamma, \forall \begin{bmatrix} x \\ \beta \end{bmatrix} \in E_1 \\ < [p,\lambda], \begin{bmatrix} x \\ \beta \end{bmatrix} > \leq \gamma, \forall \begin{bmatrix} x \\ \beta \end{bmatrix} \in E_2 \\ \text{existieren } \begin{bmatrix} \bar{x} \\ \bar{\beta} \end{bmatrix} \in E_1 \text{ und } \begin{bmatrix} \bar{\bar{x}} \\ \bar{\bar{\beta}} \end{bmatrix} \in E_2, \\ \text{so daß } < [p,\lambda], \begin{bmatrix} \bar{x} \\ \bar{\beta} \end{bmatrix} >>< [p,\lambda], \begin{bmatrix} \bar{\bar{x}} \\ \bar{\bar{\beta}} \end{bmatrix} > . \end{cases}$$

Daraus können wir schließen, daß $\lambda \geq 0$, und mit $l(x) :=< -\frac{1}{\lambda}p, x > +\frac{\gamma}{\lambda}$ zeigen, daß

$$f_1(x) \geq l(x), \forall x \in C_1,$$
$$f_2(x) \leq l(x), \forall x \in C_2.$$

Anmerkung. Die Annahme daß $\lambda = 0$ führt unter Benutzung von Theorem 2.19 zum Widerspruch mit $\mathrm{ri}(C_1) \cap \mathrm{ri}(C_2) \neq \emptyset$.

□

Satz 2.4. *Sei $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ eine eigentliche konvexe Funktion. Dann gilt:*

1. *$\forall x \in ri\ domf$ ist $\partial f(x) \neq \emptyset$.*
2. *$\forall x \in \mathbb{R}^n$ ist $\partial f(x)$ eine konvexe, abgeschlossene Menge.*

Beweis. Sei $\bar{x} \in \mathrm{ri\ dom}f$.

1. Wir setzen $C_1 := \mathrm{dom}f$, $C_2 := \{\bar{x}\}$. $f_1(x) := f(x)$, $f_2(x) := f(\bar{x}) = const.$

   Offensichtlich sind für C_1, C_2, f_1, f_2 die Voraussetzungen von Hilfssatz 2.3 erfüllt. Deshalb existieren ein $a \in \mathbb{R}^n$ und ein $\alpha \in \mathbb{R}$, so daß für $l(x) =< a, x > +\alpha$ gilt

   $$\begin{cases} f_1(x) \geq l(x), \forall x \in C_1 \\ f_2(x) \leq l(x), \forall x \in C_2. \end{cases}$$

   Damit haben wir

   $$\begin{cases} f(x) \geq< a, x > +\alpha, \forall x \in \mathrm{dom}f \\ f(\bar{x}) \leq< a, \bar{x} > +\alpha, \end{cases}$$

d.h. es gilt

$$f(x) - f(\bar{x}) \geq < a, x - \bar{x} >, \forall x \in \mathrm{dom} f$$

und somit auch

$$f(x) - f(\bar{x}) \geq < a, x - \bar{x} >, \forall x \in \mathbb{R}^n.$$

Folglich ist $a \in \partial f(\bar{x})$, d.h. $\partial f(\bar{x}) \neq \emptyset$.

2. Laut Definition von ∂f ist

$$\partial f(\bar{x}) = \{a \in \mathbb{R}^n : f(x) - f(\bar{x}) \geq < a, x - \bar{x} >, \forall x \in \mathbb{R}^n\}.$$

Damit ist

$$\partial f(\bar{x}) = \cap_{x \in \mathbb{R}^n} \{a \in \mathbb{R}^n : f(x) - f(\bar{x}) \geq < a, x - \bar{x} >\}.$$

Hier ist $\{a : ...\}$ die Lösungsmenge (bzgl. a) einer linearen Ungleichung in a (x ist fixiert), d.h. $\{a : ...\}$ ist eine konvexe, abgeschlossene Menge. Diese Eigenschaften gehen nicht verloren, wenn wir den Durchschnitt aller solcher Mengen bilden.

$\Rightarrow \partial f(\bar{x})$ ist konvex und abgeschlossen.

□

Bemerkung. 1. Bei $x \in \mathrm{dom} f \setminus \mathrm{ri}\ \mathrm{dom} f$ ist es möglich, daß $\partial f(\bar{x}) = \emptyset$. Beispiel:

$$f(x) = \begin{cases} -\sqrt{1-x^2}, -1 \leq x \leq 1 \\ +\infty, \text{ sonst} \end{cases}$$

Hier ist $\partial f(\pm 1) = \emptyset$, weil in diesen Punkten keine nichtvertikale Stützhyperebene an epi$f$ existiert (vgl. 5.Bemerkung nach der Definition der Subgradienten).

2. Wenn $f : \mathbb{R}^n \to \mathbb{R}$, f - konvex im $\mathbb{R}^n$, dann ist somit $\partial f(x) \neq \emptyset$, $\forall x \in \mathbb{R}^n$. Wie aus späteren Überlegungen (vgl. Folgerung 2.12) folgt, ist in diesem Fall $\forall x \in \mathbb{R}^n$ $\partial f(x)$ eine beschränkte Menge.

Satz 2.5. *Sei $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ eine eigentliche Funktion und sei dom$f$ eine konvexe Menge. Dann gilt: Wenn $\forall x \in$ dom$f$ gilt $\partial f(x) \neq \emptyset$, dann ist f im $\mathbb{R}^n$ konvex.*

Beweis. Seien $x, y \in \mathbb{R}^n$, $\lambda \in (0, 1)$. Wir setzen $z := \lambda x + (1-\lambda)y$.

1. Wenn wenigstens eine der Beziehungen $x \notin \mathrm{dom} f$ bzw. $y \notin \mathrm{dom} f$ wahr ist, dann gilt

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y).$$

2. Sei deshalb $x, y \in \mathrm{dom} f$. Da $\mathrm{dom} f$ konvex vorausgesetzt wurde, so ist $z \in \mathrm{dom} f$ und damit $\partial f(z) \neq \emptyset$. Folglich existiert ein $a \in \mathbb{R}^n$, so daß

$$f(w) - f(z) \geq < a, w - z >, \forall w \in \mathbb{R}^n.$$

Somit gilt

$$\begin{aligned} f(x) - f(z) &\geq < a, x - z >, \\ f(y) - f(z) &\geq < a, y - z > . \end{aligned}$$

Nach Multiplikation mit λ bzw. $(1 - \lambda)$ und anschließender Addition der beiden Ungleichungen erhalten wir

$$\lambda f(x) + (1 - \lambda) f(y) - f(z) \geq < a, \lambda x + (1 - \lambda) y - z >,$$

d.h.

$$\lambda f(x) + (1 - \lambda) f(y) \geq f(\lambda x + (1 - \lambda) y),$$

wenn wir wieder $z = \lambda x + (1 - \lambda) y$ rücksubstituieren.

1., 2. liefern die Konvexität von f. $\square$

Bemerkung. 1. Ohne die Konvexität von $\mathrm{dom} f$ muß die Behauptung des Satzes nicht wahr sein.

Beispiel:

$$f(x) = \begin{cases} c, & x \in A \cup B \\ +\infty, & \text{sonst} \end{cases},$$

wobei $A = [1, 2]$, $B = [3, 4]$, $c = const.$

Wir haben

$$\begin{cases} \mathrm{dom} f = A \cup B, \mathrm{dom} f \text{ ist nicht konvex, } f \text{ - eigentlich,} \\ \text{da } 0 \in \partial f(x), \forall x \in \mathrm{dom} f, \text{ so ist } \partial f(x) \neq \emptyset \text{ bei } x \in \mathrm{dom} f. \end{cases}$$

Aber f ist nicht konvex im $\mathbb{R}^1$, weil z.B. für $x = 2$, $y = 3$, $\lambda = \frac{1}{2}$ die Forderung

$$f(\lambda x + (1 - \lambda) y) \leq \lambda f(x) + (1 - \lambda) f(y)$$

verletzt ist.

Satz 2.6. *Es sei $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ eine konvexe, eigentliche Funktion. Dann gilt*

1. $\partial f(\bar{x}) = \{a \in \mathbb{R}^n : f'(\bar{x}; r) \geq < a, r >, \forall r \in \mathit{Lin\ dom} f\}, \forall \bar{x} \in \mathit{ri\ dom} f.$

2. $\forall \bar{x} \in \mathit{ri\ dom} f$, $\forall r \in \mathit{Lin\ dom} f$ *ist*

$$f'(\bar{x}; r) = \max\{< a, r >: a \in \partial f(\bar{x})\}.$$

Anmerkung. Wenn wir in der Definition von $f'(x; r)$ auch den Wert $+\infty$ zulassen, dann gilt Behauptung 1. auch für $\bar{x} \in \mathrm{dom} f$.

Beweis. Sei $\bar{x} \in \text{ri dom} f$. Wir setzen

$$D := \{a \in \mathbb{R}^n : f'(\bar{x}; r) \geq < a, r >, \forall r \in \text{Lin dom} f\}.$$

1. Wir zeigen zuerst, daß $\partial f(\bar{x}) \subset D$. Sei $a \in \partial f(\bar{x})$, $r \in \text{Lin dom} f$.
   Nach Satz 2.1 existiert $f'(\bar{x}; r)$ und es ist

   $$f'(\bar{x}; r) = \lim_{\alpha \to +0} \frac{f(\bar{x} + \alpha r) - f(\bar{x})}{\alpha}.$$

   Nun ist aber nach Definition des Subgradienten

   $$f(\bar{x} + \alpha r) - f(\bar{x}) \geq < a, \bar{x} + \alpha r - \bar{x} >, \forall \alpha$$

   $\Rightarrow$

   $$f'(\bar{x}; r) \geq \lim_{\alpha \to +0} \frac{\alpha < a, r >}{\alpha} = < a, r >,$$

   d.h. $a \in D$.

2. Wir zeigen, daß auch $D \subset \partial f(\bar{x})$.
   Sei jetzt $a \in D$. $\forall r \in \text{Lin dom} f$ ist dann

   $$f'(\bar{x}; r) \geq < a, r > . \tag{2.2.3}$$

   Da $\bar{x} \in \text{ri dom} f$, so ist $(\bar{x} + \alpha r) \in \text{dom} f$, wenn nur $\alpha > 0$, α - genügend klein. Nach Satz 2.1 ist deshalb $\psi(\alpha) \geq f'(\bar{x}; r)$.

   Zusammen mit (2.2.3) erhalten wir also

   $$\psi(\alpha) := \frac{f(\bar{x} + \alpha r) - f(\bar{x})}{\alpha} \geq < a, r >,$$

   d.h. $f(\bar{x} + \alpha r) - f(\bar{x}) \geq < a, \alpha r > = < a, \bar{x} + \alpha r - \bar{x} >$.

   Bewiesen haben wir diese Ungleichung unter der Annahme $(\bar{x} + \alpha r) \in \text{dom} f$. Da aber bei $(\bar{x} + \alpha r) \notin \text{dom} f$ gilt, daß $f(\bar{x} + \alpha r) = +\infty$, so gilt diese Ungleichung auch für $(\bar{x} + \alpha r) \notin \text{dom} f$.

   Somit haben wir für $\alpha = 1$, $r \in \text{Lin dom} f$ gezeigt, daß stets

   $$f(\bar{x} + r) - f(\bar{x}) \geq < a, r > . \tag{2.2.4}$$

   Das genügt noch nicht, um zu behaupten, daß $a \in \partial f(\bar{x})$.

   Sei deshalb $y \in \mathbb{R}^n$. Wir setzen $r = y - \bar{x}$.

   Wenn $y \in \text{dom} f$, dann ist $r \in \text{Lin dom} f$ (weil $\bar{x} \in \text{ri dom} f$, d.h. $\bar{x} \in \text{dom} f$ und damit $(y - \bar{x}) \in \text{Lin dom} f$). Nach (2.2.4) ist damit

   $$f(y) - f(\bar{x}) \geq < a, y - \bar{x} >, \forall y \in \text{dom} f.$$

   Wenn $y \notin \text{dom} f$, dann ist $f(y) = +\infty$ und die letzte Ungleichung gilt auch in diesem Fall. Folglich ist $a \in \partial f(\bar{x})$, d.h. $D \subset \partial f(\bar{x})$. Damit ist (1) gezeigt.

3. Bei $r = 0$ ist Behauptung (2) wahr. Sei deshalb im weiteren $r \in \text{Lin dom} f$, $r \neq 0$.

   Nach (1) ist
   $$f'(\bar{x}; r) \geq \sup\{< a, r >: a \in \partial f(\bar{x})\}.$$

   Um (2) zu beweisen genügt es, ein $\hat{a} \in \partial f(\bar{x})$ zu finden, bei dem $f'(\bar{x}; r) \leq < \hat{a}, r >$. Um Hilfssatz 2.3 über die dazwischen liegende Hyperebene zur Anwendung zu bringen setzen wir

   $$C_1 := \text{dom} f, \quad f_1(x) = f(x),$$
   $$C_2 := \{\bar{x} + \alpha r : \alpha \geq 0\}, \quad f_2(x) = f(\bar{x}) + \frac{\|x - \bar{x}\|}{\|r\|} f'(\bar{x}; r)$$

   (Vermerke, daß $f'(\bar{x}; r)$ unter unseren Voraussetzungen existiert).

   Da $\bar{x} \in \text{ri dom} f$, $r \in \text{Lin dom} f$, so ist $(\bar{x} + \alpha r) \in \text{dom} f$, wenn $\alpha > 0$, α - genügend klein. Nach dem Zugänglichkeitslemma (Lemma 2.11 der Grundvorlesung "Optimierung I") ist dann $(\bar{x} + \beta r) \in \text{ri dom} f = \text{ri} C_1$ für β mit $0 \leq \beta < \alpha$. Folglich ist $\text{ri} C_1 \cap \text{ri} C_2 \neq \emptyset$.

   Außerdem haben wir: f_1 ist auf C_1 konvex, f_2 hat auf C_2 die Gestalt

   $$f_2(\bar{x} + \alpha r) = f(\bar{x}) + \alpha f'(\bar{x}; r),$$

   d.h. ist affin-linear und damit ist f_2 auf C_2 konkav.

   Außerdem haben wir für $x \in C_1 \cap C_2$, d.h. für $x = (\bar{x} + \alpha r) \in \text{dom} f$ (weil nach Satz 2.1 $\psi(\alpha) \geq f'(\bar{x}; r)$).

   $$f_1(x) = f(x) = f(\bar{x} + \alpha r) \geq f(\bar{x}) + \alpha f'(\bar{x}; r) = f_2(x).$$

   Somit ist
   $$f_1(x) \geq f_2(x) \text{ bei } x \in C_1 \cap C_2.$$

   Nach Hilfssatz 2.3 existieren somit ein $\hat{a} \in \mathbb{R}^n$ und ein $\beta \in \mathbb{R}$, so daß für $l(x) =< \hat{a}, x > +\beta$ gilt

   $$\begin{cases} f_1(x) \geq l(x), \forall x \in C_1 \\ f_2(x) \leq l(x), \forall x \in C_2. \end{cases}$$

   Das bedeutet

   $$\begin{cases} f(x) \geq < \hat{a}, x > +\beta, \forall x \in \text{dom} f \\ f(\bar{x}) + \alpha f'(\bar{x}; r) \leq < \hat{a}, \bar{x} + \alpha r > +\beta, \forall \alpha \geq 0, \end{cases} \tag{2.2.5}$$

   woraus wir bei $\alpha = 0$ durch Substraktion voneinander

   $$f(x) - f(\bar{x}) \geq < \hat{a}, x - \bar{x} >, \forall x \in \text{dom} f$$

   erhalten. Diese Ungleichung ist aber auch für $x \notin \text{dom} f$ wahr. Also, $\hat{a} \in \partial f(\bar{x})$.

Die zweite Zeile von (2.2.5) liefert $\forall \alpha \geq 0$, daß

$$\alpha(< \hat{a}, r > -f'(\bar{x}; r)) \geq f(\bar{x}) - \beta - < \hat{a}, \bar{x} >,$$

woraus sich $< \hat{a}, r > -f'(\bar{x}; r) \geq 0$ ergibt, w.z.b.w.

□

Hilfssatz 2.7. *Sei $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ eine konvexe, eigentliche Funktion. Dann ist f auf ri dom$f$ stetig bezüglich dom$f$, d.h. für $\bar{x} \in$ ri dom$f$ und beliebiges $\varepsilon > 0$ existiert ein $\delta > 0$, so daß*

$$|f(x) - f(\bar{x})| \leq \varepsilon, \forall x \in U_\delta(\bar{x}) \cap dom f.$$

Beweis. Sei $\bar{x} \in$ ri dom$f$. Durch die Transformationen $x := x - \bar{x}$, $f(x) := f(x) - f(\bar{x})$ können wir stets erreichen, daß $\bar{x} = 0$, $f(\bar{x}) = 0$. O.E.d.A. sei deshalb für das weitere $\bar{x} = 0$, $f(\bar{x}) = 0$.

In diesem Fall ist Lin dom$f$ = aff dom$f$.

1. Es sei int dom$f \neq \emptyset$ und damit ri dom$f$ = int dom$f$.

   Da $\bar{x} = 0 \in$ int dom$f$, so existiert ein $\alpha_0 > 0$, so daß der Würfel

   $$Q := \{x \in \mathbb{R}^n : -\alpha_0 \leq x_j \leq \alpha_0, j = 1, 2, ..., n\}$$

   in dom$f$ enthalten ist.

   Es seien $x^1, x^2, ..., x^m$ mit $m = 2^n$ die Extremalpunkte(=Ecken) von Q, d.h. die Punkte x der Gestalt $x = [\pm\alpha_0, \pm\alpha_0, ..., \pm\alpha_0] \in \mathbb{R}^n$. Wir setzen $\hat{f} = \max_{1 \leq i \leq m} f(x^i)$.

   Sei $x \in Q$. Dann existieren λ_i, $i = 1, 2, ..., m$, mit $\lambda_i \geq 0, \forall i$ und $\sum_{i=1}^m \lambda_i = 1$, so daß $x = \sum_{i=1}^m \lambda_i x^i$.

   Nach der Jensenschen Ungleichung (Satz 2.23 der Grundvorlesung "Optimierung I") erhalten wir

   $$f(x) \leq \sum_{i=1}^m \lambda_i f(x^i) \leq \hat{f}, \forall x \in Q.$$

   Folglich ist f auf Q nach oben beschränkt.

   Es sei nun $\varepsilon > 0$, $\varepsilon < 1$ gewählt. Wir setzen $U_\varepsilon(\bar{x}) = U_\varepsilon(0) = \varepsilon Q$. Für $x \in U_\varepsilon(\bar{x})$ sind damit $\pm\frac{1}{\varepsilon}x \in Q$.

   Aufgrund der Konvexität von f ergibt sich

   $$f(x) = f(\varepsilon\frac{x}{\varepsilon} + (1-\varepsilon)0) \leq \varepsilon f(\frac{x}{\varepsilon}) + (1-\varepsilon)f(0) = \varepsilon f(\frac{x}{\varepsilon}) \leq \varepsilon\hat{f}. \quad (2.2.6)$$

   (Wir haben hier benutzt, daß $f(0) = f(\bar{x}) = 0$ und $\frac{x}{\varepsilon} \in Q$ gelten.)

Außerdem ist

$$\begin{aligned} 0 = f(0) = f(&\frac{1}{1+\varepsilon}x + \frac{\varepsilon}{1+\varepsilon}(-\frac{x}{\varepsilon})) \\ &\leq \frac{1}{1+\varepsilon}f(x) + \frac{\varepsilon}{1+\varepsilon}f(-\frac{x}{\varepsilon}) \\ &\leq \frac{1}{1+\varepsilon}f(x) + \frac{\varepsilon}{1+\varepsilon}\hat{f}. \end{aligned}$$

(Wir haben hier benutzt, daß $-\frac{x}{\varepsilon} \in Q$ gilt.)

Somit ist $0 \leq f(x) + \varepsilon\hat{f}$, woraus wir zusammen mit (2.2.6) erhalten

$$-\varepsilon\hat{f} \leq f(x) \leq \varepsilon\hat{f}, \forall x \in U_\varepsilon(\bar{x}),$$

d.h. $|f(x) - f(\bar{x})| \leq \varepsilon\hat{f}$, $\forall x \in U_\varepsilon(\bar{x})$, woraus wir unmittelbar schließen können, daß f in $\bar{x}$ bzgl. $\mathbb{R}^n$ stetig ist.

2. Weiterhin sei vereinbart $\bar{x} = 0$, $f(\bar{x}) = 0$, aber jetzt sei int dom$f = \emptyset$.

   Es sei $\{x^1, x^2, ..., x^m\}$ eine Basis von Lin dom$f$ = aff dom$f$.

   Dann ist Lin dom$f = \{\sum_{i=1}^m \lambda_i x^i : \lambda_i \in \mathbb{R}, i = 1, 2, ..., m\}$. Damit ist über $x = F(\lambda) := \sum_{i=1}^m \lambda_i x^i$ eine lineare Abbildung $F : \mathbb{R}^m \to$ Lin dom$f$ definiert. Aufgrund der linearen Unabhängigkeit der $\{x^i\}_{i=1}^m$ ist F eine eineindeutige Abbildung. Aufgrund der Linearität von F muß die Abbildung F^{-1} stetig bzgl. $\mathbb{R}^m$ sein.

   Wir setzen $\Lambda := F^{-1}(\text{dom}f) \subset \mathbb{R}^m$. Da auch F^{-1} linear ist, ist Λ als Bild der konvexen Menge dom$f$ selbst auch konvex. Schließlich, weil $\bar{x} = 0 \in$ ri dom$f$, so ist $F^{-1}(0) \in \text{int}\Lambda$ (interior bzgl. $\mathbb{R}^m$), d.h. $0 = F^{-1}(0) \in \text{int}\Lambda$ (F^{-1} ist linear).

   Wir setzen $\varphi(\lambda) := f(F(\lambda))$.

   Als Superposition einer konvexen Funktion f und einer linearen Funktion F ist φ auf Λ konvex (vgl. Satz 2.27 der Grundvorlesung "Optimierung I").

   φ ist also eine konvexe Funktion des $\mathbb{R}^m$ mit dom$\varphi = \Lambda$. Weil $0 \in$ int domφ und $\varphi(0) = f(F(0)) = f(0) = 0$, so ist nach Teil (1) φ in 0 stetig (stetig bzgl. des $\mathbb{R}^m$).

   Da nun aber $f(x) = \varphi(F^{-1}(x))$ ($x = F(\lambda)$), so ist f als Superposition der stetigen linearen Funktion F^{-1} mit der in 0 stetigen Funktion φ selbst stetig in $x = \bar{x} = 0$ (bzgl. dom$f$, da F^{-1} nur auf Lin dom$f$ erklärt ist).

□

Folgerung 2.8. *1. Sei $f : \mathbb{R}^n \to \bar{\mathbb{R}}$, f - konvex und eigentlich. Sei $\bar{x} \in$ int dom$f$. Dann ist f in $\bar{x}$ stetig.*

2. Damit ist eine konvexe, eigentliche Funktion $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ auf int dom$f$ lokal Lipschitzstetig.

3. *Sei $f : \mathbb{R}^n \to \mathbb{R}$, f konvex (d.h. in jedem Punkt des $\mathbb{R}^n$ ist $f(x)$ endlich). Dann ist f im $\mathbb{R}^n$ stetig.*

Bemerkung. 1. Wie das Beispiel

$$f(x) = \begin{cases} x, & x > 0 \\ 1, & x = 0 \\ +\infty, & x < 0 \end{cases}$$

zeigt, muß die konvexe Funktion $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ in $x \in \text{dom} f$ nicht stetig sein.

Satz 2.9. *Sei $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ eine konvexe, eigentliche Funktion. Sei $\bar{x} \in$ int dom$f$. Dann gilt: $\partial f(\bar{x})$ ist ein singleton $\Leftrightarrow$ f ist in $\bar{x}$ differenzierbar.*

Beweis. $\Leftarrow$ Sei f in $\bar{x}$ differenzierbar. Da $\bar{x} \in$ int dom$f$, so ist int dom$f \neq \emptyset$ und damit Lin dom$f = \mathbb{R}^n$. Nach Satz 2.6 haben wir somit $\forall a \in \partial f(\bar{x})$

$$f'(\bar{x}, r) \geq < a, r >, \forall r \in \mathbb{R}^n.$$

Da f in $\bar{x}$ differenzierbar, so ist (vgl. 3.Bemerkung zur Definition der Richtungsableitung) $f'(\bar{x}, r) = < \nabla f(\bar{x}), r >$. Also haben wir

$$< \nabla f(\bar{x}), r > \geq < a, r >, \forall r \in \mathbb{R}^n, \forall a \in \partial f(\bar{x}).$$

Das ist aber nur bei $a = \nabla f(\bar{x})$ möglich. (Wähle $r := -\nabla f(\bar{x}) + a$). $\Rightarrow \partial f(\bar{x}) = \{\nabla f(\bar{x})\}$.

$\Rightarrow$ Sei $\partial f(\bar{x})$ ein singleton. Sei $\partial f(\bar{x}) = \{a\}$.

Da Lin dom$f = \mathbb{R}^n$, so haben wir nach Satz 2.6

$$f'(\bar{x}; r) = < a, r >, \forall r \in \mathbb{R}^n. \tag{2.2.7}$$

Weil $\bar{x} \in$ int dom$f$, so existiert ein $\delta > 0$ mit $B_\delta(\bar{x}) \subset$ int dom$f$.

Sei $x \in B_\delta(\bar{x})$, $\alpha \in (0, 1]$. Wir setzen $r := x - \bar{x}$.

Wir betrachten

$$\varphi(\alpha, r) = \frac{f(\bar{x} + \alpha r) - f(\bar{x})}{\alpha} - < a, r > .$$

Nach Folgerung 2.8 ist f auf $B_\delta(\bar{x})$ stetig. Aus $(\bar{x} + \alpha r) \in B_\delta(\bar{x})$, $\alpha > 0$ folgt somit die Stetigkeit von φ auf $[(0, 1] \times (B_\delta(\bar{x}) - \{\bar{x}\})]$.

Weil aber nach Definition von $f'(\bar{x}; r)$

$$\psi(\alpha) := \frac{f(\bar{x} + \alpha r) - f(\bar{x})}{\alpha} \to f'(\bar{x}; r) \stackrel{(2.2.7)}{=} < a, r >,$$

so gilt $\varphi(\alpha, r) \to_{\alpha \to +0} 0$ bei beliebigem r. Somit ist $\varphi(\alpha, r)$ auf $(0, 1] \times (B_\delta(\bar{x}) - \{\bar{x}\})$ beschränkt. Weil $(0, 1] \times (B_\delta(\bar{x}) - \{\bar{x}\})$ kompakt ist, so muß $\varphi(\alpha, r)$ auf $(0, 1] \times (B_\delta(\bar{x}) - \{\bar{x}\})$ gleichmäßig stetig sein.

Zu $\varepsilon > 0$ existiert damit ein $\gamma > 0$, so daß

$$0 \leq \varphi(\alpha, r) \leq \varepsilon, \forall \alpha \in (0, \gamma), \forall r \in (B_\delta(\bar{x}) - \{\bar{x}\}). \tag{2.2.8}$$

($\psi(\alpha)$ ist nach Satz 2.1 monoton nichtfallend).

Sei $r \in (B_\delta(\bar{x}) - \{\bar{x}\})$ und $\|r\| \leq \gamma\delta$ fixiert.

Wir setzen

$$\alpha' = \frac{\|r\|}{\delta}, \;\; r' = \frac{\delta r}{\|r\|}.$$

Dann ist $\alpha' \in (0, \gamma)$ und $\|r'\| = \delta$, d.h. $r' \in (B_\delta(\bar{x}) - \{\bar{x}\})$, womit nach (2.2.8) gilt $0 \leq \varphi(\alpha', r') \leq \varepsilon$.

Weil aber $\bar{x} + r = \bar{x} + \alpha' r'$, so ist damit

$$0 \leq \frac{f(\bar{x} + r) - f(\bar{x}) - < a, r >}{\|r\|} \leq \frac{\varepsilon}{\delta}, \tag{2.2.9}$$

Hier ist δ fixiert und mit $\varepsilon \to 0$ wird $\gamma \to 0$ streben, d.h. wird auch $\|r\| \to 0$.

Folglich existiert

$$\lim_{\|r\| \to 0} \frac{f(\bar{x} + r) - f(\bar{x}) - < a, r >}{\|r\|}$$

und dieser Limes ist nach (2.2.9) gleich Null.

Das aber bedeutet die Differenzierbarkeit von f in $\bar{x}$ und es muß $a = \nabla f(\bar{x})$ gelten.

□

Bemerkung.

1. Nach dem Beweis (und auch nach der ersten Bemerkung zur Definition des Subgradienten) gilt also unter den Bedingungen des Satzes $\partial f(\bar{x}) = \{\nabla f(\bar{x})\}$.
2. Wenn die konvexe Funktion f in $\bar{x} \in \text{int dom} f$ nicht differenzierbar ist, dann ist nach dem Satz 2.9 $\partial f(\bar{x})$ kein singleton.

   Weil $\partial f(\bar{x})$ aber nach Satz 2.4 eine konvexe Menge ist, so muß in diesem Fall $\partial f(\bar{x})$ unendlich viele Elemente enthalten.

Wir wollen abschließend untersuchen, wann $\partial f(x)$ eine beschränkte Menge ist.

Definition. Es sei $A \subset \mathbb{R}^n$, sei $\bar{x} \in A$. Die Menge

$$N_A(\bar{x}) = \{a \in \mathbb{R}^n : < a, x - \bar{x} > \leq 0, \forall x \in A\}$$

heißt Normalenkegel an A im Punkt $\bar{x}$.

Bemerkung.

1. Offensichtlich ist $0 \in N_A(x), \forall x \in A$.

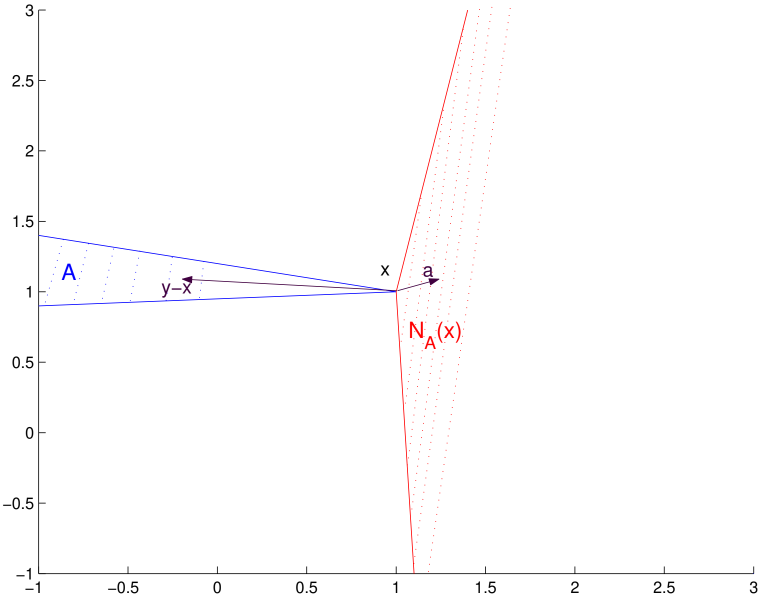
2. Wenn $a \in N_A(\bar{x})$, dann ist $\forall \alpha \geq 0$ auch $\alpha a \in N_A(\bar{x})$. Folglich ist $N_A(\bar{x})$ ein Kegel mit "Spitze" in 0.

3. Wenn $\bar{x} \in \text{int} A$, dann ist $N_A(\bar{x}) = \{0\}$.

   Beweis. Nach 1.Bemerkung ist $0 \in N_A(\bar{x})$. Entgegen dem zu Beweisenden sei auch $a \in N_A(\bar{x})$, $a \neq 0$. Da $\bar{x} \in \text{int} A$, so ist $x = (\bar{x} + \alpha a) \in A$, wenn nur $\alpha > 0$, α - genügend klein $\Rightarrow < a, \bar{x} + \alpha a - \bar{x} > \leq 0 \Rightarrow \alpha \|a\|^2 \leq 0 \overset{\alpha \geq 0}{\Rightarrow} a = 0$, was $a \neq 0$ widerspricht. □

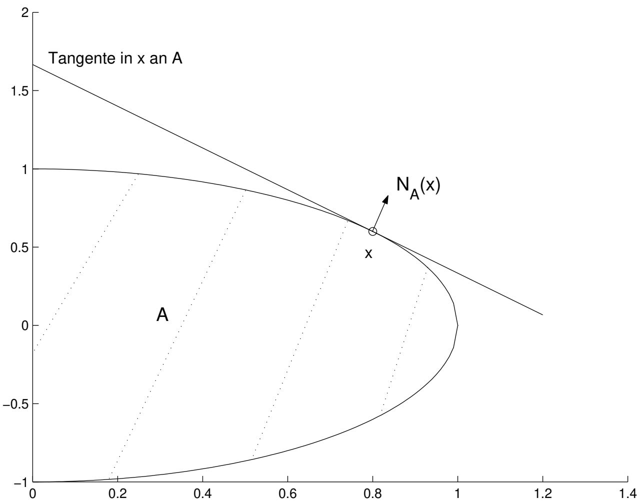
4. Geometrische Illustration im $\mathbb{R}^2$:

   Beispiel 1:



Zu $N_A(x)$ gehören alle Richtungen a, die mit Richtungen $(y-x)$, y beliebig aus A, einen nichtspitzen Winkel bilden.

Beispiel 2:



Wenn A konvex und in einer Umgebung des Randpunktes $\bar{x}$ der Rand von A glatt ist, dann bilden die äußeren Normalen zur Tangentialhyperebene in $\bar{x}$ an den Rand von A den Normalenkegel.

5. In relativen Randpunkten einer konvexen Menge besteht der Normalenkegel nicht nur aus dem Nullelement. Genauer: Es sei $C \subset \mathbb{R}^n$, C - konvex. Sei $\bar{x} \in \text{cl}C \setminus \text{int}C$. Dann ist $N_C(\bar{x}) \neq \{0\}$.

   Beweis. Nach dem Satz über die Existenz einer Stützhyperebene (vgl. Satz 2.18 der Grundvorlesung "Optimierung I") existiert ein $p \neq 0$, so daß

   $$< p, x - \bar{x} > \geq 0, \forall x \in C.$$

   Das aber bedeutet, $(-p) \in N_C(\bar{x})$. □

6. Sei $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ eine konvexe, eigentliche Funktion. Sei $\bar{x} \in \text{dom}f$ und sei $a \in \partial f(\bar{x})$. Dann gilt $a + N_{\text{dom}f}(\bar{x}) \subset \partial f(\bar{x})$.

   Beweis. Nach $a \in \partial f(\bar{x})$ ist

   $$f(x) - f(\bar{x}) \geq < a, x - \bar{x} >, \forall x \in \mathbb{R}^n.$$

   Sei $b \in N_{\text{dom}f}(\bar{x})$. Dann ist $< b, x - \bar{x} > \leq 0$, $\forall x \in \text{dom}f$.

   $$\Rightarrow f(x) - f(\bar{x}) \geq < a + b, x - \bar{x} >, \forall x \in \text{dom}f.$$

   Da aber bei $x \notin \text{dom}f$ diese Ungleichung auch wahr ist, gilt $(a + b) \in \partial f(\bar{x})$. □

Satz 2.10. *Sei $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ eine konvexe, eigentliche Funktion. In $\bar{x} \in domf$ sei $\partial f(\bar{x}) \neq \emptyset$. Wenn $\partial f(\bar{x})$ beschränkt ist, dann ist $\bar{x} \in int\ domf$.*

Beweis. Entgegen dem zu Beweisenden sei $\bar{x} \notin \text{int dom}f$. Dann ist $\bar{x} \in \text{dom}f \setminus \text{int dom}f$, d.h. nach der 5.Bemerkung von oben ist $N_{\text{dom}f}(\bar{x}) \neq \{0\}$. Folglich gibt es ein $p \neq 0$, $p \in N_{\text{dom}f}(\bar{x})$. Nach der zweiten Bemerkung zur Definition des Normalenkegels ist $\forall \alpha > 0$ auch $\alpha p \in N_{\text{dom}f}(\bar{x})$.

Sei $a \in \partial f(\bar{x})$. Nach der 6.Bemerkung von oben ist $(a+\alpha p) \in \partial f(\bar{x})$, $\forall \alpha \geq 0$. Das ist ein Widerspruch zur Beschränkheit von $\partial f(\bar{x})$. □

Bemerkung. Wie wir in Folgerung 2.12 zeigen werden gilt auch die Umkehrung der Behauptung von Satz 2.10.

2.3 Die Subdifferentialabbildung

Wie wir gesehen haben ist i.a. das Subdifferential nicht einelementig. Damit ist ∂f eine Punkt-Menge-Abbildung(PMA).

Definition. Mit 2^P bezeichnen wir die Menge aller Teilmengen von P.

Bemerkung.

1. Es gilt stets $\emptyset \in 2^P$, $P \in 2^P$.
2. Sei $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ eine konvexe Funktion. Die PMA $\partial f : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ werden wir als Subdifferentialabbildung bezeichnen.

Definition. Es sei $D \subset \mathbb{R}^n$ und $F : D \to 2^{\mathbb{R}^m}$ sei eine PMA. Wir sagen, daß F abgeschlossen ist, wenn aus

$$\begin{cases} x^k \in D, y^k \in F(x^k), k = 1, 2, ... \\ \lim_{k\to\infty} x^k = x \in D, \lim_{k\to\infty} y^k = y \in \mathbb{R}^m \end{cases}$$

stets folgt, daß $y \in F(x)$.

Bemerkung.

1. Wenn F eine abgeschlossene PMA ist, dann ist $\forall x \in D$ die Menge $F(x)$ eine abgeschlossene Menge.

   Beweis. Wähle $x^k = x$, $\forall k$. □

2. Anstelle von Abgeschlossenheit spricht man in der Literatur häufig auch von Oberhalbstetigkeit der PMA.

Definition. Sei $D \subset \mathbb{R}^n$, $F : D \to 2^{\mathbb{R}^m}$. Wir sagen, daß die PMA F lokal beschränkt ist, wenn aus

$$\begin{cases} x^k \in D, y^k \in F(x^k), k = 1, 2, ... \\ \lim_{k\to\infty} x^k = x \in D \end{cases}$$

stets folgt, daß $\{y^k\}_{k=1}^{\infty}$ beschränkt ist.

Bemerkung.

Wenn F eine lokal beschränkte PMA ist, dann ist $\forall x \in D$ die Menge $F(x)$ beschränkt.

Beweis. Entgegen dem zu Beweisenden sei für ein $\bar{x} \in D$ die Menge $F(\bar{x})$ nicht beschränkt. Dann existiert in $F(\bar{x})$ eine Folge von Elemente y^k, $k = 1, 2, ...$ mit $\left\|y^k\right\| \to \infty$. Wenn wir nun wählen $x^k := \bar{x}$, $\forall k$, dann erhalten wir einen Widerspruch zur Definition. □

Wenn für die Folge $\left\{x^k\right\}$ aus der Definition gilt $x = \lim_{k\to\infty} x^k$, aber $x \notin D$, dann ist es für die lokale Beschränkheit von F nicht erforderlich, daß $\left\{y^k\right\}_{k=1}^{\infty}$ beschränkt ist.

Definition. Es sei $D \subset \mathbb{R}^n$, $F : D \to 2^{\mathbb{R}^m}$. Wir sagen, daß die PMA F konvexwertig ist, wenn bei beliebigem $x \in D$ die Menge $F(x)$ eine konvexe Menge ist.

Definition. Es sei $D \subset \mathbb{R}^n$, $F : D \to 2^{\mathbb{R}^n}$. Wir sagen, daß die PMA F monoton ist, wenn $\forall x^1, x^2 \in D$, $\forall y^1, y^2$ mit $y^1 \in F(x^1)$, $y^2 \in F(x^2)$ stets gilt $< y^1 - y^2, x^1 - x^2 > \geq 0$.

Bemerkung. Die Monotonie einer PMA ist eine Verallgemeinerung des Begriffes monoton nichtfallend bei Funktionen einer Variablen. Genauer: $f : \mathbb{R} \to \mathbb{R}$ heißt nichtfallend auf $D \subset \mathbb{R}$, wenn aus $x_1, x_2 \in D$, $x_1 \leq x_2$ stets folgt, daß $f(x_1) \leq f(x_2)$. In diesem Fall ist

$$(f(x_1) - f(x_2))(x_1 - x_2) \geq 0.$$

Satz 2.11. *Sei $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ eine konvexe, eigentliche Funktion. Es sei dom $\partial f = \{x \in \mathbb{R}^n : \partial f(x) \neq \emptyset\}$. Für die PMA ∂f : dom $\partial f \to 2^{\mathbb{R}^n}$ gilt dann:*

1. *∂f ist konvexwertig und monoton,*
2. *Auf Teilmengen $D \subset$ dom ∂f, auf denen f stetig bzgl. dom$f$ ist, ist ∂f eine abgeschlossene PMA,*
3. *Auf offenen Teilmengen $D \subset$ dom ∂f ist ∂f eine lokal beschränkte PMA.*

Bemerkung.
1. Da aus $x \notin$ dom$f$ stets folgt $x \notin$ dom ∂f, so ist dom $\partial f \subset$ dom$f$.
2. f stetig bzgl. dom$f$ bedeutet u.a., daß aus $\left\{x^k\right\} \subset$ dom$f$, $\lim_{k\to\infty} x^k = x \in$ dom$f$ stets folgt, daß $f(x^k) \to f(x)$ bei $k \to \infty$.
3. Nach Hilfssatz 2.7 ist ein konvexes f auf $D :=$ ri dom$f$ bzgl. dom$f$ stetig.

Beweis von Satz 2.11.
1. Nach Satz 2.4 ist $\partial f(x)$ bei beliebigem $x \in \mathbb{R}^n$ eine konvexe Menge $\Rightarrow \partial f$ ist eine konvexwertige PMA.
2. Seien $x^1, x^2 \in$ dom ∂f. Seien $a^1 \in \partial f(x^1)$, $a^2 \in \partial f(x^2)$. Dann gilt per Definition der Subgradienten

$$\begin{cases} f(x^2) - f(x^1) \geq < a^1, x^2 - x^1 > \\ f(x^1) - f(x^2) \geq < a^2, x^1 - x^2 >, \end{cases}$$

woraus wir sofort durch Addition erhalten, daß $0 \geq < a^1 - a^2, x^2 - x^1 >$, d.h. ∂f ist monotone PMA.

3. Sei f auf $D \subset \text{dom}\, \partial f$ bzgl. dom$f$ stetig. Wir überprüfen die Abgeschlossenheit von ∂f.

   Seien $x^k \in D$, $a^k \in \partial f(x^k)$, $k = 1, 2, ...$ Weiterhin sei $x = \lim_{k\to\infty} x^k$, $x \in D$, $a = \lim_{k\to\infty} a^k$. Dann haben wir $\forall k$

$$f(y) - f(x^k) \geq < a^k, y - x^k >, \forall y \in \mathbb{R}^n. \tag{2.3.1}$$

   Wegen der Stetigkeit von f bzgl. dom$f$ und da gilt, $D \subset \text{dom}\, \partial f \subset \text{dom} f$ ergibt der Grenzübergang in (2.3.1) bei fixiertem y, daß

$$f(y) - f(x) \geq < a, y - x > .$$

   Da y beliebig war ist somit $a \in \partial f(x)$, d.h. ∂f ist auf D eine abgeschlossene PMA.

4. Wir überprüfen die lokale Beschränktheit von ∂f.

   Sei $D \subset \text{dom}\, \partial f$, D - offen. Es sei

$$\begin{cases} x^k \in D, a^k \in \partial f(x^k), k = 1, 2, ... \\ \lim_{k\to\infty} x^k = x \in D. \end{cases}$$

   Dann gilt $\forall k$

$$f(y) - f(x^k) \geq < a^k, y - x^k >, \forall y \in \mathbb{R}^n. \tag{2.3.2}$$

   Entgegen dem zu Beweisenden sei $\left\{a^k\right\}_{k=1}^{\infty}$ nicht beschränkt. O.E.d.A. gelte bereits $\left\|a^k\right\| \to \infty$. Außerdem können wir o.E.d.A. annehmen, daß $\frac{a^k}{\|a^k\|} \to \bar{a}$ bei $k \to \infty$. Dann ist $\|\bar{a}\| = 1$.

   Aus (2.3.2) erhalten wir $\forall y \in \mathbb{R}^n$, daß

$$\frac{f(y) - f(x^k)}{\|a^k\|} \geq < \frac{a^k}{\|a^k\|}, y - x^k >, \forall k. \tag{2.3.3}$$

   Da D offen ist, so ist nach Folgerung 2.8 f auf D stetig. Deshalb folgt aus $x^k \to x$, daß $f(x^k) \to f(x)$, d.h. $\{f(x^k)\}_{k=1}^{\infty}$ ist eine beschränkte Folge. Da schließlich für $y \in \text{dom} f$ auch $f(y)$ endlich ist, so folgt aus (2.3.3) für fixiertes $y \in \text{dom} f$

$$0 \geq < \bar{a}, y - x > . \tag{2.3.4}$$

   Weil $x \in D$ und D offen ist, so ist $(x + \alpha\bar{a}) \in D$, wenn $\alpha > 0$, α - genügend klein.

   Nach $D \subset \text{dom} f$ erhalten wir aus (2.3.4), daß $0 \geq < \bar{a}, x + \alpha\bar{a} - x >$ für gewisse $\alpha > 0$, d.h. $0 \geq \|\bar{a}\|^2 \Rightarrow \bar{a} = 0$, was $\|\bar{a}\| = 1$ widerspricht.

$\square$

Bemerkung.

1. Im obigen Beispiel $f(x) = |x|$ mit

$$\partial f(x) = \begin{cases} 1, & x > 0 \\ [-1, 1], & x = 0 \\ -1, & x < 0 \end{cases}$$

beobachten wir, daß die Gradienten von f für $x \neq 0$ in $\partial f(0)$ enthalten sind. Nach Satz 2.11 ist das nicht zufällig so, denn wenn das konvexe f auf einer Teilmenge D bzgl. dom$f$ stetig ist, so folgt aus

$$\begin{cases} x = \lim_{k\to\infty} x^k, x \in D, \{x^k\}_{k=1}^{\infty} \subset D \\ a = \lim_{k\to\infty} a^k, a^k \in \partial f(x^k), k = 1, 2, ..., \end{cases} \tag{2.3.5}$$

daß $a \in \partial f(x)$ (die sogenannte Oberhalbstetigkeit der PMA ∂f).

Das Beispiel $f(x) = |x|$ zeigt aber auch, daß auf diese Weise nicht alle Elemente von $\partial f(0)$ entstehen müssen (d.h. $\partial f(x^k) \to \partial f(x)$ ist nicht zu beobachten). Das liegt an der fehlenden Unterhalbstetigkeit der PMA ∂f (siehe weiter unten).

2. Im Beispiel $f(x) = |x|$ ist f in $x \neq 0$ differenzierbar und $\partial f(x) = \{\nabla f(x)\}$ bei $x \neq 0$. Die Forderungen (2.3.5) bedeuten also in diesem Fall

$$\begin{cases} x = \lim_{k\to\infty} x^k, x \in D, \{x^k\}_{k=1}^{\infty} \subset D \\ a = \lim_{k\to\infty} \nabla f(x^k) \end{cases}$$

(d.h. der Grenzwert soll existieren) und die Aussage $a \in \partial f(x)$ bleibt.

Diese Aussage kann man zum Ausgangspunkt einer Verallgemeinierung des Subdifferentials für konvexe Funktion auf nichtkonvexe Funktionen nehmen:

Sei $f : \mathbb{R}^n \to \mathbb{R}$ eine sogenannte lokal Lipschitzstetige Funktion, d.h. $\forall x \in \mathbb{R}^n$ existiere ein $\varepsilon > 0$ und ein $L = L(x)$, so daß

$$|f(y) - f(z)| \leq L \|y - z\|, \forall y, z \in U_\varepsilon(x).$$

Nach einem Satz von Rademacher sind solche Funktionen fast überall differenzierbar. Deshalb scheint folgende Definition sinnvoll:

Definition. Sei $f : \mathbb{R}^n \to \mathbb{R}$ und $\bar{x} \in \mathbb{R}^n$. Unter der Menge $S(\bar{x})$ verstehen wir alle Vektoren $a \in \mathbb{R}^n$ für die folgendes gilt: Es existiert eine Folge $\{x^k\}_{k=1}^{\infty}$ mit $\bar{x} = \lim_{k\to\infty} x^k$ so, daß f in allen x^k, $k = 1, 2, ...,$ differenzierbar ist und $a = \lim_{k\to\infty} \nabla f(x^k)$ gilt.

Als Analogon des Subdifferentials $\partial f(x)$ konvexer Funktionen f führte F.Clarke für lokal Lipschitzstetige $f : \mathbb{R}^n \to \mathbb{R}$ die Menge conv$(S(\bar{x}))$ als Subdifferential ein. Üblich ist die Bezeichnung

$$\partial_{\text{Cl}} f(\bar{x}) := \text{conv}(S(\bar{x})).$$

Wir vermerken: Wenn $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ konvex und eigentlich ist, dann ist f auf ri dom$f$ lokal Lipschitzstetig und es gilt

$$\partial f(x) = \partial_{\mathrm{Cl}} f(x), \forall x \in \text{ri dom} f.$$

Folgerung 2.12. *Sei $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ konvexe, eigentliche Funktion. In $\bar{x} \in$ dom$f$ sei $\partial f(\bar{x}) \neq \emptyset$. Dann ist die Menge $\partial f(\bar{x})$ beschränkt $\Leftrightarrow \bar{x} \in$ int dom$f$.*

Beweis. 1. Die Notwendigkeit der Aussage wurde bereits mit Satz 2.10 bewiesen.

2. Sei deshalb $\bar{x} \in$ int dom$f$. Dann existiert ein $\varepsilon > 0$, so daß $U_\varepsilon(\bar{x}) \subset$ dom$f$. O.E.d.A. sei $U_\varepsilon(\bar{x})$ offen. Wir setzen

$$F(x) = \begin{cases} f(x), & x \in U_\varepsilon(\bar{x}) \\ +\infty, & \text{sonst.} \end{cases}$$

Nach Satz 2.11 ist ∂F auf dom$F = U_\varepsilon(\bar{x}) =$ dom∂F eine lokal beschränkte PMA. Nach der Bemerkung zur Definition der lokalen Beschränktheit ist $\partial F(\bar{x})$ eine beschränkte Menge. Wenn wir zeigen, daß $\partial f(\bar{x}) = \partial F(\bar{x})$, so ist der Beweis geführt.

3. Wir zeigen $\partial f(\bar{x}) \subset \partial F(\bar{x})$:

Aus $a \in \partial f(\bar{x})$ haben wir

$$f(x) - f(\bar{x}) \geq < a, x - \bar{x} >, \forall x \in \mathbb{R}^n.$$

Da $f(\bar{x}) = F(\bar{x})$ und $F(x) \geq f(x)$, $\forall x$, so ergibt sich $F(x) - F(\bar{x}) \geq < a, x - \bar{x} >, \forall x \in \mathbb{R}^n$, d.h. $a \in \partial F(\bar{x})$.

4. Wir zeigen $\partial F(\bar{x}) \subset \partial f(\bar{x})$:

Aus $a \in \partial F(\bar{x})$ erhalten wir

$$F(x) - F(\bar{x}) \geq < a, x - \bar{x} >, \forall x \in \mathbb{R}^n. \tag{2.3.6}$$

Für $z \in U_\varepsilon(\bar{x})$ haben wir $F(z) = f(z)$ und damit gilt nach (2.3.6)

$$f(z) - f(\bar{x}) \geq < a, z - \bar{x} >, \forall z \in U_\varepsilon(\bar{x}). \tag{2.3.7}$$

Sei jetzt $z \notin U_\varepsilon(\bar{x})$, d.h. (o.E.d.A. sei $U_\varepsilon(\bar{x})$ eine Kugelumgebung)

$$\|z - \bar{x}\| \geq \varepsilon. \tag{2.3.8}$$

Dann existiert ein $\alpha > 0$, so daß $\bar{z} = [\bar{x} + \frac{\alpha}{\|z-\bar{x}\|}(z - \bar{x})] \in U_\varepsilon(\bar{x})$, d.h. $\|\bar{z} - \bar{x}\| = \left\| \frac{\alpha}{\|z-\bar{x}\|}(z - \bar{x}) \right\| = \alpha < \varepsilon$.

Nach (2.3.8) ist damit $\|z - \bar{x}\| > \alpha$, d.h. $\frac{\alpha}{\|z-\bar{x}\|} < 1$. Weil $\bar{z} \in U_\varepsilon(\bar{x})$, so ist nach dem bereits bewiesenen Teil (Formel (2.3.7))

$$f(\bar{z}) - f(\bar{x}) \geq < a, \bar{z} - \bar{x} > . \tag{2.3.9}$$

Mit $\beta := \frac{\alpha}{\|z-\bar{x}\|} \in (0,1)$ haben wir per Definition von $\bar{z}$, daß $\bar{z} = \bar{x} + \beta(z - \bar{x}) = \beta z + (1-\beta)\bar{x}$, d.h. nach (2.3.9) erhalten wir unter Benutzung der Konvexität von f

$$\beta f(z) + (1-\beta) f(\bar{x}) - f(\bar{x}) \geq f(\bar{z}) - f(\bar{x}) \geq < a, \bar{z} - \bar{x} > = < a, \beta z - \beta \bar{x} >,$$

d.h. $\beta[f(z) - f(\bar{x})] \geq \beta < a, z - \bar{x} >$.

Aufgrund von $\beta > 0$ bedeutet das

$$f(z) - f(\bar{x}) \geq < a, z - \bar{x} >, \forall z \notin U_\varepsilon(\bar{x}),$$

woraus zusammen mit (2.3.7) folgt $a \in \partial f(\bar{x})$.

□

Folgerung 2.13. *Sei $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ eine konvexe, eigentliche Funktion. Sei $D \subset \mathrm{dom} f$ eine offene Menge (bzgl. $\mathbb{R}^n$). f sei auf D differenzierbar. Dann ist ∇f auf D stetig.*

Beweis. Nach Satz 2.9 ist $\forall x \in D$ $\partial f(x) = \{\nabla f(x)\}$.

Die Abbildung $\nabla f : D \to \mathbb{R}^n$ ist nach Satz 2.11 auf D lokal beschränkt und da nach Satz 2.9 f auf D stetig (bzgl. $\mathbb{R}^n$) ist, so ist wiederum nach Satz 2.11 die Abbildung $\nabla f : D \to \mathbb{R}^n$ auch abgeschlossen.

Wie man leicht überprüft, ist eine eindeutige Abbildung (unser ∇f), die abgeschlossen ist, auf int dom$f$ eine stetige Abbildung. □

Die oben eingeführte Abgeschlossenheit (auch Oberhalbstetigkeit genannt) einer PMA ist bei Funktionen (d.h. einwertigen PMA) äquivalent mit der Stetigkeit. Bei wirklichen PMA (d.h. mehrwertigen PMA) folgt die analoge Eigenschaft nicht aus der Abgeschlossenheit. Genauer: Wenn $F : D \to 2^{\mathbb{R}^m}$ abgeschlossen ist, so folgt aus $x^k \to \bar{x}$ in allgemeinen nicht, daß $F(x^k) \to F(\bar{x})$.

Das führt auf den Begriff der Unterhalbstetigkeit.

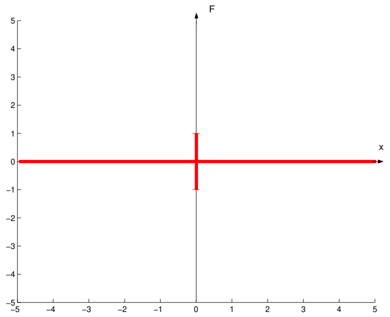
Definition. Sei $F : D \to 2^{\mathbb{R}^m}$ mit $D \subset \mathbb{R}^n$ eine PMA. Wir sagen, daß F in $\bar{x} \in D$ unterhalbstetig ist, wenn aus $x^k \to \bar{x}$ und $\bar{y} \in F(\bar{x})$ stets die Existenz einer Folge $\{y^k\}$ mit $y^k \in F(x^k)$, $k = 1, 2, ...$, folgt, so daß $y^k \to \bar{y}$.

Bemerkung.

1. Bei Funktionen (d.h. einwertigen PMA) ist die Unterhalbstetigkeit im obigen Sinn wieder äquivalent mit der Stetigkeit: wähle $y^k := F(x^k), \forall k$.

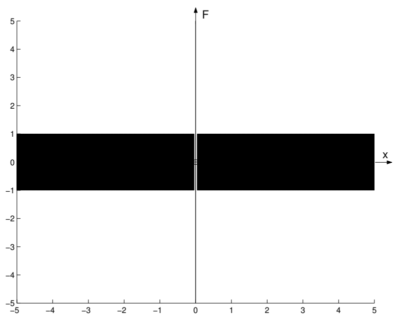
2. Beispiele:

$$F_1(x) = \begin{cases} \{0\}, & x \neq 0 \\ [-1,1], & x = 0. \end{cases}$$



F_1 ist in $\bar{x} = 0$ nicht unterhalbstetig.

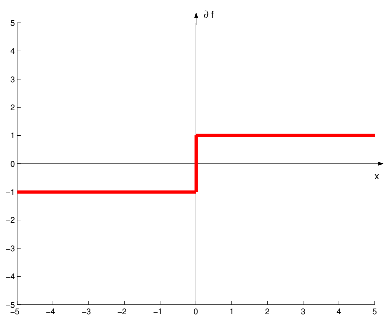
$$F_2(x) = \begin{cases} [-1,1], & x \neq 0 \\ \{0\}, & x = 0. \end{cases}$$



F_2 ist in $\bar{x} = 0$ unterhalbstetig.

3. Wie das Beispiel $f(x) = |x|$, $x \in \mathbb{R}$, zeigt, muß die Subdifferentialabbildung keine unterhalbstetige PMA sein:

$$\partial f(x) = \begin{cases} 1, & x > 0 \\ [-1,1], & x = 0 \\ -1, & x < 0. \end{cases}$$



∂f ist in $\bar{x} = 0$ nicht unterhalbstetig, da z.B. bei $\bar{y} = \frac{1}{2} \in \partial f(0)$ und $x^k = \frac{1}{k}, \forall k$, zwar $x^k \to \bar{x} = 0$ gilt, aber keine Elemente $y^k \in \partial f(x^k)$ existieren, für die $y^k \to \frac{1}{2}$.

Durch ε-Aufweitung des Subdifferentials kann man aber zu einer unterhalbstetigen PMA kommen.

Definition. Sei $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ eine konvexe, eigentliche Funktion. Sei $\varepsilon \geq 0$. Die Menge

$$\partial_\varepsilon f(\bar{x}) = \{a \in \mathbb{R}^n : f(x) - f(\bar{x}) \geq < a, x - \bar{x} > -\varepsilon, \forall x \in \mathbb{R}^n\}$$

heißt ε-Subdifferential von f in $\bar{x}$. Jedes $a \in \partial_\varepsilon f(\bar{x})$ heißt ε-Subgradient von f in $\bar{x}$.

Bemerkung.

1. Bei $\varepsilon = 0$ ist $\partial f(\bar{x}) = \partial_\varepsilon f(\bar{x}), \forall \bar{x} \in \mathbb{R}^n$.
2. $\partial f(\bar{x}) \subset \partial_\varepsilon f(\bar{x}), \forall \varepsilon \geq 0, \forall \bar{x} \in \mathbb{R}^n$.
3. Analog wie im Satz 2.4 zeigt man, daß $\forall x \in \mathbb{R}^n$ die Menge $\partial_\varepsilon f(x)$ konvex und abgeschlossen ist.
4. Analog wie in der 4.Bemerkung zur Definition des Subgradienten (vgl. Abschnitt 2.2.1) zeigt man, daß a ein ε-Subgradient von f in $\bar{x}$ genau dann ist, wenn die durch den Punkt $[\bar{x}, f(\bar{x}) - \varepsilon]$ gehende Hyperebene mit dem Stellungsvektor $\begin{bmatrix} a \\ -1 \end{bmatrix}$ überall unterhalb des Epigraphen von f verläuft.
5. Analog wie im Satz 2.11 zeigt man, daß $\partial_\varepsilon f : D \to 2^{\mathbb{R}^n}$ eine abgeschlossene PMA ist, wenn $D \subset \mathrm{dom}\partial_\varepsilon f$ und f auf D bzgl. $\mathrm{dom}f$ stetig ist.

Satz 2.14 (Asplund/Rockafellar, 1969). *Sei $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ eine konvexe, eigentliche Funktion. Sei $domf$ abgeschlossen. Dann ist $\partial_\varepsilon f : [0, \infty) \times \mathbb{R}^n \to 2^{\mathbb{R}^n}$ in jedem Punkt $[\varepsilon^0, x^0]$ mit $\varepsilon^0 > 0$, $x^0 \in domf$ unterhalbstetig.*

Bemerkung. Wenn die PMA F in $\bar{x}$ sowohl abgeschlossen als auch unterhalbstetig ist, so sagt man, daß F in $\bar{x}$ Kakutani-stetig ist.

2.4 Die Subdifferentialrechnung

Wir wollen für das Weitere folgende abkürzende Bezeichnungen vereinbaren: Es seien $F_1, F_2 : D \to 2^{\mathbb{R}^n}$. Wir schreiben $F_1 \cap F_2 \neq \emptyset$, wenn $\forall x \in D$ gilt $F_1(x) \cap F_2(x) \neq \emptyset$. Wir schreiben $F_1 \subset F_2$, wenn $\forall x \in D$ gilt $F_1(x) \subset F_2(x)$. Analog soll $F_1 = F_2$ für $F_1(x) = F_2(x)$, $\forall x \in D$ stehen. Für $F : D \to 2^{\mathbb{R}^n}$ bezeichnen wir mit $\mathrm{conv}(F) : D \to 2^{\mathbb{R}^n}$ die PMA, welche jedem $x \in D$ die Menge $\mathrm{conv}(F(x))$ zuordnet.

Bei konvexwertiger PMA F ist dann $F = \mathrm{conv}(F)$.

Hilfssatz 2.15. *Sei $D \subset \mathbb{R}^n$ eine offene Menge. Es seien $F_1, F_2 : D \to 2^{\mathbb{R}^n}$, wobei*

$$\begin{cases} F_1\text{- eine abgeschlossene, lokal beschränkte PMA} \\ F_2\text{- eine monotone PMA} \end{cases}$$

sind. Aus $F_1 \cap F_2 \neq \emptyset$ folgt dann, daß $conv(F_2) \subset conv(F_1)$.

Beweis. 1. Wenn es uns gelingt zu zeigen, daß $F_2 \subset \mathrm{conv}(F_1)$, dann ist auch die Behauptung bewiesen.

Sei $x \in D$. Nach der jeweils ersten Bemerkung zur Definition der Abgeschlossenheit bzw. der lokalen Beschränktheit (vgl. Anfang von Abschnitt 2.3) ist $F_1(x)$ eine abgeschlossene, beschränkte Menge. Damit ist $\forall x \in D$ die Menge $F_1(x)$ kompakt. Nach dem Satz von Caratheodory (Folgerung 2.8 der Grundvorlesung "Optimierung I") ist dann bei $x \in D$ auch $\mathrm{conv}(F_1(x))$ eine kompakte Menge.

2. Sei $x \in D$ und $y \in F_2(x)$. Um $F_2 \subset \mathrm{conv}(F_1)$ zu zeigen, müssen wir nachweisen, daß $y \in \mathrm{conv}(F_1(x))$. Entgegen dem zu Beweisenden sei das nicht so. Dann sind die Mengen $\{y\}$ und $\mathrm{conv}(F_1(x))$ stark trennbar (vgl. Folgerung 2.18 der Grundvorlesung "Optimierung I"), d.h. es existiert ein $p \in \mathbb{R}^n$, so daß

$$< p, z >< < p, y >, \forall z \in \mathrm{conv}(F_1(x)). \qquad (2.4.1)$$

Wir betrachten die Elemente

$$x^k = x + \frac{1}{k}p, \;\; k = 1, 2, ...$$

Da $x \in D$ und D offen ist, so ist $x^k \in D$ für genügend große k.

Da $F_1 \cap F_2 \neq \emptyset$, so existiert für solche k ein y^k mit $y^k \in F_1(x^k) \cap F_2(x^k)$.

Da F_1 lokal beschränkt ist, so muß die Folge $\left\{y^k\right\}$ beschränkt sein. Folglich existiert für $\left\{y^k\right\}$ eine konvergente Teilfolge. O.B.d.A. sei bereits

$$\bar{y} = \lim_{k\to\infty} y^k. \qquad (2.4.2)$$

Da per Konstruktion der x^k gilt $x = \lim_{k\to\infty} x^k$, $x \in D$ laut Voraussetzung und da F_1 eine abgeschlossene PMA ist, so muß

$$\bar{y} \in F_1(x) \qquad (2.4.3)$$

gelten.

Weil laut Voraussetzung $y \in F_2(x)$ und weil per Wahl der y^k stets $y^k \in F_2(x^k)$, so gilt aufgrund der Monotonie von F_2, daß

$$< y^k - y, x^k - x > \geq 0, \;\; \forall k\text{- genügend groß.}$$

Damit (weil $x^k = x + \frac{1}{k}p$) ist $< y^k - y, p > \geq 0$, wenn nur k genügend groß ist. Nach (2.4.2) erhalten wir $< \bar{y} - y, p > \geq 0$.

Das aber widerspricht (2.4.1) und (2.4.3), w.z.b.w.

□

Folgerung 2.16. *Es sei $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ eine konvexe, eigentliche Funktion. Es sei $\mathrm{dom} f$ eine offene Menge und $F : \mathrm{dom} f \to 2^{\mathbb{R}^n}$ eine PMA, wobei $F(x) \neq \emptyset$ für beliebiges $x \in \mathrm{dom} f$ sei. Dann gilt*

1. *Wenn F abgeschlossen und lokal beschränkt ist und $F \cap \partial f \neq \emptyset$, dann ist $\partial f \subset conv(F)$.*

2. *Wenn F monoton ist und $F \cap \partial f \neq \emptyset$, dann ist $conv(F) \subset \partial f$.*

3. *Wenn F abgeschlossen, lokal beschränkt und monoton sowie $F \cap \partial f \neq \emptyset$, dann ist $\partial f = conv(F)$.*

4. *Wenn F abgeschlossen und $F \subset \partial f$, dann ist $\partial f = conv(F)$.*

5. *Wenn F monoton ist und $\partial f \subset F$, dann ist $\partial f = conv(F)$.*

Bemerkung. 1. Behauptung (4) bedeutet, daß in der Klasse der abgeschlossenen, konvexwertigen PMA die Subdifferentialabbildungen die Minimalelemente sind.

2. Behauptung (5) bedeutet, daß in der Klasse der monotonen PMA, die Subdifferentialabbildungen die Maximalelemente sind.

Beweis der Folgerungen 2.16. 1. Wir setzen $F_1 := F$ und $F_2 := \partial f$. Nach Satz 2.11 ist F_2 monoton und unter den Voraussetzungen von (1) erfüllen F_1, F_2 die Bedingungen des Hilfssatzes 2.15. Danach gilt $\mathrm{conv}(\partial f) \subset \mathrm{conv}(F)$. Nach Satz 2.11 ist aber ∂f konvexwertig, d.h. es ist $\mathrm{conv}(\partial f) = \partial f$.

2. Wir setzen $F_1 := \partial f$, $F_2 := F$. Weil $\mathrm{dom} f$ als offen vorausgesetzt wird, so ist F_1 nach Behauptung (3) von Satz 2.11 lokal beschränkt. Nach Folgerung 2.8 ist f auf $\mathrm{dom} f$ stetig und damit ist nach Behauptung (2) von Satz 2.11 F_1 abgeschlossen. Damit erfüllen F_1, F_2 die Bedingungen von Hilfssatz 2.15 und es gilt folglich

$$\mathrm{conv}(F) \subset \mathrm{conv}(\partial f) = \partial f.$$

3. folgt aus 1. und 2.

4. Aus $F \subset \partial f$ folgt

$$conv(F) \subset \mathrm{conv}(\partial f) = \partial f. \tag{2.4.4}$$

Mit derselben Begründung wie im Beweis von 2. ist ∂f unter unseren Bedingungen lokal beschränkt und abgeschlossen.

Da $F \subset \partial f$ vorausgesetzt, so folgt aus $F(x) \neq \emptyset, \forall x \in \mathrm{dom} f$, daß $F \cap \partial f \neq \emptyset$. Damit sind die Voraussetzungen von (1) erfüllt und wir haben deshalb $\partial f \subset \mathrm{conv}(F)$. Zusammen mit (2.4.4) ergibt sich daraus die Behauptung.

5. Es sind die Voraussetzungen von 2. erfüllt (nach Satz 2.4 ist $\partial f \neq \emptyset$, so daß aus $\partial f \subset F$ folgt, daß $\partial f \cap F \neq \emptyset$). Deshalb haben wir $\mathrm{conv}(F) \subset \partial f$. Da die umgekehrte Inklusion aus der Voraussetzung $\partial f \subset F$ folgt, so gilt die Behauptung (5).

$\square$

Satz 2.17. *Sei $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ eine konvexe, eigentliche Funktion. Es sei $\lambda \geq 0$. Wir setzen $F(x) = \lambda f(x)$. Dann gilt:*

1. *Bei $\lambda > 0$ ist $\partial F(x) = \lambda \partial f(x)$, $\forall x \in \mathbb{R}^n$.*
2. *Bei $\lambda = 0$ ist*
$$\{0\} = \partial F(x) = \lambda \partial f(x), \forall x \in dom\ \partial f,$$
$$\{0\} = \partial F(x) \supset \lambda \partial f(x) = \emptyset, \forall x \notin dom\ \partial f.$$

Beweis. 1. Wir betrachten zuerst den Fall $\lambda = 0$.

Nach unserer Verabredung vom Ende des Abschnittes 1.9 ist $\lambda f(x) = 0$ unabhängig davon ob $f(x) < +\infty$ oder $f(x) = +\infty$. Deshalb ist $F(x) = 0$ $\forall x \in \mathbb{R}^n$ und damit $\partial F(x) = \{0\}$ (wende z.B. Satz 2.9 und die 1.Bemerkung dazu an). Bei $x \in$ dom ∂f ist per Definition $\partial f \neq \emptyset$ und damit ist $\lambda \partial f = \{0\}$.

Bei $x \notin$ dom ∂f ist $\partial f = \emptyset$ und damit auch $\lambda \partial f = \emptyset$.

2. Sei $\lambda > 0$. Wir zeigen $\lambda \partial f(x) \subset \partial F(x)$. Sei $a \in \lambda \partial f(x) \Rightarrow a = \lambda b$, wobei $b \in \partial f(x)$.
$$\begin{aligned}
&\Rightarrow f(y) - f(x) \geq < b, y - x >, \forall y \in \mathbb{R}^n \\
&\Rightarrow \lambda f(y) - \lambda f(x) \geq < \lambda b, y - x >, \forall y \in \mathbb{R}^n, (\text{weil } \lambda > 0) \\
&\Rightarrow F(y) - F(x) \geq < a, y - x >, \forall y \in \mathbb{R}^n \\
&\Rightarrow a \in \partial F(x).
\end{aligned}$$

3. Sei $\lambda > 0$. Wir zeigen $\partial F(x) \subset \lambda \partial f(x)$. Sei $a \in \partial F(x)$
$$\begin{aligned}
&\Rightarrow F(y) - F(x) \geq < a, y - x >, \forall y \in \mathbb{R}^n \\
&\Rightarrow \lambda f(y) - \lambda f(x) \geq < a, y - x >, \forall y \in \mathbb{R}^n \\
&\Rightarrow f(y) - f(x) \geq < \frac{a}{\lambda}, y - x >, \forall y \in \mathbb{R}^n, (\text{weil } \lambda > 0) \\
&\Rightarrow \frac{a}{\lambda} \in \partial f(x) \\
&\Rightarrow a = \lambda \frac{a}{\lambda} \in \lambda \partial f(x).
\end{aligned}$$

□

Satz 2.18. *Seien $f_1, f_2, ..., f_m : \mathbb{R}^n \to \bar{\mathbb{R}}$ konvexe, eigentliche Funktionen. Wir setzen $F(x) = f_1(x) + f_2(x) + ... + f_m(x)$. Dann gilt:*

(i) $\forall x \in \mathbb{R}^n$ *ist*
$$\partial F(x) \supset \partial f_1(x) + \partial f_2(x) + ... + \partial f_m(x).$$

(ii) Seien $f_1, f_2, ..., f_k$ polyedral (d.h. epif_i ist eine polyedrale Menge - vgl. Definition der Polyedralität im Abschnitt 1.4.1 der Grundvorlesung "Optimierung I"). Wenn $(\cap_{i=1}^k domf_i) \cap (\cap_{i=k+1}^m ri\ domf_i) \neq \emptyset$, dann gilt $\forall x \in \mathbb{R}^n$
$$\partial F(x) = \partial f_1(x) + \partial f_2(x) + ... + \partial f_m(x).$$

Beweis. 1. Sei $x \in \mathbb{R}^n$. Wenn bei wenigstens einem i_0 die Menge $\partial f_{i_0} = \emptyset$, dann ist $\sum_{i=1}^m \partial f_i(x) = \emptyset$ und Behauptung (i) gilt.

Sei deshalb $\forall i = 1, 2, ..., m$ die Menge $\partial f_i(x) \neq \emptyset$. Sei $a \in \sum_{i=1}^m \partial f_i(x)$. Dann existieren $a^i \in \partial f_i(x), i = 1, 2, ..., m$, so daß $a = \sum_{i=1}^m a^i$.

$a^i \in \partial f_i(x)$ bedeutet: $f_i(y) - f_i(x) \geq < a^i, y - x >, \forall y$.

$$\Rightarrow \sum_{i=1}^m f_i(y) - \sum_{i=1}^m f_i(x) \geq < \sum_{i=1}^m a^i, y - x >, \forall y \in \mathbb{R}^n,$$

d.h. $F(y) - F(x) \geq < a, y - x >, \forall y \in \mathbb{R}^n$, was $a \in \partial F(x)$ bedeutet und damit gilt (i) auch in diesem Fall.

2. Wir zeigen (ii) nur für den Fall $k = 0$.

   (a) Mit Hilfe der Folgerung 2.16 können wir die Behauptung unter der verschärften Voraussetzung $\cap_{i=1}^m \text{int dom} f_i \neq \emptyset$ sofort zeigen.

   Wir setzen dazu
   $$\begin{cases} D := \cap_{i=1}^m \text{int dom} f_i, \\ \mathbb{F}(x) = \sum_{i=1}^m \partial f_i(x). \end{cases}$$

   Dann ist D offen, $\mathbb{F} : D \to 2^{\mathbb{R}^n}$ eine PMA. Nach Teil (i) ist $\mathbb{F}(x) \subset \partial F(x), \forall x$, d.h. $\mathbb{F} \subset \partial F$.

   Nach Satz 2.4 ist $\mathbb{F}(x) \neq \emptyset, \forall x \in D$.

   Unter Hinzuziehung von Folgerung 2.8 ergibt sich aus Behauptung (2) von Satz 2.11, daß $\mathbb{F}$ eine abgeschlossene PMA ist. Damit sind alle Voraussetzungen der Behauptung (4) von Folgerung 2.16 erfüllt und es gilt damit $\partial F(x) = \text{conv}(\mathbb{F}(x)), \forall x \in \mathbb{R}^n$.

   Da aber jedes $\partial f_i(x)$ eine konvexe Menge ist (vgl. Satz 2.4), so ist $\mathbb{F}(x) = \sum_{i=1}^m \partial f_i(x)$ eine konvexe Menge und damit $\text{conv}(\mathbb{F}(x)) = \mathbb{F}(x)$. Also gilt (ii).

   (b) Im Falle $\cap_{i=1}^m \text{ri dom} f_i \neq \emptyset$ zeigen wir die Behauptung per vollständiger Induktion nach m.

   Wenn wir sie für $m = 2$ bewiesen haben, dann ist einerseits die Induktionsbasis gezeigt und der Induktionsschluß folgt aus den Eigenschaften, daß $(\text{ri dom} f_{k+1}) \cap (\cap_{i=1}^k)\text{ri dom} f_i) = \cap_{i=1}^{k+1} \text{ri dom} f_i$ und daß damit nach der Induktionsbasis

   $$\partial(f_{k+1} + \sum_{i=1}^k f_i)(x) = \partial f_{k+1}(x) + \partial(\sum_{i=1}^k f_i)(x) = \partial f_{k+1}(x) + \sum_{i=1}^k \partial f_i(x),$$

   wobei die letzte Gleichung aus der Induktionsannahme folgt.

   Es sei deshalb $m = 2$, d.h. $F(x) = f_1(x) + f_2(x)$. Behauptung (ii) ist aufgrund der bereits bewiesenen Behauptung (i) gezeigt, wenn wir nachweisen, daß aus $a \in \partial F(x)$ folgt $a \in \partial f_1(x) + \partial f_2(x)$. Sei deshalb $x \in \mathbb{R}^n$, $a \in \partial F(x)$.

Wenn $x \notin \text{dom} f_1 \cap \text{dom} f_2$, dann gilt $\partial F(x) \subset \mathbb{F}(x)$. Sei deshalb im weiteren $x \in \text{dom} f_1 \cap \text{dom} f_2$.
Wir setzen

$$\begin{aligned}
C_1 &:= \text{dom} f_1 - \{x\} \\
C_2 &:= \text{dom} f_2 - \{x\} \\
g_1(y) &:= f_1(x+y) - f_1(x) - < a, y > \\
g_2(y) &:= f_2(x) - f_2(x+y) \\
G(y) &:= g_1(y) - g_2(y) \\
&= f_1(x+y) + f_2(x+y) - (f_1(x) + f_2(x)) - < a, y > .
\end{aligned}$$

Nach der letzten Bemerkung aus Abschnitt 1.9 sind C_1 und C_2 konvexe Mengen.
Wir haben

$$\left\{ \begin{array}{l}
g_1 : C_1 \to \mathbb{R}, g_1 \text{ auf } C_1 \text{ konvex} \\
g_2 : C_2 \to \mathbb{R}, g_2 \text{ auf } C_2 \text{ konkav} \\
0 \in C_1, g_1(0) = 0 \\
0 \in C_2, g_2(0) = 0 \\
\text{ri} C_1 \cap \text{ri} C_2 = \text{ri dom} f_1 \cap \text{ri dom} f_2 - \{x\} \neq \emptyset \text{ laut Voraussetzung} \\
\partial G(y) = \partial F(x+y) - a, \text{ was man direkt aus der Definition} \\
\text{des Subdifferentials erhalten kann(d.h. ohne Benutzung} \\
\text{der Behauptung (ii) des Satzes).} \\
\text{Damit ist } \partial G(0) = \partial F(x) - a. \\
\text{Weil } a \in \partial F(x), \text{ so ist } 0 \in \partial G(0). \\
\text{Da } G(y) \text{ eine auf } C_1 \cap C_2 \text{ konvexe Funktion ist,} \\
\text{so bedeutet } 0 \in \partial G(0) \text{nach der Definition des Subgradienten} \\
G(y) \geq G(0), \forall y \in C_1 \cap C_2.
\end{array} \right.$$

Also haben wir

$$g_1(y) \geq g_2(y), \forall y \in C_1 \cap C_2.$$

Damit sind alle Voraussetzungen von Hilfssatz 2.3 erfüllt. Es existiert deshalb eine affin-lineare Funktion $l(y) =< b, y > +\beta$ mit

$$\left\{ \begin{array}{l} g_1(y) \geq l(y), \forall y \in C_1 \\ g_2(y) \leq l(y), \forall y \in C_2. \end{array} \right.$$

Da $x \in \text{dom} f_1 \cap \text{dom} f_2$, so ist $0 \in C_1 \cap C_2$.
Aufgrund $g_1(0) = g_2(0) = 0$ ergibt sich $\beta = 0$. Damit bedeuten die Ungleichungen, daß

$$\left\{ \begin{array}{l} f_1(x+y) - f_1(x) - < a, y > \geq < b, y >, \forall y \in (\text{dom} f_1 - \{x\}) \\ f_2(x) - f_2(x+y) \leq < b, y >, \forall y \in (\text{dom} f_2 - \{x\}), \end{array} \right.$$

d.h.

$$\left\{ \begin{array}{l} f_1(z) - f_1(x) \geq < a + b, z - x >, \forall z \in \text{dom} f_1 \\ f_2(x) - f_2(z) \leq < b, z - x >, \forall z \in \text{dom} f_2. \end{array} \right.$$

Damit ist

$$\begin{cases} f_1(z) - f_1(x) \geq < a+b, z-x >, \forall z \in \mathbb{R}^n \\ f_2(z) - f_2(x) \geq < -b, z-x >, \forall z \in \mathbb{R}^n, \end{cases}$$

was bedeutet

$$\begin{cases} (a+b) \in \partial f_1(x) \\ -b \in \partial f_2(x). \end{cases}$$

Daraus ergibt sich aber $a \in (\partial f_1(x) + \partial f_2(x))$, w.z.b.w.

□

Bemerkung. 1. Die Voraussetzung aus (ii) des Satzes ist relativ schwach. Dazu überzeuge man sich, daß $\mathrm{dom} f_1 \cap \mathrm{dom} f_2 \neq \emptyset$ genau dann gilt, wenn $f_1 + f_2$ nicht identisch $+\infty$ ist. Bei $F = f_1 + f_2 \equiv +\infty$ ist $\partial F(x) = \emptyset$, $\forall x \in \mathbb{R}^n$. Deshalb ist $\mathrm{dom} f_1 \cap \mathrm{dom} f_2 \neq \emptyset$ eine Minimalforderung dafür, daß $\partial F(x) = \partial f_1(x) + \partial f_2(x)$ für alle x gelten kann.

2. Auch unter der Voraussetzung

$$\begin{cases} \exists \bar{x} \in \cap_{i=1}^m \mathrm{dom} f_i \text{ mit der Eigenschaft} \\ f_1, f_2, ..., f_{m-1} \text{ sind in } \bar{x} \text{ bzgl. } \mathbb{R}^n \text{ stetig} \end{cases} \tag{2.4.5}$$

läßt sich die Gleichheit aus (ii) zeigen (siehe z.B. Theorem 6.6.7 in P.-J. Laurent: Approximation of Optimization, Hermann, Paris 1972).

Mit Satz 2.25 der Grundvorlesung "Optimierung I" hatten wir gezeigt, daß die obere Einhüllende konvexer Funktionen wieder eine konvexe Funktion ist. Da diese Hüllenoperation eine wichtige Quelle des Entstehens von nichtdifferenzierbaren konvexen Funktionen ist, ist die Berechnung des Subdifferentials solcher Funktionen von Interesse.

Satz 2.19. *Sei $Y \subset \mathbb{R}^m$ eine kompakte Menge. $\forall y \in Y$ sei $\varphi(\cdot, y) : \mathbb{R}^n \to \bar{\mathbb{R}}$ eine konvexe, eigentliche und im folgenden Sinne unterhalbstetige Funktion: aus $x^k \to x$ folgt $\varphi(x,y) \leq \underline{\lim}_{k\to\infty} \varphi(x^k, y)$. Es sei $\forall y \in Y$ $D = dom_x \varphi(x,y)$, wobei D eine konvexe, offene Menge sei. $\forall x \in D$ sei $\varphi(x,y)$ bezüglich y auf Y oberhalbstetig (d.h. für $y^k \to y$ ist $\varphi(x,y) \geq \overline{\lim}_{k\to\infty} \varphi(x, y^k)$). Dann gilt $\forall x \in D$ für $f(x) = \max_{y \in Y} \varphi(x,y)$, daß*

$$\partial f(x) = conv(\cup_{y \in Y(x)} \partial_x \varphi(x,y)),$$

wobei

$$Y(x) = \left\{ y \in Y : \varphi(x,y) = \max_{z \in Y} \varphi(x,z) \right\}.$$

Beweis. Es sei zusätzlich vorausgesetzt, daß φ auf $D \times Y$ stetig bzgl. des vereinten Arguments $[x, y]$ ist.

1. Wir setzen $F(x) = \cup_{y \in Y(x)} \partial_x \varphi(x, y)$, $F : D \to 2^{\mathbb{R}^n}$.

   Als erstes überprüfen wir, daß F eine abgeschlossene PMA ist.

   Dazu sei $x \in D$ und $a \in \mathbb{R}^n$ gegeben und Folgen $\{x^k\} \subset D$ sowie $\{a^k\}$ mit $x = \lim_{k\to\infty} x^k$, $a^k \in F(x^k)$, $\forall k$ und $a = \lim_{k\to\infty} a^k$. Zu zeigen ist $a \in F(x)$.

   $a^k \in F(x^k)$ bedeutet, es existiert ein $y^k \in Y(x^k)$ so, daß $a^k \in \partial_x \varphi(x^k, y^k)$. Nach Definition von f und $Y(x)$ haben wir $\varphi(x^k, y^k) = f(x^k)$.

   Da $\{y^k\} \subset Y$ und Y kompakt, so können wir o.E.d.A. annehmen, daß $\exists y \in Y$ und $\exists \lim_{k\to\infty} y^k$, wobei $y = \lim_{k\to\infty} y^k$.

   Da wir φ als stetig angenommen hatten, so haben wir $\varphi(x^k, y^k) \to \varphi(x, y)$ bei $k \to \infty$.

   Nach Satz 2.25 der Grundvorlesung "Optimierung I" ist f auf D konvex. Nun ist aber D eine offene Menge, $x \in D$. Somit ist f nach Folgerung 2.8 in x stetig, d.h. $f(x^k) \to f(x)$. Da aber $f(x^k) = \varphi(x^k, y^k) \to \varphi(x, y)$, so muß $f(x) = \varphi(x, y)$ sein, d.h. $y \in Y(x)$ gelten. Schließlich gilt nach $a^k \in \partial \varphi_x(x^k, y^k)$ daß

   $$\varphi(z, y^k) - \varphi(x^k, y^k) \geq < a^k, z - x^k >, \forall z.$$

   Für fixiertes z erhalten wir für $k \to \infty$ (da φ stetig bzgl. $[x, y]$, $< \cdot, \cdot >$ - stetig), daß

   $$\varphi(z, y) - \varphi(x, y) \geq < a, z - x >, \forall z,$$

   was $a \in \partial \varphi_x(x, y)$ bedeutet. Zusammen mit $y \in Y(x)$ heißt das $a \in F(x)$.

2. Wir zeigen, daß $F \subset \partial f$ (d.h. $F(x) \subset \partial f(x), \forall x$). Es sei $x \in D$, $a \in F(x)$. Per Definition von F existiert ein $y \in Y(x)$, so daß $a \in \partial \varphi_x(x, y)$ und $f(x) = \varphi(x, y)$. Damit haben wir ($f(z) = \max_{w \in Y} \varphi(z, w) \geq \varphi(z, y)$)

   $$f(z) - f(x) \geq \varphi(z, y) - \varphi(x, y) \geq < a, z - x >, \forall z,$$

   d.h. $a \in \partial f(x)$.

3. Folgerung 2.16 (Behauptung (4) der Folgerung anwenden) ergibt $\partial f(x) = \text{conv} F(x)$.

□

Folgerung 2.20. *Seien $f_1, f_2, ..., f_m : \mathbb{R}^n \to \bar{\mathbb{R}}$ konvexe, eigentliche Funktionen, wobei $dom f_i = D$, $\forall i$ und D eine offene Menge des $\mathbb{R}^n$ sei. Sei $f(x) = \max_{1 \leq i \leq m} f_i(x)$. Dann gilt*

$$\partial f(x) = conv(\cup_{i \in I(x)} \partial f_i(x)), \forall x \in D.$$

Hierbei bezeichnet $I(x) = \{i : f_i(x) = f(x)\}$.

Bemerkung. 1. Die konvexe Hülle ist die Menge aller endlichen konvexen Linearkombinationen (Satz 2.5 der Grundvorlesung "Optimierung I"). Deshalb läßt sich die Behauptung der Folgerung auch wie folgt aufschreiben:

$$\partial f(x) = \left\{ \sum_{i \in I(x)} \lambda_i a^i : a^i \in \partial f_i(x), \lambda_i \geq 0, \forall i \in I(x), \sum_{i \in I(x)} \lambda_i = 1 \right\}.$$

2. Wir betrachten das Beispiel

$$f(x) = \max \left\{ |x|, e^{2x} - 1 \right\}.$$

Mit $f_1(x) = x$, $f_2(x) = -x$, $f_3(x) = e^{2x} - 1$ haben wir $f(x) = \max_{1 \leq i \leq 3} f_i(x)$. Da $I(0) = \{1, 2, 3\}$, so ist

$$\partial f(0) = \text{conv} \{\nabla f_1(0), \nabla f_2(0), \nabla f_3(0)\} = \text{conv} \{1, -1, 2\} = [-1, 2].$$

3. Sei $f(x) = \max_{1 \leq i \leq m} \left\{ < a^i, x > -b_i \right\}$, $x \in \mathbb{R}^n$. Nach der ersten Bemerkung zu Satz 2.9 ist $\partial f_i(x) = \left\{ a^i \right\}$. Mit $I(x) = \left\{ i :< a^i, x > -b_i = f(x) \right\}$ ergibt die Folgerung, daß

$$\partial f(x) = \text{conv} \left\{ a^i \right\}_{i \in I(x)}.$$

Wie Satz 2.26 der Grundvorlesung "Optimierung I" aussagt, ergibt die Superposition konvexer Funktionen wieder eine konvexe Funktion, wenn die äußere Funktion noch Monotonieeigenschaft hat. Deshalb steht die Frage nach einer Formel für das Subdifferential solcher zusammengesetzter Funktionen. Es geht also um die Verallgemeinerung der aus der Differentialrechnung bekannten Kettenregel.

Zur Herleitung dieser verallgemeinerten Kettenregel benötigen wir zwei Hilfsaussagen.

Hilfssatz 2.21. *Seien $C_1, C_2, ..., C_m \subset \mathbb{R}^n$ konvexe Mengen. Sei $P \subset \mathbb{R}^m_+$, P - konvex. Bei $p \in P$ sei $p = [p_1, p_2, ..., p_m]$. Dann ist*

$$C = \cup_{p \in P} (\sum_{i=1}^{m} p_i C_i)$$

eine konvexe Menge.

Beweis. Seien $x, y \in C$, $\lambda \in (0, 1)$. Dann existieren $p = [p_1, p_2, ..., p_m]$, $q = [q_1, q_2, ..., q_m] \in P$ und $a^i, b^i \in C_i$, $i = 1, 2, ..., m$, so daß

$$x = \sum_{i=1}^{m} p_i a^i, \quad y = \sum_{i=1}^{m} q_i b^i.$$

Wir setzen

$$\begin{cases} r := \lambda p + (1-\lambda)q \\ I = \{i : r_i > 0\}. \end{cases}$$

Bei $i \notin I$ ist dann $p_i = q_i = 0$ (weil $\lambda > 0, 1-\lambda > 0$ und $r_i \geq 0$, sowie $p_i, q_i \geq 0, \forall i$).

Wir haben dann

$$\begin{aligned} \lambda x + (1-\lambda)y &= \sum_{i=1}^{m} \lambda p_i a^i + \sum_{i=1}^{m} (1-\lambda) q_i b^i \\ &= \sum_{i \in I} r_i \Big(\frac{\lambda p_i}{r_i} a^i + \frac{(1-\lambda)q_i}{r_i} b^i\Big) \in \sum_{i \in I} r_i C_i = \sum_{i=1}^{m} r_i C_i \subset C, \end{aligned}$$

womit die Konvexität von C bewiesen ist.

Die hier aufgeschriebenen Inklusionen begründen sich wie folgt:

Für $\gamma_i := \frac{\lambda p_i}{r_i} = \frac{\lambda p_i}{\lambda p_i + (1-\lambda) q_i}$ haben wir bei $i \in I$, daß $0 < \gamma_i < 1$, da $\lambda \in (0,1)$, $p_i > 0$, $q_i > 0$. Nun ist aber

$$1 - \gamma_i = \frac{\lambda p_i + (1-\lambda)q_i - \lambda p_i}{\lambda p_i + (1-\lambda) q_i} = \frac{(1-\lambda)q_i}{\lambda p_i + (1-\lambda)q_i} = \frac{(1-\lambda)q_i}{r_i}.$$

Weil $a^i, b^i \in C_i$ und C_i konvex, so muß $(\frac{\lambda p_i}{r_i} a^i + \frac{(1-\lambda)q_i}{r_i} b^i) \in C_i$ gelten.

Schließlich sind $p, q \in P$. Aufgrund der Konvexität von P ist deshalb $r = (\lambda p + (1-\lambda)q) \in P$, woraus per Definition von C folgt, daß $\sum_{i=1}^m r_i C_i \subset C$. $\square$

<u>**Bemerkung.**</u> Nach Satz 2.2 der Grundvorlesung "Optimierung I" ist $\sum_{i=1}^m r_i C_i$ eine konvexe Menge (selbst wenn $p_i \geq 0$, $\forall i$ nicht gelten würde), aber, wie bekannt, die Vereinigung konvexer Mengen ergibt im allgemeinen keine konvexe Menge.

Für konvexes, differenzierbares $f : \mathbb{R} \to \mathbb{R}$ ist bekanntlich die erste Ableitung f' eine monoton nichtfallende Funktion (vgl. Satz 2.29 der Grundvorlesung "Optimierung I"). Wenn f selbst monoton nichtfallend ist, dann muß (vgl. Grundkurs Analysis) f' nichtnegativ sein (unabhängig von der Konvexität von f). Das Analogon gilt für den Subgradienten einer konvexen, monotonen Funktion:

Hilfssatz 2.22. *Sei $U \subset \mathbb{R}^m$, U - eine konvexe, offene Menge. Sei $\varphi : U \to \mathbb{R}$ eine konvexe, monoton nichtfallende Funktion (letzteres bedeutet: aus $u^1, u^2 \in U$ mit $u^1 \leq u^2$ folgt $\varphi(u^1) \leq \varphi(u^2)$). Dann gilt*

$$\partial \varphi(u) \subset \mathbb{R}^m_+, \forall u \in U.$$

Beweis. Sei $u \in U$. Sei $p \in \partial\varphi(u)$. Per Definition des Subgradienten ist dann

$$\varphi(u') - \varphi(u) \geq < p, u' - u >, \forall u' \in U \tag{2.4.6}$$

Für $\alpha > 0$, α genügend klein ist $(u - \alpha e^i) \in U$ (weil U als offen vorausgesetzt wurde). Somit haben wir einerseits aufgrund der Monotonie (da $u \geq u - \alpha e^i$) und andererseits nach (2.4.6), daß

$$0 \geq \varphi(u - \alpha e^i) - \varphi(u) \geq < p, -\alpha e^i >= -\alpha p_i,$$

woraus $p_i \geq 0$ folgt. Somit ist $p \geq 0$. □

Satz 2.23 (Kettenregel). *Es seien $g_1, g_2, ..., g_m : \mathbb{R}^n \to \bar{\mathbb{R}}$ konvexe, eigentliche Funktionen mit $D = dom(g_i)$, $i = 1, 2, ..., m$. Hier sei D eine offene Menge des $\mathbb{R}^n$. Wir bezeichnen mit $g = [g_1, g_2, ..., g_m]$ die Vektorfunktion $g : D \to \mathbb{R}^m$. Es sei $\varphi : U \to \mathbb{R}$, wobei $U \subset \mathbb{R}^m$, U - eine konvexe offene Menge des $\mathbb{R}^m$ mit der Eigenschaft $g(D) \subset U$ ist. φ sei auf U konvex und monoton nichtfallend. Dann gilt für $f(x) = \varphi(g(x))$, daß*

$$\partial f(x) = \cup_{p \in \partial\varphi(u)} (\sum_{i=1}^{m} p_i \partial g_i(x)), \forall x \in D.$$

In dieser Formel wurde $p = [p_1, p_2, ..., p_m]$, $u = g(x)$ gesetzt.

Beweis.

1. Wir setzen

$$F(x) = \cup_{p \in \partial\varphi(u)} (\sum_{i=1}^{m} p_i \partial g_i(x)),$$

   mit $u = g(x)$. Dann ist F eine PMA, $F : D \to 2^{\mathbb{R}^n}$.

2. F ist konvexwertig, denn

   - nach Satz 2.4 ist $\partial\varphi(u)$ eine konvexe Menge, welche nach Hilfssatz 2.22 im $\mathbb{R}^m_+$ liegt,
   - da die $\partial g_i(x)$ ebenso nach Satz 2.4 konvexe Mengen sind, so ist nach dem ersten Anstrich und Hilfssatz 2.21 $F(x)$ eine konvexe Menge.

3. F ist eine abgeschlossene PMA, denn

   - nach Satz 2.11 sind ∂g_i und $\partial\varphi$ abgeschlossene, lokal beschränkte PMA (da D und U offen sind, so sind die g_i und φ als konvexe Funktionen auf ihren Definitionsgebieten stetig (vgl. Folgerung 2.8)),
   - nach Folgerung 2.12 und Satz 2.4 sind alle beteiligten Subdifferentiale kompakte Mengen,
   - mit diesen Eigenschaften überprüft man die Definition der Abgeschlossenheit einer PMA bei der PMA F.

4. Wir zeigen, daß $F \subset \partial f$.

   Es sei $x \in D$ und $a \in F(x)$. Dann existiert ein $p \in \partial\varphi(u)$, $p = [p_1, p_2, ..., p_m] \geq 0$ (was nach dem ersten Anstrich von 2. gilt), so daß $a = \sum_{i=1}^{m} p_i a^i$. Hierbei sind $a^i \in \partial g_i(x)$ und $u = g(x)$.

Für $y \in D$ gilt dann

$$\begin{aligned} f(y) - f(x) &= \varphi(g(y)) - \varphi(g(x)) \\ &\geq < p, g(y) - g(x) >= \sum_{i=1}^{m} p_i (g_i(y) - g_i(x)) \\ &\geq \sum_{i=1}^{m} p_i < a^i, y - x >=< \sum_{i=1}^{m} p_i a^i, y - x >=< a, y - x > . \end{aligned}$$

Folglich ist $a \in \partial f(x)$.

Die vorkommenden Ungleichungen gelten, weil $p \in \partial\varphi(u)$, $u = g(x)$ bzw. weil $a^i \in \partial g_i(x)$, $i = 1, 2, ..., m$.

5. Die vierte Behauptung von Folgerung 2.16 beweist, daß nach 3. und 4. gelten muß $\partial f = \mathrm{conv} F$. Nach 2. ist aber $F = \mathrm{conv} F$, was den Beweis abschließt.

□

Bemerkung. 1. Wir erwähnen nochmals, daß unter den Bedingungen von Satz 2.23 die Funktion $f(x) = \varphi(g(x))$ auf D konvex ist(vgl. Satz 2.26 der Grundvorlesung "Optimierung I").

2. Wenn φ in $u = g(x)$ differenzierbar ist, so haben wir nach Satz 2.9 (weil U offen), daß $\partial\varphi(u) = \{\nabla\varphi(u)\}$. Mit $\nabla\varphi(u) = [\frac{\partial\varphi(u)}{\partial u_1}, \frac{\partial\varphi(u)}{\partial u_2}, ..., \frac{\partial\varphi(u)}{\partial u_m}]$ ergibt unser Satz, daß
$$\partial f(x) = \sum_{i=1}^{m} \frac{\partial\varphi(u)}{\partial u_i} \partial g_i(x).$$

3. Wenn zusätzlich zur Annahme der zweiten Bemerkung noch alle g_i bei x differenzierbar sind, so ist $\partial g_i(x) = \{\nabla g_i(x)\}$, $i = 1, 2, ..., m$ (weil D offen) und wir erhalten
$$\nabla f(x) = \sum_{i=1}^{m} \frac{\partial\varphi(u)}{\partial u_i} \nabla g_i(x) = \nabla\varphi(u) G(x),$$
wobei
$$G(x) = \begin{bmatrix} \text{—}\ \nabla g_1(x)\ \text{—} \\ \text{—}\ \nabla g_2(x)\ \text{—} \\ \vdots \\ \text{—}\ \nabla g_m(x)\ \text{—} \end{bmatrix},$$
d.h. die Formel des Satzes geht in die aus der Analysis-Vorlesung bekannte Kettenregel der Differentialrechnung über.

Nach Satz 2.27 der Grundvorlesung "Optimierung I" ist für die Konvexität von $f(x) = \varphi(g(x))$ die Monotonie von φ nicht erforderlich, wenn g eine affinlineare Vektorfunktion ist.

Eine abgerüstete Variante des Beweises von Satz 2.23 liefert für den erwähnten Fall einen Beweis von folgendem Satz 2.24

Satz 2.24. *Es sei $\varphi : \mathbb{R}^m \to \bar{\mathbb{R}}$ eine konvexe, eigentliche Funktion, wobei $U := dom\varphi$ eine offene Menge des $\mathbb{R}^m$ sei.*

Sei A eine Matrix vom Typ $[m, n]$, $b \in \mathbb{R}^m$. Dabei sei

$$D := \{x \in \mathbb{R}^n : (Ax + b) \in dom\varphi\} \neq \emptyset.$$

Dann gilt für $f(x) = \varphi(Ax + b)$, daß $\forall x \in D$

$$\partial f(x) = \partial\varphi(u) \cdot A = \{pA : p \in \partial\varphi(u)\}.$$

Hierbei wurde $u = Ax + b$ gesetzt.

Bemerkung.

1. Wenn φ konvex und eigentlich, dann gilt für $f(x) = \varphi(Ax+b)$ generell nur: Wenn $(Ax+b) \in \mathrm{dom}\varphi$, dann ist $\partial\varphi(u) \cdot A \subset \partial f(x)$. Wenn $\mathrm{dom}\varphi$ zusätzlich noch offen ist, dann garantiert Satz 2.24 die Gleichheit.

2. Wenn zusätzlich zu den Annahmen der ersten Bemerkung noch gilt φ ist polyedral (d.h. φ ist obere Einhüllende von endlich vielen affin-linearen Funktionen), dann ist für x mit $(Ax + b) \in \mathrm{dom}\varphi$ stets $\partial f(x) = \partial\varphi(u) \cdot A$.

Abschließende Bemerkung zu den Rechenregeln: Wir beobachten eine starke Analogie zu den Rechenregeln der Differentialrechnung. Leider folgt aus f_1, f_2 konvex im allgemeinen nicht, daß $f_1 \cdot f_2$ bzw. $\frac{f_1}{f_2}$ konvex sind. Deshalb sind für das Produkt und den Quotienten konvexer Funktionen Subdifferentialformeln i.a. nur unter relativ einschränkenden Voraussetzungen möglich.

2.5 Weitere Rechenbeispiele zur Subdifferentialrechnung

2.5.1

$$f(x) = \max\{|x|, |x - 1|\}, x \in \mathbb{R}.$$

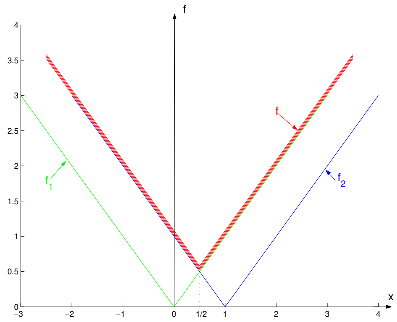
Wir setzen

$$f_1(x) = |x|, \;\; f_2(x) = |x - 1|.$$

Damit ist

$$f_1(x) = \begin{cases} x, & x \geq 0 \\ -x, & x < 0 \end{cases}, \; f_2(x) = \begin{cases} x - 1, & x \geq 1 \\ -x + 1, & x < 1 \end{cases},$$

$$\mathrm{dom} f_i = \mathbb{R}, i = 1, 2.$$



$$f(x) = \begin{cases} x, & x \geq \frac{1}{2} \\ -x+1, & x \leq \frac{1}{2} \end{cases}$$

Nach Folgerung 2.20 haben wir $\forall x \in \mathbb{R}\ \partial f(x) = \operatorname{conv}(\cup_{i \in I(x)} \partial f_i(x))$, wobei $I(x) = \{i : f(x) = f_i(x)\}$.

In Anwendung auf f_1 und f_2 ergibt dieselbe Folgerung

$$\partial f_1(x) = \begin{cases} 1, & x > 0 \\ \operatorname{conv}\{-1, 1\}, & x = 0 \\ -1, & x < 0 \end{cases}$$

bzw.

$$\partial f_2(x) = \begin{cases} 1, & x > 1 \\ \operatorname{conv}\{-1, 1\}, & x = 1 \\ -1, & x < 1 \end{cases}.$$

Da

$$I(x) = \begin{cases} \{2\}, & x < \frac{1}{2} \\ \{1, 2\}, & x = \frac{1}{2} \\ \{1\}, & x > \frac{1}{2} \end{cases},$$

so ergibt sich

$$\partial f(x) = \begin{cases} -1, & x < \frac{1}{2} \\ \operatorname{conv}\{-1, 1\}, & x = \frac{1}{2} \\ 1, & x > \frac{1}{2} \end{cases}.$$

2.5.2

$$F(x) = f(x) + \sum_{i=1}^{m} M_i \max\{0, g_i(x)\},$$

wobei f, g_i auf dom$f$ bzw. domg_i konvex und stetig sein sollen, $M_i > 0$, $\forall i$ (vgl. Abschnitt 1.3 "Exakte Strafen").

Wir setzen

$$\begin{aligned} f_0(x) &= f(x), \\ f_i(x) &= M_i \max\{0, g_i(x)\}, \quad i = 1, 2, ..., m. \end{aligned}$$

Wenn $(\text{ri dom}f) \cap \cap_{i=1}^{m} (\text{ri dom}g_i) \neq \emptyset$, dann ist nach Satz 2.18

$$\partial F(x) = \sum_{i=0}^{m} \partial f_i(x).$$

Für $i \geq 1$ ergibt sich nach Folgerung 2.20

$$\partial f_i(x) = \left\{ \begin{array}{rl} \{0\}\,, & g_i(x) < 0 \\ \text{conv}\,\{0, M_i \partial g_i(x)\}\,, & g_i(x) = 0 \\ M_i \partial g_i(x), & (g_i(x) > 0) \wedge (x \in \text{dom}g_i) \\ \emptyset, & x \notin \text{dom}g_i. \end{array} \right.$$

Für ∂f_0 haben wir

$$\partial f_0(x) = \left\{ \begin{array}{rl} \emptyset, & x \notin \text{dom}f \\ \partial f(x), & x \in \text{dom}f, \end{array} \right.$$

wobei es sich erweisen kann, daß für $x \in (\text{dom}f \setminus \text{ri dom}f)$ ebenfalls $\partial f(x) = \emptyset$ (vgl. Satz 2.4).

Wir setzen

$$\begin{aligned} I_1(x) &= \{i : g_i(x) = 0\} \\ I_2(x) &= \{i : +\infty > g_i(x) > 0\} \\ I_3(x) &= \{i : x \notin \text{dom}g_i\} \cup \{0 : x \notin \text{dom}f\}\,. \end{aligned}$$

Wir erhalten

$$\partial F(x) = \left\{ \begin{array}{rl} & \emptyset, \quad I_3(x) \neq \emptyset \\ \partial f(x) + \sum_{i \in I_2(x)} M_i \partial g_i(x) & \\ + \sum_{i \in I_1(x)} \text{conv}\,\{0, \partial g_i(x)\}\,, & I_3(x) = \emptyset. \end{array} \right.$$

2.5.3

$$f(\lambda) = \sup\,\{< c, x > + < \lambda, Ax - b >: Bx = d, x \geq 0\}$$

(vgl. Abschnitt 1.7 "Duale Dekomposition")

Hier sei A eine Matrix vom Typ $[m, n]$. Dann ist i.a. $f : \mathbb{R}^m \to \bar{\mathbb{R}}$. Um Satz 2.19 zur Anwendung bringen zu können, wollen wir annehmen, daß $\lambda \in \text{int dom}f$ und

$$X := \{x \in \mathbb{R}^n : Bx = d, x \geq 0\}$$

nichtleer und beschränkt ist.

Wir haben, daß

$$\varphi(x, \lambda) =< c, x > + < \lambda, Ax - b >$$

konvex bzgl. λ und stetig bzgl. $[x, \lambda]$ ist. Deshalb ist nach Satz 2.19

$$\partial f(\lambda) = \text{conv}(\cup_{x \in X(\lambda)} (Ax - b)), \tag{2.5.1}$$

wobei

$$X(\lambda) = \{x \in X : f(\lambda) =< c, x > + < \lambda, Ax - b >\}.$$

Wenn wir eine optimale Lösung x^* der $f(\lambda)$ definierenden Aufgabe

$$\sup \{\varphi(x, \lambda) : x \in X\}$$

berechnet haben (d.h. $x^* \in X(\lambda)$), dann ist also $(Ax^* - b)$ ein Subgradient von f an der Stelle λ.

Wir wollen zwei Bemerkungen zu Formel (2.5.1) anfügen

1. In unserem Falle vereinfacht sich Formel (2.5.1), denn $X(\lambda)$ ist eine konvexe Menge und damit ist das affin-lineare Bild $\cup_{x \in X(\lambda)}(Ax - b)$ dieser Menge ebenfalls konvex.

   Deshalb ist das "conv" in (2.5.1) überflüssig.

2. Man kann zeigen, daß Formel (2.5.1) auch für nichtbeschränktes X sowie für $\lambda \in (\mathrm{dom} f) \setminus (\mathrm{int}\ \mathrm{dom} f)$ wahr ist (vgl. z.B. Vorlesung "Dekompositionsverfahren").

2.5.4

$$f(y) = \inf \{< c, x >: Ax \leq b - By, x \geq 0\}$$

(vgl. Abschnitt 1.8 "Primale Dekomposition") Um Beispiel 2.5.3 zur Anwendung bringen zu können, nehmen wir an, daß

$$\left\{ \begin{array}{l} y \in \mathrm{int}\ \mathrm{dom} f \text{ und daß} \\ Z = \{z \in \mathbb{R}^m : zA + c \geq 0, z \geq 0\} \text{ eine beschränkte Menge ist.} \end{array} \right. \tag{2.5.2}$$

Bei y nach (2.5.2) ist nach der Dualitätstheorie der linearen Optimierung (vgl. 3.Bemerkung im Abschnitt 4.3. der Grundvorlesung "Optimierung I")

$$f(y) = \max \{< z, b - By >: z \in Z\}.$$

Das in 2.5.3. durchgerechnete Beispiel läßt sich folglich auf Beispiel 2.5.4 anwenden und wir erhalten

$$\partial f(y) = - \cup_{z \in Z(y)} zB,$$

wobei

$$Z(y) = \{z \in Z : f(y) =< z, b - By >\}.$$

Kapitel 3

Optimalitätskriterien

3.1 Die allgemeine konvexe Aufgabe

Gegeben seien

$$\begin{cases} C \subset \mathbb{R}^n, \\ f : \mathbb{R}^n \to \bar{\mathbb{R}}, \end{cases} \tag{3.1.1}$$

C - konvex, sowie f eine eigentliche konvexe Funktion.

Wir betrachten die Aufgabe

$$\begin{cases} f(x) \to \min \\ x \in C. \end{cases} \tag{3.1.2}$$

Satz 3.1. *Seien die Bedingungen (3.1.1) erfüllt. Außerdem sei ri dom$f$ $\cap$ ri$C \neq \emptyset$. Dann gilt: $x^* \in C \cap \mathrm{dom} f$ ist global-optimale Lösung von (3.1.2) genau dann, wenn existiert $a \in \partial f(x^*)$, so daß $< a, x - x^* > \geq 0$, $\forall x \in C$.*

Beweis." $\Rightarrow$ " Es sei $x^* \in C \cap \mathrm{dom} f$ global-optimale Lösung von (3.1.2), d.h.

$$f(x) \geq f(x^*), \forall x \in C. \tag{3.1.3}$$

Wir setzen

$$\begin{aligned} f_1(x) &= f(x) - f(x^*), & C_1 &= \mathrm{dom} f, \\ f_2(x) &= 0, & C_2 &= C. \end{aligned}$$

Nach (3.1.3) gilt dann $f_1(x) \geq f_2(x)$, $\forall x \in C_1 \cap C_2$. Außerdem ist laut Voraussetzung ri$C_1 \cap$ ri$C_2 \neq \emptyset$, f_1 auf C_1 konvex, f_2 auf C_2 konkav. Nach Hilfssatz 2.3 existiert deshalb eine affin-lineare Funktion $l(x) =< a, x > +\alpha$, so daß

$$\begin{cases} f_1(x) \geq l(x), \forall x \in C_1 \\ f_2(x) \leq l(x), \forall x \in C_2. \end{cases} \tag{3.1.4}$$

Da $x^* \in C \cap \mathrm{dom} f = C_1 \cap C_2$ und $f_1(x^*) = f_2(x^*) = 0$, so ergibt (3.1.4), daß $l(x^*) = 0$, d.h. $\alpha = - < a, x^* >$ bzw. $l(x) =< a, x - x^* >$.

Die erste Ungleichung von (3.1.4) lautet also

$$f(x) - f(x^*) \geq < a, x - x^* >, \forall x \in \mathrm{dom} f,$$

was $a \in \partial f(x^*)$ bedeutet.

Die zweite Ungleichung von (3.1.4) lautet

$$< a, x - x^* > \geq 0, \forall x \in C.$$

" $\Leftarrow$ " Sei $a \in \partial f(x^*)$ und $< a, x - x^* > \geq 0, \forall x \in C$.

Per Definition des Subgradienten ist damit

$$f(x) - f(x^*) \geq < a, x - x^* >, \forall x \in \mathbb{R}^n.$$

Zusammen mit der Voraussetzung $< a, x - x^* > \geq 0$ bei $x \in C$ heißt das also, daß

$$f(x) - f(x^*) \geq 0, \forall x \in C,$$

was globale Optimalität von x^* in (3.1.2) bedeutet.

□

Bemerkung. 1. Die Aussage des Satzes verallgemeinert das aus Theorem 3.1 der Grundvorlesung "Optimierung I" für (3.1.2) mit differenzierbarem (nicht notwendiger Weise konvexem) f folgende notwendige Optimalitätskriterium $< \nabla f(x^*), x - x^* > \geq 0$, $\forall x \in C$.

2. Im Falle $C = \mathbb{R}^n$ und ri dom$f \neq \emptyset$ (was bei konvexem f genau dann gilt, wenn dom$f \neq \emptyset$, vgl. Satz 2.10 der Grundvorlesung "Optimierung I" und letzte Bemerkung im Abschnitt 1.9) bedeutet das Kriterium des Satzes: $0 \in \partial f(x^*)$.

   Wie in diesem Fall läßt sich das Kriterium des Satzes in vielen weiteren speziellen Fällen für C (z.B. $C = \mathbb{R}^n_+$, $C = \{x : a \leq x \leq b\}$, $C = \{x : Ax = b\}$) durch eine endliche Anzahl von Relationen ersetzen. Diese können dann benutzt werden, um x^* zu berechnen (das allgemeine Kriterium des Satzes beinhaltet eine unendliche Anzahl von zu erfüllenden Ungleichungen).

   Vergleichen Sie hierzu auch Folgerungen 3.2 bis 3.4 aus der Grundvorlesung "Optimierung I".

3. Aus der Behauptung des Satzes folgt, daß im optimalen Punkt x^* gilt $f'(x^*; x - x^*) \geq 0$, $\forall x \in C$.

   Beweis. Nach Satz 2.6 ist bei $x^* \in \mathrm{dom} f$, $r \in \mathrm{Lin}\ \mathrm{dom} f$, $b \in \partial f(x^*)$ stets $f'(x^*, r) \geq < b, r >$.

   Somit ist $f'(x^*, r) \geq < a, r >$, da $a \in \partial f(x^*)$.

   Sei $x \in C$. Wir setzen $r := x - x^*$.

Wenn $r \in \text{Lin dom}f$, dann gilt also $f'(x^*, r) \geq < a, r >$, woraus aufgrund des Satzes (es ist $< a, x - x^* > \geq 0$) folgt $f'(x^*, r) \geq 0$.

Wenn $r \notin \text{Lin dom}f$, dann ist $f'(x^*, r) = +\infty$ und damit gilt $f'(x^*, r) \geq 0$ auch in diesem Fall. □

4. Für $x^* \in C \cap \text{ri dom}f$ gilt auch die Umkehrung der Aussage der dritten Bemerkung. Genauer: Wenn für ein $\bar{x} \in C \cap \text{ri dom}f$ gilt, daß

$$f'(\bar{x}, x - \bar{x}) \geq 0, \forall x \in C,$$

dann ist $\bar{x}$ global-optimale Lösung von (3.1.2) und damit existiert nach Satz 3.1 ein $a \in \partial f(\bar{x})$, so daß $< a, x - \bar{x} > \geq 0$, $\forall x \in C$.

Beweis. Sei $x \in C \cap \text{dom}f$. Dann ist $(x - \bar{x}) \in \text{Lin dom}f$.

Aus $f'(\bar{x}; x - \bar{x}) \geq 0$ folgt nach Satz 2.6, daß

$$\max \{< a, x - \bar{x} >: a \in \partial f(\bar{x})\} \geq 0.$$

Somit existiert ein $a = a(x) \in \partial f(\bar{x})$, so daß für dieses x $< a, x - \bar{x} > \geq 0$, woraus per Definition des Subgradienten folgt, daß $f(x) - f(\bar{x}) \geq 0$.

Wenn $x \in C$, aber $x \notin \text{dom}f$, dann ist $f(x) = +\infty$, weshalb auch in diesem Fall $f(x) - f(\bar{x}) \geq 0$.

Deshalb ist $f(x) \geq f(\bar{x})$, $\forall x \in C$. □

3.2 Aufgaben mit einer konvexen Restriktion

Seien $f, g : \mathbb{R}^n \to \bar{\mathbb{R}}$ konvexe, eigentliche Funktionen. Wir betrachten die Aufgabe

$$f^* = \inf \{f(x) : g(x) \leq 0\} . \tag{3.2.1}$$

Wir wollen (3.2.1) unter folgenden Annahmen untersuchen:

(V1) (Slaterpunkt): Es existiert ein $\bar{x} \in \mathbb{R}^n$, so daß $g(\bar{x}) < 0$.

(V2) (Lösbarkeit): Es existiert ein $x^* \in (\text{dom}f \cap \text{dom}g)$, so daß $g(x^*) \leq 0$, $f(x^*) = f^*$.

(V3) $\text{dom}f \cap \text{dom}g$ ist eine offene Menge.

Bemerkung. 1. Unter Voraussetzung (V2) ist die Funktion

$$F(x) := \max \{f(x) - f^*, g(x)\}$$

eine konvexe, eigentliche Funktion, $F : \mathbb{R}^n \to \bar{\mathbb{R}}$.

2. Außerdem ist $F(x) \geq 0, \forall x \in \mathbb{R}^n$, denn wenn $g(x) \leq 0$, dann ist x zulässig in (3.2.1), woraus $f(x) \geq f^*$, d.h. $f(x) - f^* \geq 0$ folgt. Dann ist aber $F(x) \geq 0$. Wenn $g(x) > 0$, dann ist $F(x) \geq g(x) > 0$.

   Neben (3.2.1) wollen wir noch die (hypothetische, da wir f^* nicht kennen) Aufgabe
   $$F^* = \inf \{F(x) : x \in \mathbb{R}^n\}. \tag{3.2.2}$$
   betrachten.

3. Unter (V2) ist $F^* = 0$, denn $\forall x \in \mathbb{R}^n$ ist $F(x) \geq 0 = F(x^*)$.

4. (3.2.1) und (3.2.2) haben die gleichen optimalen Lösungen.

   Beweis.

   "$\Rightarrow$" Wenn x^* optimale Lösung von (3.2.1) ist, dann ist $f(x^*) = f^*$, $g(x^*) \leq 0$. Folglich ist $F(x^*) = \max\{f(x^*) - f^*, g(x^*)\} = 0$, d.h. $F(x^*) = F^*$.

   "$\Leftarrow$" Sei $\bar{x}$ optimale Lösung von (3.2.2), d.h. $F(\bar{x}) = 0$. Dann ist $0 = \max\{f(\bar{x}) - f^*, g(\bar{x})\}$. Damit ist
   $$\begin{cases} g(\bar{x}) \leq 0 \\ f(\bar{x}) - f^* \leq 0. \end{cases}$$
   Aus der ersten Zeile folgt aber, daß $f(\bar{x}) \geq f^*$. Zusammen mit der zweiten Zeile bedeutet das $f(\bar{x}) = f^*$. Folglich ist $\bar{x}$ optimale Lösung von (3.2.1).

   □

Wir führen die Punkt-Menge-Abbildung (PMA)
$$G(x) = \begin{cases} \partial f(x), & \text{wenn } g(x) < 0 \\ \operatorname{conv}\{\partial f(x), \partial g(x)\}, & \text{wenn } g(x) = 0 \\ \partial g(x), & \text{wenn } g(x) > 0 \end{cases}$$
ein.

Satz 3.2. *Es seien (V1)-(V3) erfüllt. Dann gilt: x^* ist optimal in (3.2.2) (und damit zulässig sowie optimal in (3.2.1)) genau dann, wenn $0 \in G(x^*)$.*

Beweis. 1. Nach (V2) ist $\operatorname{dom} F \neq \emptyset$. Damit ist auch $\operatorname{ri}\operatorname{dom} F \neq \emptyset$(vgl. Satz 2.10 der Grundvorlesung "Optimierung I"). Somit erfüllt (3.2.2) die Voraussetzung von Satz 3.1. Deshalb ist (vgl. 2.Bemerkung zu Satz 3.1) $x^* \in \mathbb{R}^n$ optimal in (3.2.2) genau dann, wenn $0 \in \partial F(x^*)$. Wenn wir zeigen, daß für ein $x^* \in \mathbb{R}^n$ $0 \in \partial F(x^*)$ genau dann, wenn $0 \in G(x^*)$ gilt, dann haben wir somit Satz 3.2 bewiesen. Das weisen wir in 3.-4. nach.

2. Aufgrund von (V3) können wir aus Folgerung 2.20 ableiten, daß $\forall x \in (\mathrm{dom} f \cap \mathrm{dom} g)$
$$\partial F(x) = \mathrm{conv}(\cup_{i \in I(x)}) \partial f_i(x),$$
wobei $f_1(x) = f(x) - f^*$, $f_2(x) = g(x)$, $I(x) = \{i \in \{1,2\} : f_i(x) = F(x)\}$ gesetzt wurde.

   Für $x \notin (\mathrm{dom} f \cap \mathrm{dom} g)$ ist $\partial F(x) = \emptyset$, da $F(x) = +\infty$ für solche x.

   Unter Beachtung, daß $\forall x \in \mathbb{R}^n$ stets $F(x) \geq 0$ folgt, erhalten wir folgende Möglichkeiten für $I(x)$
$$I(x) = \begin{cases} \{1\}, & \text{wenn } F(x) = f_1(x) > f_2(x) \\ \{1,2\}, & \text{wenn } F(x) = f_1(x) = f_2(x) \geq 0 \\ \{2\}, & \text{wenn } 0 \leq F(x) = f_2(x) > f_1(x). \end{cases}$$
   Somit haben wir
$$\partial F(x) = \begin{cases} \partial f(x), & \text{wenn } f(x) - f^* > g(x) \\ \mathrm{conv}\{\partial f(x), \partial g(x)\}, & \text{wenn } f(x) - f^* = g(x) \geq 0 \\ \partial g(x), & \text{wenn } 0 \leq g(x) > f(x) - f^*. \end{cases}$$

3. " $\Rightarrow$ " Sei $0 \in \partial F(x^*)$. Dann ist nach Teil 1. x^* optimale Lösung von (3.2.2) und damit (vgl. 4.Bemerkung von oben) auch optimale Lösung von (3.2.1). Das bedeutet:
$$g(x^*) \leq 0, \ f(x^*) - f^* = 0.$$
   (a) Wenn $g(x^*) < 0$, dann haben wir $0 = f(x^*) - f^* > g(x^*)$, was nach Teil (2) bedeutet $\partial F(x^*) = \partial f(x^*)$ und damit ist (wir haben $g(x^*) < 0$) $\partial F(x^*) = G(x^*)$. Somit folgt aus $0 \in \partial F(x^*)$ die Behauptung $0 \in G(x^*)$.

   (b) Wenn $g(x^*) = 0$, dann liegt $0 = f(x^*) - f^* = g(x^*)$ vor, was nach Teil (2) bedeutet $\partial F(x^*) = \mathrm{conv}\{\partial f(x^*), \partial g(x^*)\}$. Nach Definition von G ist aber (wir haben $g(x^*) = 0$) $G(x^*) = \mathrm{conv}\{\partial f(x^*), \partial g(x^*)\}$. Aus $0 \in \partial F(x^*)$ folgt also wiederum die Behauptung $0 \in G(x^*)$.

4. " $\Leftarrow$ " Es sei $0 \in G(x^*)$.

   (a) Wäre $g(x^*) > 0$, so wäre nach Definition von G $0 \in \partial g(x^*)$. Nach Satz 3.1 bedeutet das aber, x^* ist optimale Lösung der Aufgabe $\inf\{g(x) : x \in \mathbb{R}^n\}$, d.h. $0 < g(x^*) = \inf\{g(x) : x \in \mathbb{R}^n\}$. Das widerspricht (V1). Folglich muß $g(x^*) \leq 0$ gelten.

   (b) Wenn $g(x^*) < 0$, dann bedeutet $0 \in G(x^*)$ nach Definition von G, daß $0 \in \partial f(x^*)$.

   Da bei $g(x^*) < 0$ nach der Formel für ∂F aus Teil (2) für $x = x^*$ nur die erste Zeile dieser Formel gelten kann, so ist $\partial F(x^*) = \partial f(x^*)$, womit aus $0 \in \partial f(x^*)$ folgt, daß $0 \in \partial F(x^*)$.

   Nach Teil (1) bedeutet das aber: x^* ist global-optimale Lösung von (3.2.2).

(c) Wenn $g(x^*) = 0$, dann bedeutet $0 \in G(x^*)$ nach Definition von G, daß $0 \in \operatorname{conv}\{\partial f(x^*), \partial g(x^*)\}$.
Somit existieren $a \in \partial f(x^*)$, $b \in \partial g(x^*)$ und ein $\lambda \in [0,1]$, so daß $0 = \lambda a + (1-\lambda)b$. Hier ist der Fall $\lambda = 0$ nicht möglich, denn bei $\lambda = 0$ wäre $0 = b \in \partial g(x^*)$, was analog wie im Fall (4.a) der Bedingung (V1) widerspricht (wir sind im Falle $g(x^*) = 0$).
Somit haben wir $\lambda > 0$. Außerdem gilt nach Definition des Subgradienten

$$f(y) - f(x^*) \geq < a, y - x^* >, \forall y \in \mathbb{R}^n,$$
$$g(y) - g(x^*) \geq < b, y - x^* >, \forall y \in \mathbb{R}^n.$$

Nach Multiplikation mit $\lambda > 0$ bzw. $(1-\lambda) \geq 0$ und nachfolgender Addition der Ungleichungen erhalten wir unter Berücksichtigung von $g(x^*) = 0$, $0 = \lambda a + (1-\lambda)b$, daß

$$\lambda f(y) - \lambda f(x^*) + (1-\lambda) g(y) \geq 0, \forall y \in \mathbb{R}^n$$

Somit ist (es ist $\lambda > 0$)

$$f(y) - f(x^*) \geq -\frac{1-\lambda}{\lambda} g(y), \forall y \in \mathbb{R}^n. \tag{3.2.3}$$

Wenn wir hier als y eine optimale Lösung x^{**} von (3.2.1) wählen (sie existiert nach (V2)), so ist $g(x^{**}) \leq 0$, d.h. $-\frac{1-\lambda}{\lambda} g(x^{**}) \geq 0$, und $f(x^{**}) = f^*$.
Für dieses y bedeutet (3.2.3) aber

$$f^* - f(x^*) \geq 0.$$

Aufgrund der Zulässigkeit von x^* (wir sind im Falle $g(x^*) = 0$) in (3.2.1) ist in der letzten Ungleichung eine echte Ungleichheit nicht möglich. Es gilt also $f(x^*) = f^*$, was zusammen mit $g(x^*) = 0$ bedeutet, daß x^* optimale Lösung von (3.2.1) und damit auch von (3.2.2) ist. Nach 1. heißt das aber $0 \in \partial F(x^*)$.

□

3.3 Aufgaben mit endlich vielen konvexen Restriktionen

Es seien $f, g_i : \mathbb{R}^n \to \bar{\mathbb{R}}$, $i = 1, 2, ..., m$, wobei alle diese Funktionen im $\mathbb{R}^n$ konvex und eigentlich seien.

Wir betrachten die Aufgabe

$$f^* = \inf \{f(x) : g_i(x) \leq 0, i = 1, 2, ..., m\}, \tag{3.3.1}$$

unter den Annahmen:

$(\overline{V1})$ (Slaterpunkt): Es existiert ein $\bar{x} \in \mathbb{R}^n$ mit $g_i(\bar{x}) < 0,\ i = 1, 2, ..., m$.

$(\overline{V2})$ (Lösbarkeit): Es existiert ein $x^* \in (\mathrm{dom} f \cap (\cap_{i=1}^m \mathrm{dom} g_i))$, so daß $g_i(x^*) \leq 0,\ i = 1, 2, ..., m,\ f(x^*) = f^*$.

$(\overline{V3})$ $\mathrm{dom} f \cap (\cap_{i=1}^m \mathrm{dom} g_i)$ ist eine offene Menge.

Bemerkung.

1. Die Funktion $g(x) = \max_{1 \leq i \leq m} g_i(x)$ ist offensichtlich eine konvexe, eigentliche Funktion (letzteres aufgrund $(\overline{V2})$, ersteres nach Satz 2.25 der Grundvorlesung "Optimierung I").

2. Wir vermerken, daß $g(x)$ i.a. eine nichtdifferenzierbare Funktion ist (selbst wenn alle g_i differenzierbar sind).

3. Offensichtlich ist
$$\{x \in \mathbb{R}^n : g_i(x) \leq 0, i = 1, 2, ..., m\} = \{x \in \mathbb{R}^n : g(x) \leq 0\}.$$
Deshalb ist Aufgabe (3.3.1) äquivalent mit der Aufgabe (3.2.1)
$$f^* = \inf \{f(x) : g(x) \leq 0\}.$$

4. Wir vermerken, daß aus der Gültigkeit von $(\overline{V1})$-$(\overline{V3})$ das Erfülltsein der Annahmen (V1)-(V3) aus Abschnitt 3.2 für (3.2.1) folgt.

Im weiteren sei $I(x) = \{i \in \{1, 2, ..., m\} : g_i(x) = 0\}$.

Satz 3.3. *Es seien* $(\overline{V1}) - (\overline{V3})$ *erfüllt. Sei* $g_i(x)$ *in* x^* *stetig für* $\forall i = 1, 2, ..., m$. *Dann gilt: Die zulässige Lösung* x^* *von Aufgabe (3.3.1) ist optimal genau dann, wenn existieren* $y_i^* \geq 0\ \forall i \in I(x^*)$ *und* $a^0 \in \partial f(x^*)$ *sowie* $a^i \in \partial g_i(x^*)\ \forall i \in I(x^*)$, *so daß*
$$0 = a^0 + \sum_{i \in I(x^*)} y_i^* a^i.$$

Beweis." $\Leftarrow$ " Es sei gegeben
$$\begin{cases} x^* \text{ ist zulässige Lösung von (3.3.1)} \\ \text{Es existieren } y_i^* \geq 0, \forall i \in I(x^*), \text{ sowie} \\ a^i \in \partial g_i(x^*), \forall i \in I(x^*) \text{ und ein } a^0 \in \partial f(x^*), \text{ so daß} \\ 0 = a^0 + \sum_{i \in I(x^*)} y_i^* a^i. \end{cases} \tag{3.3.2}$$

Per Definition der eingehenden Größen gilt dann
$$f(y) - f(x^*) \geq < a^0, y - x^* >, \forall y \in \mathbb{R}^n$$
$$g_i(y) - g_i(x^*) \geq < a^i, y - x^* >, \forall y \in \mathbb{R}^n, \forall i \in I(x^*).$$

Nach Multiplikation der zweiten Zeile mit $y_i^* \geq 0$ und Addition aller Ungleichungen erhalten wir $\forall y \in \mathbb{R}^n$, daß
$$f(y) - f(x^*) + \sum_{i \in I(x^*)} y_i^* (g_i(y) - g_i(x^*)) \geq < a^0 + \sum_{i \in I(x^*)} y_i^* a^i, y - x^* > .$$

Nach der letzten Zeile von (3.3.2) und da $g_i(x^*) = 0$ bei $i \in I(x^*)$ gilt damit

$$f(y) - f(x^*) \geq - \sum_{i \in I(x^*)} y_i^* g_i(y), \forall y \in \mathbb{R}^n.$$

Wenn y zulässig in (3.3.1) ist, haben wir $g_i(y) \leq 0, \forall i$, woraus sich ergibt, daß $f(y) - f(x^*) \geq 0$ bei beliebigem zulässigen y. Folglich ist x^* optimal in (3.3.1).

" $\Rightarrow$ " Es sei x^* optimale Lösung von (3.3.1). Nach der vierten Bemerkung vor Satz 3.3 und Satz 3.2 haben wir $0 \in G(x^*)$. Wie in früheren Beweisen schon mehrfach gezeigt, ist dann $g(x^*) > 0$ nicht möglich (Widerspruch zu $(\overline{V1})$). Wir unterscheiden die Fälle $g(x^*) < 0$ und $g(x^*) = 0$.

Fall 1: $g(x^*) < 0$. Dann ist $g_i(x^*) < 0$, $\forall i$. Aufgrund der Stetigkeit der g_i in x^* bedeutet das, daß x^* innerer Punkt des zulässigen Bereichs von (3.3.1) ist. Dann ist x^* aber auch optimale Lösung der Aufgabe $\min\{f(x) : x \in \mathbb{R}^n\}$. Aufgrund der zweiten Bemerkung zu Satz 3.1 heißt das aber $0 \in \partial f(x^*)$. Somit existiert ein $a^0 \in \partial f(x^*)$ mit $a^0 = 0$. Da im Falle 1 $I(x^*) = \emptyset$, so gilt die Behauptung des Satzes im Falle 1.

Fall 2: $g(x^*) = 0$. Die Eigenschaft $0 \in G(x^*)$ bedeutet dann

$$0 \in \operatorname{conv}\{\partial f(x^*), \partial g(x^*)\}.$$

Folglich existieren ein $\lambda \in [0,1]$ sowie $a^0 \in \partial f(x^*)$, $b \in \partial g(x^*)$, so daß $0 = \lambda a^0 + (1-\lambda)b$. Wäre hier $\lambda = 0$, so wäre $b = 0$. Das ist aber nicht möglich, denn wäre $0 \in \partial g(x^*)$, was (wie oben mehrfach gezeigt) bedeutet, daß x^* optimale Lösung von $\min\{g(x) : x \in \mathbb{R}^n\}$ wäre. Da aber im Falle 2 $g(x^*) = 0$, so steht das im Widerspruch zu $(\overline{V1})$.

Per Definition ist $g(x) = \max\{g_i(x) : 1 \leq i \leq m\}$. Deshalb bedeutet $b \in \partial g(x^*)$ nach Folgerung 2.20 (wähle $D = U_\varepsilon(x^*)$, da die g_i in x^* stetig, so existiert ein $\varepsilon > 0$ mit $U_\varepsilon(x^*) \subset \operatorname{dom} g_i$, $\forall i$), daß $b \in \operatorname{conv}(\cup_{i \in I(x^*)} \partial g_i(x^*))$. Somit existieren für $i \in I(x^*)$ Zahlen $\lambda_i \geq 0$ mit $\sum_{i \in I(x^*)} \lambda_i = 1$ und $a^i \in \partial g_i(x^*)$, $i \in I(x^*)$, so daß $b = \sum_{i \in I(x^*)} \lambda_i a^i$.

Deshalb haben wir

$$0 = \lambda a + (1-\lambda)b = \lambda a + (1-\lambda) \sum_{i \in I(x^*)} \lambda_i a^i,$$

wobei $\lambda_i \geq 0$, $\forall i$, $0 < \lambda \leq 1$.

Das ergibt

$$0 = a + \frac{1-\lambda}{\lambda} \sum_{i \in I(x^*)} \lambda_i a^i.$$

Mit $y_i^* = \frac{1-\lambda}{\lambda}\lambda_i$ für $i \in I(x^*)$ haben wir $y_i^* \geq 0$, $\forall i$ und es gilt die Behauptung des Satzes auch im Falle 2.

$\square$

Kapitel 4

Lösungsverfahren für die Aufgabe des freien Optimums

Wir betrachten in diesem Abschnitt die Aufgabe

$$\left\{\begin{array}{l} f(x) \to \min \\ x \in \mathbb{R}^n \end{array}\right. \tag{4.0.1}$$

für den Fall $f : \mathbb{R}^n \to \mathbb{R}$, f im $\mathbb{R}^n$ konvex. Damit ist $\mathrm{dom} f = \mathbb{R}^n$, aber i.a. wird f nicht in jedem Punkt des $\mathbb{R}^n$ differenzierbar sein.

Wir vermeiden vorerst bewußt den Fall $f : \mathbb{R}^n \to \bar{\mathbb{R}}$, weil dann Aufgaben mit Nebenbedingungen ebenfalls vom Typ (4.0.1) wären.

Im weiteren bezeichne

$$f^* = \inf\{f(x) : x \in \mathbb{R}^n\}$$

den Optimalwert von Aufgabe (4.0.1) (Vermerke, daß $f^* = -\infty$ möglich ist),

$$X^* = \{x \in \mathbb{R}^n : f(x) = f^*\}$$

die Menge der (global-)optimalen Lösungen von (4.0.1) (Vermerke, daß $X^* = \emptyset$ möglich ist).

4.1 Einige nützliche Beobachtungen (Motivation)

4.1.1 Rückführung auf glatte Aufgaben

Man kann sich von der eventuellen Nichtdifferenzierbarkeit einer konvexen Funktion befreien und von f in (4.0.1) zu einer geglätteten Funktion F übergehen, deren freies Minimum im $\mathbb{R}^n$ mit der Lösung von (4.0.1) übereinstimmt.

Solch ein Übergang erlaubt, das Minimum von $F(x)$ mit den Methoden der differenzierbaren Optimierung zu berechnen. Ein Beispiel solch einer geglätteten Funktion ist die sogenannte Moreau-Yosida-Regularisierte

$$F(x) = \min\left\{f(y) + \frac{1}{\gamma}\|y-x\|^2 : y \in \mathbb{R}^n\right\}. \tag{4.1.1}$$

Hier sei $\gamma > 0$ eine Konstante.

Wenn $f : \mathbb{R}^n \to \bar{\mathbb{R}}$, f im $\mathbb{R}^n$ konvex und eigentlich ist sowie für jedes $\alpha \in \mathbb{R}$ die Niveaumenge $\{x \in \mathbb{R}^n : f(x) \leq \alpha\}$ abgeschlossen ist, dann ist $F(x)$ eine im $\mathbb{R}^n$ stetig differenzierbare konvexe Funktion und jede optimale Lösung von

$$\left\{\begin{array}{l} F(x) \to \min \\ x \in \mathbb{R}^n \end{array}\right. \tag{4.1.2}$$

ist auch optimale Lösung von (4.0.1) und umgekehrt.

Leider erfordert aber jede Funktionswert- und Gradientenberechnung von F das Lösen der F definierenden freien Optimierungsaufgabe (4.1.1) mit einer i.a. nichtdifferenzierbaren Zielfunktion. Wir wollen auf diesen Weg hier nicht näher eingehen (siehe z.B. Fukushima, M. and Qi, L.(1996): A globally and superlinearly convergent algorithm for nonsmooth convex minimization. SIAM Journal on Optimization, vol.6, pp. 1106-1120).

4.1.2 Ein Antisubgradient muß keine Abstiegsrichtung sein

Für ein differenzierbares f ist $(-\nabla f(x))$ die Richtung des lokal stärksten Abstiegs von f im Punkte x und für jede positiv definite Matrix $D \in \mathbb{R}^{n,n}$ ist im Falle $\nabla f(x) \neq 0$ $(-D\nabla f(x))$ eine Abstiegsrichtung von f in x (vgl. Grundkurs "Optimierung II").

Da wir in den Abschnitten 2. und 3. gesehen hatten, daß $\partial f(x)$ bei konvexen Funktionen das Analogon zu $\nabla f(x)$ bei differenzierbarem f ist, ist es naheliegend, Analoga zu den Verfahren vom Gradiententyp für differenzierbare Aufgaben, die im Grunde nach der Iterationsvorschrift

$$x^{k+1} = x^k - \alpha_k D_k \nabla f(x^k)$$

verlaufen, auch auf unsere Aufgabe (4.0.1) anzuwenden, indem man anstelle von $\nabla f(x^k)$ einen geeigneten Subgradienten von f benutzt.

Wir werden uns in den späteren Abschnitten davon überzeugen, daß dieses Vorgehen im Prinzip möglich ist (vgl. Subgradientenverfahren aus Abschnitt 4.2). Hier wollen wir uns davon überzeugen, daß auf diesem Weg einige Schwierigkeiten auftreten.

Ein Blick auf die Formeln für das Subdifferential aus Abschnitt 2.4 (vgl. Satz 2.19, Folgerung 2.20, Satz 2.23, Beispiele aus Abschnitt 2.5) legt es nahe davon auszugehen, daß wir in einem Iterationspunkt x^k nicht das gesamte Subdifferential von f, sondern nur einen Subgradienten kennen werden.

In den später zu beschreibenden Verfahren gehen wir davon aus, daß wir zur Funktion f über ein sogenanntes "Orakel" verfügen. Das Orakel zu f ist eine black-box (in der Software-Realisierung eines Verfahrens ein vom Aufgabensteller zu lieferndes Unterprogramm), das zu jedem $x \in \mathbb{R}^n$ am Eingang der black-box am Ausgang die Zahl $f(x)$ und einen Vektor $g(x)$ mit $g(x) \in \partial f(x)$ liefert.

Wir wollen uns an einem Beispiel davon überzeugen, daß i.a. das vom Orakel gelieferte $g(x)$ nicht mit einer Abstiegsrichtung von f im Punkte x verbunden sein muß, was im Gegensatz zum differenzierbaren Fall steht, denn im letzten Fall gilt für $\alpha > 0$, α - genügend klein, daß

$$f(x - \alpha \nabla f(x)) < f(x),$$

wenn x keine optimale Lösung von (4.0.1) ist.

<u>Beispiel 1</u>:

$$f(x) = |x_1 - x_2| + 0.1 \cdot |x_1 + x_2 - 4|$$

mit $x = [x_1, x_2] \in \mathbb{R}^2$.

Wir haben dom$f = \mathbb{R}^2$, f - konvex im $\mathbb{R}^2$ (vgl. Satz 2.25 der Grundvorlesung "Optimierung I"), weil

$$f(x) = \max\{x_1 - x_2, -x_1 + x_2\} + 0.1 \cdot \max\{x_1 + x_2 - 4, -x_1 - x_2 + 4\}.$$

Offensichtlich gilt

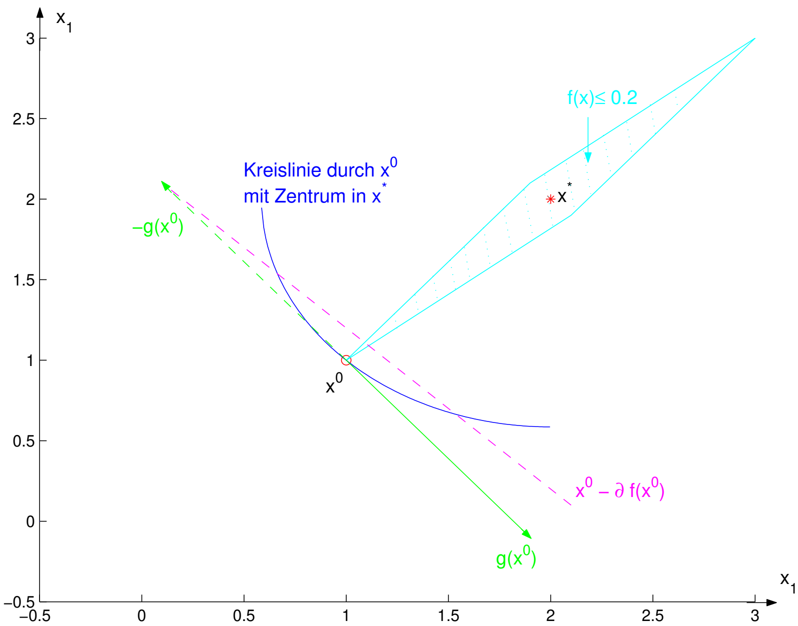
$$f^* = 0, \;\; X^* = \{[2, 2]\}.$$

Es sei $x^0 = [1, 1]$. Unter Benutzung von Satz 2.18 und Folgerung 2.20 erhalten wir

$$\partial f(x^0) = \text{conv}\left\{ \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\} + 0.1 \cdot \begin{bmatrix} -1 \\ -1 \end{bmatrix}.$$

Wir nehmen an, das Orakel zu f habe uns für $x^0 = [1, 1]$ den Funktionswert $f(x^0) = 0.2$ und einen Subgradienten $g(x^0) = \begin{bmatrix} 1 \\ -1 \end{bmatrix} + 0.1 \cdot \begin{bmatrix} -1 \\ -1 \end{bmatrix} = \begin{bmatrix} 0.9 \\ -1.1 \end{bmatrix}$ geliefert.

In der nachfolgenden Skizze sei die Richtung $g(x^0)$ im Punkt x^0 angetragen.



Das schraffierte Parallelogramm beschreibt die Menge $\{x \in \mathbb{R}^n : f(x) \leq f(x^0)\}$. Nur Richtungen d, die in x^0 angelegt, ins Innere des Parallelogramms zeigen, sind Abstiegsrichtung von f in x^0 (d.h. $f(x^0 + \alpha d) < f(x^0)$ bei $\alpha > 0$, α - genügend klein). Wir erkennen, daß $d = -g(x^0)$ keine Absteigsrichtung von f in x^0 ist.

Anhand der in der Skizze angedeuteten Kreislinie vom Radius $\|x^0 - x^*\|$ mit Zentrum im Optimalpunkt x^* erkennen wir, daß aber die Punkte $x^0 - \alpha g(x^0)$ für $\alpha > 0$, α - genügend klein, näher an x^* liegen als x^0 selbst. Wie Hilfssatz 4.1 zeigen wird, ist das eine allgemeine Gesetzmäßigkeit. Die Summe $x^0 - \partial f(x^0)$ ist der Abschnitt, der die Punkte $\begin{bmatrix} 2.1 \\ 0.1 \end{bmatrix}$ und $\begin{bmatrix} 0.1 \\ 2.1 \end{bmatrix}$ miteinander verbindet. In der Skizze ist die Summe als gestrichelte Linie eingetragen. Man erkennt, daß gewisse in x^0 angelegte Abstiegsrichtungen d von f in x^0 in die Menge $(x^0 - \partial f(x^0))$ zeigen. Auch das ist nicht zufällig, wie Hilfssatz 4.2 zeigt.

4.1.3 Eigenschaften der Antisubgradienten

Hilfssatz 4.1. *Gegeben sei Aufgabe (4.0.1) mit konvexem f, dom$f = \mathbb{R}^n$. Es sei $f(x^0) > f^*$. Es sei $X^* = \{x \in \mathbb{R}^n : f(x) = f^*\} \neq \emptyset$ und abgeschlossen. Für beliebiges $g(x^0) \in \partial f(x^0)$ gilt dann*

$$\rho(x^0 - \lambda g(x^0), X^*) < \rho(x^0, X^*),$$

wenn nur $\lambda > 0$, λ - genügend klein.

Hierbei bezeichnet ρ den Euklidischen Abstand im $\mathbb{R}^n$, d.h.

$$\rho(A, B) = \inf \{\|x - y\| : x \in A, y \in B\}.$$

Beweis. Sei $x^* \in X^*$. Dann ist $f^* = f(x^*) < f(x^0)$ und ($g(x^0)$ ist Subgradient) $f(x^*) \geq f(x^0) + < g(x^0), x^* - x^0 >$. Somit haben wir

$$< -g(x^0), x^* - x^0 >> 0, \forall x^* \in X^*. \tag{4.1.3}$$

Andererseits ist

$$\begin{aligned}\left\|(x^0 - \lambda g(x^0)) - x^*\right\|^2 &= \left\|x^0 - x^*\right\|^2 + \lambda^2 \left\|g(x^0)\right\|^2 + 2\lambda < -g(x^0), x^0 - x^* > \\ &= \left\|x^0 - x^*\right\|^2 + \lambda \left\|g(x^0)\right\|^2 [\lambda + \frac{2}{\left\|g(x^0)\right\|^2} < -g(x^0), x^0 - x^* >]\end{aligned}$$

Nach (4.1.3) ist $[\lambda + ...] < 0$, wenn $\lambda > 0$, λ - genügend klein. Somit ist für solche λ

$$\rho(x^0 - \lambda g(x^0), x^*) < \rho(x^0, x^*), \forall x^* \in X^*.$$

Nach Hilfssatz 2.14 der Grundvorlesung "Optimierung I" wird $\rho(x^0, X^*)$ ein einem Punkt $x^{**} \in X^*$ angenommen (da X^* konvex und abgeschlossen). Nach dem bereits bewiesenen ist $\rho(x^0 - \lambda g(x^0), x^{**}) < \rho(x^0, X^*)$ bei $\lambda > 0$, λ genügend klein. Um so mehr gilt die Behauptung des Hilfssatzes. □

Bemerkung.

1. Offensichtlich gilt die Behauptung des Hilfssatzes auch, wenn wir $g(x^0)$ durch $\frac{g(x^0)}{\|g(x^0)\|}$ ersetzen.

2. Ein Iterationsverfahren vom Typ

$$x^{k+1} := x^k - \lambda_k \frac{g(x^k)}{\|g(x^k)\|} \text{ mit } g(x^k) \in \partial f(x^k) \tag{4.1.4}$$

hat folglich eine gewisse Berechtigung, wenn $\lambda_k > 0$, λ_k - genügend klein (x^{k+1} liegt näher an X^* als x^k).

3. Wird $\left\{x^k\right\}_{k=1}^{\infty}$ nach (4.1.4) generiert, so haben wir einerseits $\left\|x^{k+1} - x^k\right\| = \lambda_k, \forall k$. Wir sollten also λ_k in den ersten Iterationen nicht zu klein wählen.

   Andererseits ist nach (4.1.4)

$$\begin{aligned}\left\|x^{k+1} - x^0\right\| &= \left\|x^k - \lambda_k \frac{g(x^k)}{\|g(x^k)\|} - x^0\right\| \\ &= \left\|x^{k-1} - \lambda_{k-1} \frac{g(x^{k-1})}{\|g(x^{k-1})\|} - \lambda_k \frac{g(x^k)}{\|g(x^k)\|} - x^0\right\| \\ &= ... \\ &= \left\|x^0 - \sum_{i=0}^{k} \lambda_i \frac{g(x^i)}{\|g(x^i)\|} - x^0\right\|,\end{aligned}$$

   weshalb sich $\left\|x^{k+1} - x^0\right\| \leq \sum_{i=0}^{k} \lambda_i$ ergibt.

   Da wir für größere k erreichen wollen, daß $\left\|x^{k+1} - x^0\right\| \approx \left\|x^* - x^0\right\|$, so müssen wir davon ausgehen, daß $\left\|x^{k+1} - x^0\right\|$ nicht gerade klein ist. Folglich wird auch $\sum_{i=0}^{k} \lambda_i$ nicht gerade klein sein dürfen.

Zusammen mit der ersten Bemerkung (λ_k nicht zu klein wählen) legt das nahe, die λ_i so zu wählen, daß $\sum_{i=0}^{\infty} \lambda_i = \infty$.

Dabei muß möglichst $\lambda_i \to 0$ bei $i \to \infty$ gelten, denn $\lambda_i = \left\|x^{i+1} - x^i\right\|$ und wir möchten ja erreichen, daß $x^i \to x^*$ bei $i \to \infty$.

In der weiter unten formulierten Konvergenzaussage für Verfahren vom Typ (4.1.4) werden wir die eben formulierten Forderungen an die Schrittweiten λ_k wiederfinden.

Für konvexes f mit dom$f = \mathbb{R}^n$ existiert nach Folgerung 2.2 $f'(x^0; d)$, $\forall x^0 \in \mathbb{R}^n$, $\forall d \in \mathbb{R}^n$.

Definition. Sei $f : \mathbb{R}^n \to \mathbb{R}$, f in x^0 in jeder Richtung $d \in \mathbb{R}^n$ richtungsableitbar. Wenn ein $d^* \in \mathbb{R}^n$ mit $\|d^*\| = 1$ existiert, so daß $f'(x^0; d^*) = \inf \left\{ f'(x^0; d) : \|d\| = 1 \right\}$, dann sagen wir, daß d^* Richtung des steilsten Abstiegs von f in x^0 ist.

Bemerkung. Aus der Definition der Richtungsableitung (vgl. Abschnitt 2.1) folgt sofort, daß für kleine $\lambda > 0$

$$f(x^0 + \lambda d) \approx f(x^0) + \lambda \cdot f'(x^0; d), \ \ \forall d.$$

Das motiviert die Begriffsbildung und wenn $f'(x^0; d^*) < 0$, dann ist $f(x^0 + \lambda d^*) < f(x^0)$ bei $\lambda > 0$, λ - genügend klein.

Hilfssatz 4.2. *Wir betrachten Aufgabe (4.0.1) mit $f : \mathbb{R}^n \to \mathbb{R}$, f im $\mathbb{R}^n$ konvex. Es sei $f(x^0) > f^*$. Dann existiert in $\partial f(x^0)$ ein Element $g(x^0)$ so, daß $(-g(x^0))$ Abstiegsrichtung von f in x^0 ist. Ja, es gilt sogar: Für die optimale Lösung $g^*(x^0)$ der Aufgabe*

$$\inf \left\{ \|a\| : a \in \partial f(x^0) \right\} \tag{4.1.5}$$

ist

$$0 > -\left\|g^*(x^0)\right\| = f'(x^0; -\frac{g^*(x^0)}{\|g^*(x^0)\|}) = \inf \left\{ f'(x^0; d) : \|d\| = 1 \right\},$$

d.h. $-\frac{g^*(x^0)}{\|g^*(x^0)\|}$ ist Richtung des steilsten Abstiegs von f in x^0.

Beweis.

1. Nach Satz 2.4 ist $\partial f(x^0)$ eine nichtleere, konvexe, abgeschlossene Menge. Folglich besitzt Aufgabe (4.1.5) eine optimale Lösung (siehe auch Hilfssatz 2.14 der Grundvorlesung "Optimierung I"). Da x^0 laut Voraussetzung nicht optimal in (4.0.1) ist, so ist nach Bemerkung 2 zu Satz 3.1 $0 \notin \partial f(x^0)$. Folglich gilt $\left\|g^*(x^0)\right\| > 0$.

2. Nach der bekannten Winkeleigenschaft (vgl. Hilfssatz 2.14 der Grundvorlesung "Optimierung I") für die Projektion (hier $g^*(x^0)$ eines Punktes (hier 0) auf eine konvexe Menge (hier $\partial f(x^0)$) gilt

$$< g^*(x^0), a > \geq \left\|g^*(x^0)\right\|^2 > 0, \forall a \in \partial f(x^0).$$

Nach der Formel für die Richtungsableitung (Satz 2.6) gilt somit

$$\begin{aligned} f'(x^0; -\frac{g^*(x^0)}{\|g^*(x^0)\|}) &= \frac{1}{\|g^*(x^0)\|} \max_{a \in \partial f(x^0)} < a, -g^*(x^0) > \\ &= \frac{1}{\|g^*(x^0)\|} < g^*(x^0), -g^*(x^0) > \\ &= -\left\|g^*(x^0)\right\|. \end{aligned}$$

Für jedes $d \in \mathbb{R}^n$ mit $\|d\| = 1$ gilt nach derselben Richtungsableitungsformel sowie der Ungleichung von Cauchy-Schwarz deshalb

$$\begin{aligned} f'(x^0; d) &= \max_{a \in \partial f(x^0)} < a, d > \\ &\geq \max_{a \in \partial f(x^0)} \{- \|a\| \cdot 1\} \\ &= - \min_{a \in \partial f(x^0)} \|a\| \\ &= - \left\|g^*(x^0)\right\| \\ &= f'(x^0; -\frac{g^*(x^0)}{\|g^*(x^0)\|}), \end{aligned}$$

d.h. $d = -\frac{g^*(x^0)}{\|g^*(x^0)\|}$ ist Lösung von $\inf_{\|d\|=1} f'(x^0; d)$.

□

4.1.4 Kein geeignetes Stop-Kriterium

In der differenzierbaren Optimierung ist in Iterationsverfahren für (4.0.1) der Test $\left\|\nabla f(x^k)\right\| \leq \varepsilon$ ein geeigneter Abbruchstest. Nach Satz 3.1 ist dieser Test bei konvexer, i.a. nichtdifferenzierbarer Aufgabe (4.0.1) durch

$$\inf \left\{\|a\| : a \in \partial f(x^k)\right\} \leq \varepsilon$$

zu ersetzen.

Da wir in der Regel nicht das vollständige Subdifferential $\partial f(x^k)$, sondern nur ein $g(x^k) \in \partial f(x^k)$ zur Verfügung haben werden (vgl. Abschnitt 4.1.2) und bei $x^k \to x^* \in X^*$ bei weitem nicht $g(x^k) \to 0$ gelten muß (vgl. (4.0.1) für $f(x) = |x|$ bei $x^k \to 0$), so wird der Ersatztest $\left\|g(x^k)\right\| \leq \varepsilon$ anstelle $\inf \left\{\|a\| : a \in \partial f(x^k)\right\} \leq \varepsilon$ i.a. nicht funktionieren.

4.1.5 Klassische Methoden der Schrittweitenwahl versagen

In der differenzierbaren Optimierung wird in den Verfahren vom Gradiententyp ($x^{k+1} = x^k - \alpha_k D_k \nabla f(x^k)$) die Richtung ($-D_k \nabla f(x^k)$) üblicherweise als Abstiegsrichtung gewählt. Darauf basieren dann die Vorgehensweisen, um eine Schrittweite α_k festzulegen: Cauchy, Armijo, Powell, Goldstein, ...

Wie wir mit Beispiel 1 aus Abschnitt 4.1.2 belegten, wird i.a. ein $g(x) \in \partial f(x)$ aber keine Abstiegsrichtung in x sein. Daraus resultiert der Zwang, in der Nichtdifferenzierbaren Optimierung die Strategie der Wahl von α_k weitgehend vom aktuellen Iterationspunkt x^k und der Richtung $(-g(x^k))$ unabhängig(apriori) zu determinieren (vgl. Verfahren von Shor, Abschnitt 4.2).

Dennoch gelingt es in gewissen Situationen, die im Hilfssatz 4.1 vorausgesagte Annäherung an die Optimalmenge zu erreichen. Wenn der Optimalwert f^* von (4.0.1) bekannt ist, dann sichert das z.B. die von B.T. Poljak vorgeschlagene Schrittweite $\alpha_k = \frac{f(x^k)-f^*}{\|g(x^k)\|}$.

Hilfssatz 4.3. *Gegeben sei Aufgabe (4.0.1) mit $f : \mathbb{R}^n \to \mathbb{R}$, f im $\mathbb{R}^n$ konvex. Die Folge $\left\{x^k\right\}_{k=0}^{\infty}$ werde nach der Vorschrift*

$$x^{k+1} = x^k - \alpha_k \frac{g(x^k)}{\|g(x^k)\|}, k = 0, 1, 2, ... \tag{4.1.6}$$

mit $g(x^k) \in \partial f(x^k)$ gewonnen.

Wenn dabei

$$0 < \alpha_k < \frac{2(f(x^k) - f^*)}{\|g(x^k)\|}, k = 0, 1, 2, ..., \tag{4.1.7}$$

dann ist $\left\|x^{k+1} - x^*\right\| < \left\|x^k - x^*\right\|$, $\forall x^* \in X^*$, $k = 0, 1, 2, ...$.

Beweis. Nach dem Beweis von Hilfssatz 4.1 gilt die Aussage des hiesigen Hilfssatzes für $\left\{x^k\right\}_{k=0}^{\infty}$, wenn $\forall k$

$$[\alpha_k + \frac{2}{\|g(x^k)\|} < -g(x^k), x^k - x^* >] < 0.$$

(in 4.1.6 ist im Vergleich zu Hilfssatz 4.1 $g(x^0)$ durch $\frac{g(x^0)}{\|g(x^0)\|}$ ersetzt worden).

Wird α_k so gewählt, daß (4.1.7) eingehalten ist, dann haben wir

$$\begin{aligned}[...] &< \frac{2(f(x^k) - f^*)}{\|g(x^k)\|} + \frac{2}{\|g(x^k)\|} < -g(x^k), x^k - x^* > \\ &= \frac{2}{\|g(x^k)\|}[f(x^k) - f(x^*) - < g(x^k), x^k - x^* >] \\ &\leq 0.\end{aligned}$$

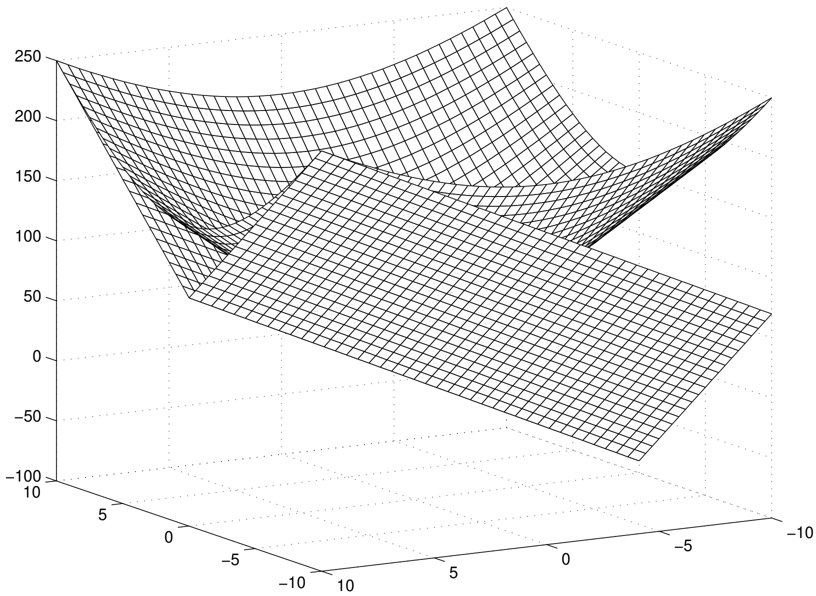
Die letzte Ungleichung gilt, weil nach Definition des Subgradienten $f(x^*) \geq f(x^k) + < g(x^k), x^* - x^k >$. ☐

4.1.6 Möglichkeit der Konvergenz gegen einen nichtoptimalen Punkt

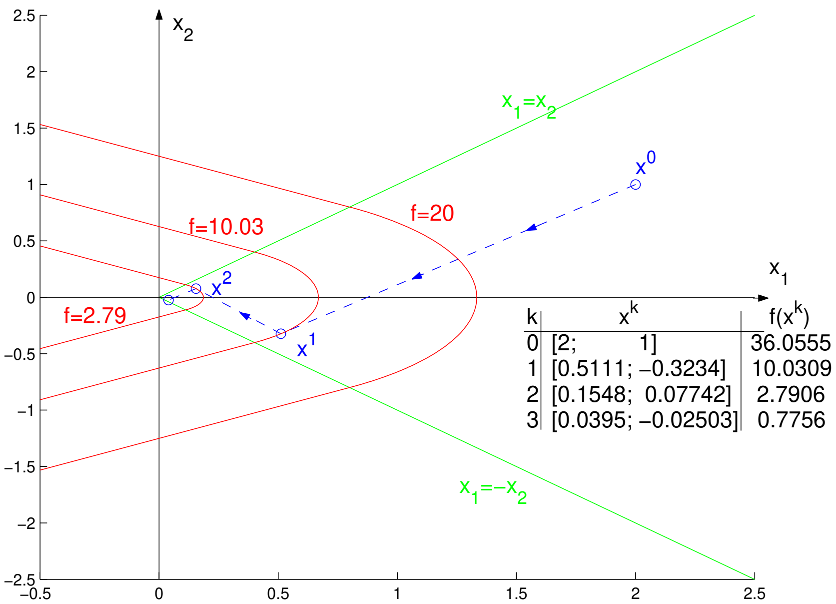
Beispiel 2: Man überzeugt sich leicht, daß sowohl $f_1(x) = 5\sqrt{9(x_1)^2 + 16(x_2)^2}$ als auch $f_2(x) = 9x_1 + 16|x_2|$ im $\mathbb{R}^2$ konvex sind. Außerdem ist auch

$$f(x) = \begin{cases} f_1(x), x_1 > |x_2| \\ f_2(x), x_1 \leq |x_2| \end{cases}$$

konvex im $\mathbb{R}^2$. f ist bei $x = [x_1, x_2]$ mit $x_1 \leq 0$, $x_2 = 0$ nicht differenzierbar (aber bei $x = [x_1, x_2]$ mit $x_1 = |x_2|$, $x \neq 0$, für diese Punkte $f_1(x) = f_2(x)$).



In der nachfolgenden Skizze sind die Niveaulinien von f dargestellt.



Die Folge $\{x^k\}_{k=0}^{\infty}$ werde nach $x^{k+1} := x^k - \alpha_k g(x^k)$, $k = 0, 1, 2, ...$ gewonnen, wobei $x^0 = [x_1^0, x_2^0]$ so gewählt sei, daß $x_1^0 > |x_2^0|$ und $\alpha_k \geq 0$ so festgelegt sei, daß

$$f(x^k - \alpha_k g(x^k)) = \min_{\alpha \geq 0} f(x^k - \alpha g(x^k)).$$

Man rechnet leicht nach, daß dann alle Iterationspunkte $x^k = [x_1^k, x_2^k]$ die Bedingung $x_1^k > |x_2^k|$ erfüllen, d.h. im ersten und vierten Quadranten zwischen den Geraden $x_1 = \pm x_2$ liegen, und daß $x^k \to 0$ bei $k \to \infty$. Deshalb ergibt sich $f(x^k) \to 0$ bei $k \to \infty$.

Demgegenüber ist

$$f^* = \lim_{x_1 \to \infty} f(x_1, 0) = -\infty.$$

Für die eben benutzten Behauptungen seien die wesentlichen Schritte ihres Beweises hier angegeben.

Eine kleine Rechnung ergibt, daß

$$x^{k+1} = [x_1^{k+1}, x_2^{k+1}] = [(1 - 9\delta_k)x_1^k, (1 - 16\delta_k)x_2^k],$$

wobei

$$\delta_k = \frac{9^2(x_1^k)^2 + 16^2(x_2^k)^2}{9^3(x_1^k)^2 + 16^3(x_2^k)^2}. \tag{4.1.8}$$

Daraus ergibt sich

$$\frac{|x_1^{k+1}|}{|x_2^{k+1}|} = \frac{16^2}{9^2}\frac{|x_2^k|}{|x_1^k|}. \tag{4.1.9}$$

Da $\frac{1}{16} < \delta_k < \frac{1}{9}$, $\forall k$, so folgt aus (4.1.8) und (4.1.9) sowie der Annahme $1 < |\frac{x_1^0}{x_2^0}| < \frac{16^2}{9^2}$, daß $1 < |\frac{x_1^k}{x_2^k}| < \frac{16^2}{9^2}$, $\forall k$ und daß bei $x_1^0 > 0$, $x_2^0 > 0$, stets $x_1^k > 0$, $(-1)^k x_2^k > 0$, $\forall k$, sowie

$$|x_1^{k+1}| < \frac{7}{16}|x_1^k|, |x_2^{k+1}| < \frac{7}{9}|x_2^k|, \forall k.$$

Damit sind die obigen Behauptungen bestätigt.

4.2 Subgradientenverfahren von N.Z. Shor

4.2.1 Das Verfahren SGV und seine Konvergenz bzgl. der Zielfunktionswerte

Seien gegeben

1. ein Startpunkt $x^0 \in \mathbb{R}^n$,
2. eine im $\mathbb{R}^n$ konvexe Funktion $f : \mathbb{R}^n \to \mathbb{R}$,
3. ein Orakel, welches zu x den Wert $f(x)$ und einen Vektor $g(x) \in \partial f(x)$ berechnet,
4. eine Abbruchgenauigkeit $\varepsilon > 0$.

Zur Lösung von
$$f^* = \inf \{f(x) : x \in \mathbb{R}^n\} \tag{4.2.1}$$
werde folgendes Verfahren SGV betrachtet:

1.Schritt: Setze $k := 0$, $\varphi_0 := +\infty$.

2.Schritt: Teile dem Orakel x^k mit und erhalte $f(x^k)$ sowie $g(x^k)$.

3.Schritt: Wenn $\left\|g(x^k)\right\| \leq \varepsilon$, dann Stop. Andernfalls wähle ein $\lambda_k > 0$.

4.Schritt: Setze
$$x^{k+1} = x^k - \lambda_k \frac{g(x^k)}{\|g(x^k)\|}$$
$$\varphi_{k+1} = \min \left\{f(x^k), \varphi_k\right\}$$
und gehe mit $k := k+1$ wieder zum 2. Schritt über.

Satz 4.4. *1. Wenn das Verfahren SGV in der k-ten Iteration stoppt, dann gilt*
$$0 \leq f(x^k) - f(x) \leq \varepsilon \left\|x^k - x\right\|, \forall x \in T^f_{x^k} = \left\{x \in \mathbb{R}^n : f(x) \leq f(x^k)\right\},$$
d.h. bei Existenz einer optimalen Lösung x^* von (4.2.1) gilt u.a.
$$0 \leq f(x^k) - f^* \leq \varepsilon \left\|x^k - x^*\right\|.$$
Wenn d eine obere Schranke für den Durchmesser von $T^f_{x^k}$ ist, dann gilt somit
$$0 \leq f(x^k) - f^* \leq \varepsilon d.$$

2. Wenn die Schrittweiten λ_k den Bedingungen
$$\lambda_k > 0, \lim_{k\to\infty} \lambda_k = 0, \sum_{k=1}^{\infty} \lambda_k = +\infty \tag{4.2.2}$$
genügen und der Algorithmus nie stoppt, dann gilt
$$\lim_{k\to\infty} \varphi_k = f^*.$$

Beweis. 1. Sei $\left\|g(x^k)\right\| \leq \varepsilon$. Für $x \in T^f_{x^k}$ haben wir nach der Subgradientengleichung
$$\begin{aligned} 0 &\leq f(x^k) - f(x) \\ &\leq < g(x^k), x^k - x > \\ &\leq \left\|g(x^k)\right\| \left\|x^k - x\right\| \\ &\leq \varepsilon \left\|x^k - x\right\|, \end{aligned}$$
d.h. (1) gilt.

2. Wir nehmen an, der Algorithmus stoppe nie (z.B. bei $\varepsilon = 0$).

   Wir vermerken, daß $\lim_{k\to\infty} \varphi_k$ existiert, wenn die monotone Zahlenfolge $\{\varphi_k\}$ beschränkt ist. Andernfalls haben wir $\varphi_k \to -\infty$.

   Wir setzen die Zahlen α_j wie folgt fest:
$$\alpha_j = \begin{cases} f^* + \frac{1}{j}, \text{ wenn } f^* > -\infty, \\ -j, \text{ wenn } f^* = -\infty, j = 1, 2, ... \end{cases}$$

   Es sei $T_j^f = \{x \in \mathbb{R}^n : f(x) < \alpha_j\}$.

   Nach Definition von α_j ist $T_j^f \neq \emptyset, \forall j$. Außerdem ist T_j^f eine offene Menge, weil nach Folgerung 2.8 f im $\mathbb{R}^n$ stetig ist. Schließlich gilt die Behauptung $\lim_{k\to\infty} \varphi_k = f^*$ offensichtlich, wenn wir nachweisen, daß $\forall j$ in der Folge $\{x^k\}_{k=0}^{\infty}$ sich ein Element x^{k_j} mit $x^{k_j} \in T_j^f$ findet.

3. Entgegen dem zu Beweisenden existiere ein j_0, so daß $x^k \notin T_{j_0}^f, \forall k = 0, 1, 2, ...$ Damit ist $f(x^k) \geq \alpha_{j_0}$. Für $y \in T_{j_0}^f$ ist $f(y) < \alpha_{j_0}$. Folglich ist für beliebiges k
$$0 \geq f(y) - f(x^k) \geq < g(x^k), y - x^k > .$$
   Folglich ist auch
$$< \frac{g(x^k)}{\|g(x^k)\|}, y - x^k > \leq 0, \forall k, \forall y \in T_{j_0}^f. \tag{4.2.3}$$
   Sei $x' \in T_{j_0}^f$. Da $T_{j_0}^f$ offen ist, so existiert ein $\delta > 0$, so daß $B_\delta(x') \subset T_{j_0}^f$.

   Wir wählen $y^k := (x' + \delta \frac{g(x^k)}{\|g(x^k)\|}) \in B_\delta(x')$. Folglich ist $y^k \in T_{j_0}^f$, $\forall k$ und damit nach (4.2.3)
$$< \frac{g(x^k)}{\|g(x^k)\|}, x' + \delta \frac{g(x^k)}{\|g(x^k)\|} - x^k > \leq 0, \forall k,$$
   d.h.
$$\delta + < \frac{g(x^k)}{\|g(x^k)\|}, x' - x^k > \leq 0, \forall k. \tag{4.2.4}$$

4. Da $\lambda_k \to 0$, so existiert ein Index m mit der Eigenschaft
$$\lambda_k \leq \delta, \forall k \geq m. \tag{4.2.5}$$
   Damit haben wir (nach SGV ist $x^{k+1} - x^k = -\lambda_k \frac{g(x^k)}{\|g(x^k)\|}$)
$$\begin{aligned} \left\|x^{k+1} - x'\right\|^2 &= \left\|(x^{k+1} - x^k) + (x^k - x')\right\|^2 \\ &= \left\|x^{k+1} - x^k\right\|^2 + 2 < x^{k+1} - x^k, x^k - x' > + \left\|x^k - x'\right\|^2 \\ &= \lambda_k^2 - 2\lambda_k < \frac{g(x^k)}{\|g(x^k)\|}, x^k - x' > + \left\|x^k - x'\right\|^2 \\ &= \left\|x^k - x'\right\|^2 - \lambda_k [2 < \frac{g(x^k)}{\|g(x^k)\|}, x^k - x' > -\lambda_k]. \end{aligned}$$

Nach (4.2.4) ist damit

$$\left\|x^{k+1} - x'\right\|^2 \leq \left\|x^k - x'\right\|^2 - \lambda_k(2\delta - \lambda_k),$$

woraus sich nach (4.2.5) ergibt

$$\left\|x^{k+1} - x'\right\|^2 \leq \left\|x^k - x'\right\|^2 - \delta\lambda_k, \forall k \geq m.$$

Somit ist $\forall s \in \mathbb{N}$

$$\sum_{k=m}^{m+s} \left\|x^{k+1} - x'\right\|^2 \leq \sum_{k=m}^{m+s} \left\|x^k - x'\right\|^2 - \delta \sum_{k=m}^{m+s} \lambda_k,$$

d.h.

$$\left\|x^{m+s+1} - x'\right\|^2 \leq \|x^m - x'\|^2 - \delta \sum_{k=m}^{m+s} \lambda_k,$$

was nach $\sum_{k=m}^{m+s} \lambda_k \to \infty$ bei $s \to \infty$ den Widerspruch $\left\|x^{m+s+1} - x'\right\| < 0$ für genügend großes s ergibt. Folglich müssen wir unsere Annahme vom Beginn des Abschnitts (3) fallen lassen.

□

Bemerkung. 1. Wir vermerken, daß bei Erfülltsein der Bedignung von Teil (1) des Satzes $f^* > -\infty$ gilt, wogegen in Teil (2) des Satzes der Fall $f^* = -\infty$ nicht ausgeschlossen ist.

2. Die Bedingungen (4.2.2) des Satzes sind z.B. erfüllt wenn λ_k nach

$$\lambda_k = \frac{\gamma}{k^p + c}, 0 < p \leq 1, \gamma > 0, c \geq 0$$

oder nach

$$\lambda_k = \frac{\gamma}{k \ln k}, \gamma > 0, k \geq 2$$

gewählt wird.

3. Daß der Wahl der λ_k weite Grenzen gesetzt sind, zeigt u.a. folgende Aussage: Wenn die Zahlen $\{\lambda_k\}_{k=1}^{\infty}$ die Eigenschaft (4.2.2) haben und $0 < \alpha < \beta$, dann haben auch die Zahlen $\{\lambda'_k\}_{k=1}^{\infty}$ mit $\lambda'_k \in [\alpha\lambda_k, \beta\lambda_k]$ die Eigenschaft (4.2.2).

4. In der Praxis wird man $\lambda_{k+1} := \lambda_k$ so lange wählen, so lange sich erweist, daß $f(x^{k+1}) \leq f(x^k)$ und erst bei Gestörtsein dieser Ungleichung zu einem neuen λ_k nach (4.2.2) übergehen.

5. Im SGV ist es möglich, daß eines der x^k eine optimale Lösung ist, aber weder der Abbruchtest $\left\|g(x^k)\right\| \leq \varepsilon$ anspricht noch die Folge $\left\{x^k\right\}$ von da an stationär ($x^{k+s} = x^k, s \geq 1$) ist. In solch einem Fall wird aber die Folge $\{\varphi_k\}$ stationär sein.

Derselbe Effekt wird über weite Strecken (d.h. viele Iterationen) auftreten, wenn eines der x^k sehr nahe an der Optimalmenge war.

Deshalb hat auch der Abbruchstest

$$\varphi_k - \varphi_{k+s} \leq \varepsilon \text{ mit } s >> 0$$

eine gewisse Berechtigung.

6. Der Satz erlaubt eine Aussage über die Konvergenz der Funktionswerte. Um auch etwas über die Konvergenz der Argumente $x^k \to x^*$ aussagen zu können, müssen schärfere Forderungen an die λ_k bzw. bzgl. der Optimalmenge gestellt werden. Dazu mehr im Abschnitt 4.2.2.

4.2.2 Konvergenz bzgl. der Argumente

Satz 4.5. *In der konvexen Aufgabe (4.2.1) sei die Optimalmenge*

$$X^* = \{x \in \mathbb{R}^n : f(x) = f^*\} \neq \emptyset.$$

Die Folge $\left\{x^k\right\}_{k=0}^{\infty}$ *sei nach SGV erzeugt worden.*

1. *Es sei X^* außerdem beschränkt. Dann gilt:*
   (a) *Wenn das Verfahren in der k-ten Iteration stoppt, dann gilt*
   $$0 \leq f(x^k) - f^* \leq \varepsilon\rho(x^k, X^*).$$
   (b) *Wenn das Verfahren nie stoppt und die Bedingungen (4.2.2) aus Satz 4.4 erfüllt sind, dann existiert* $\lim_{k\to\infty} \rho(x^k, X^*)$ *und es gilt*
      i. $\lim_{k\to\infty} \rho(x^k, X^*) = 0$
      ii. *Die Folge* $\left\{x^k\right\}$ *besitzt wenigstens einen Häufungspunkt und jeder Häufungspunkt ist eine optimale Lösung von (4.2.1).*
2. *Wenn Bedingungen (4.2.2) aus Satz 4.4 und wenigstens eine der Bedingungen* ($b1$) *oder* ($b2$) *erfüllt sind, sowie das Verfahren nie stoppt, dann existiert* $\lim_{k\to\infty} x^k$ *und es gilt* $\lim_{k\to\infty} x^k \in X^*$. *Hier sei*
   ($b1$) $\sum_{k=1}^{\infty} (\lambda_k)^2 < \infty$
   ($b2$) X^* *ist eine lineare Mannigfaltigkeit.*
3. *Wenn $intX^* \neq \emptyset$ und Bedingungen (4.2.2) aus Satz 4.4 erfüllt sind, dann stoppt das Verfahren entweder nach endlich vielen Iterationen, weil* $\left\|g(x^k)\right\| \leq \varepsilon$, *oder es erzeugt nach endlich vielen Iterationen einen Punkt* x^k *mit der Eigenschaft* $x^k \in X^*$ *(in diesem Fall ist* $\varphi_k = f^*$ *und* $\varphi_{k+s} = \varphi_k,\ \forall s \geq 1$*).*

Beweis. 1. Wir wollen hier nur die Ausagen aus (1) untersuchen. Für (2) und (3) sei auf [Beiko, Bublik, Zin'ko] verweisen.

2. Es sei X^* beschränkt. Da nach Folgerung 2.8 f im $\mathbb{R}^n$ stetig ist, so ist X^* eine abgeschlossene Menge. Folglich existiert für beliebiges $\bar{x} \in \mathbb{R}^n$ ein $x^* = x^*(\bar{x}) \in X^*$, so daß

$$\rho(\bar{x}, X^*) = \inf \{\|\bar{x} - x\| : x \in X^*\} = \|\bar{x} - x^*\| .$$

Bei Stop nach $\left\|g(x^k)\right\| \leq \varepsilon$ gilt nach Teil (1) von Satz 4.4, daß

$$0 \leq f(x^k) - f(x) \leq \varepsilon \left\|x^k - x\right\|, \forall x \in T^f_{x^k} = \left\{x \in \mathbb{R}^n : f(x) \leq f(x^k)\right\}.$$

Da zu x^k ein $x^* \in X^*$ existiert, für das $\left\|x^k - x^*\right\| = \rho(x^k, X^*)$ und da dieses $x^* \in T^f_{x^k}$, so gilt

$$0 \leq f(x^k) - f(x^*) \leq \varepsilon\rho(x^k, X^*).$$

3. Es sei weiterhin X^* als beschränkt vorausgesetzt. Wir zeigen, daß dann aus $\lim_{k\to\infty} \left\|z^k\right\| = +\infty$ stets folgt, daß $\lim_{k\to\infty} f(z^k) = +\infty$.

Da $X^* \neq \emptyset$ und kompakt ist, so existiert zu $x^* \in X^*$ eine Kugel $B_\rho(0)$ vom Radius ρ mit Zentrum in 0, so daß $X^* \subset \{x^*\} + B_\rho(0)$.

O.E.d.A. können wir hierbei annehmen, daß

$$X^* \cap (\{x^*\} + S_\rho(0)) = \emptyset \tag{4.2.6}$$

($S_\rho(0)$ bezeichne die Sphäre vom Radius ρ mit Zentrum in 0).

Da $S_\rho(0)$ eine kompakte Menge ist und f im $\mathbb{R}^n$ stetig ist, so existiert

$$f_\rho = \min \{f(x) : x \in \{\{x^*\} + S_\rho(0)\}\}$$

und ist endlich. Nach (4.2.6) ist außerdem

$$f^* < f_\rho < +\infty. \tag{4.2.7}$$

Es sei nun $\left\{z^k\right\}$ so gewählt, daß $\left\|z^k\right\| \to +\infty$. Dann existiert ein k_0, so daß für $k \geq k_0$ gilt

$$\left\|z^k - x^*\right\| > \rho. \tag{4.2.8}$$

Wir betrachten die Punkte $y^k := x^* + \rho\frac{z^k - x^*}{\|z^k - x^*\|}$.

Für $k \geq k_0$ ist dann $y^k \in (\{x^*\} + S_\rho(0))$ und somit haben wir nach Definition von f_ρ

$$f(y^k) \geq f_\rho, k \geq k_0. \tag{4.2.9}$$

Nach (4.2.8) ist für

$$\alpha_k := \frac{\rho}{\|z^k - x^*\|}$$

$0 < \alpha_k < 1$ bei $k \geq k_0$.

Aufgrund der Konvexität von f haben wir damit für $y^k = \alpha_k z^k + (1 - \alpha_k)x^*$, daß

$$f(y^k) \leq \alpha_k f(z^k) + (1 - \alpha_k) f(x^*), k \geq k_0,$$

weil $f(x^*) = f^*$, so haben wir also

$$\begin{aligned} f(z^k) \geq \frac{1}{\alpha_k} f(y^k) + f^* - \frac{1}{\alpha_k} f^* &= \frac{1}{\alpha_k}(f(y^k) - f^*) + f^* \\ &= f^* + \frac{\left\| z^k - x^* \right\|}{\rho}(f(y^k) - f^*). \end{aligned}$$

Unter Verwendung von (4.2.9) erhalten wir somit

$$f(z^k) \geq f^* + \frac{\left\| z^k - x^* \right\|}{\rho}(f_\rho - f^*),$$

woraus nach (4.2.7) und $\left\| z^k \right\| \to +\infty$ folgt, daß $f(z^k) \to +\infty$.

4. Wir setzen nun voraus, daß das Verfahren (SGV) eine unendliche Folge $\left\{ x^k \right\}_{k=1}^\infty$ erzeuge.

   Nach Satz 4.4 haben wir $\lim_{k\to\infty} \varphi_k = f^*$. Weil $\varphi_k = \min_{i \leq i \leq k} f(x^i)$, so existiert folglich eine Teilfolge $\left\{ x^{k_i} \right\}_{i=1}^\infty$ von $\left\{ x^k \right\}_{k=1}^\infty$ mit der Eigenschaft

$$\lim_{i\to\infty} f(x^{i_k}) = f^*. \tag{4.2.10}$$

   Da $f^* < +\infty$, so muß nach Teil (3) des Beweises die Folge $\left\{ \left\| x^{k_i} \right\| \right\}_{i=1}^\infty$ beschränkt sein. Folglich existiert in $\left\{ x^{k_i} \right\}_{i=1}^\infty$ eine konvergente Teilfolge $\left\{ x^{k_{i_s}} \right\}_{s=1}^\infty$, d.h. es existiert ein $\bar{x}$, so daß

$$\lim_{s\to\infty} x^{k_{i_s}} = \bar{x}.$$

   Aus der Stetigkeit von f und aus (4.2.10) erhalten wir deshalb

$$\lim_{s\to\infty} f(x^{k_{i_s}}) = f(\bar{x}) = f^*.$$

   Damit ist Behauptung 1.(b).ii für die so konstruierte Folge bewiesen ($\left\{ x^k \right\}_{k=1}^\infty$ besitzt Häufungspunkt $\bar{x}$, $\bar{x}$ ist optimale Lösung). Daß jeder Häufungspunkt eine optimale Lösung ist, folgert man auf analoge Weise aus

$$\lim_{k\to\infty} \varphi_k = f^*.$$

□

4.2.3 Negativaussage zur Konvergenzgeschwindigkeit im algemeinen Fall

Eine Aussage zur Konvergenz eines Verfahrens ist nur die halbe Sache, wenn nicht gleichzeitig eine Aussage über die Konvergenzgeschwindigkeit mitgeliefert wird.

Zur Sprachverständigung seien hier die Definition von linearer, quadratischer und geometrische Konvergenzgeschwindigkeit wiederholt.

Definition. Es sei $\left\{x^k\right\}_{k=1}^{\infty}$ eine Folge mit $x^k = \lim_{k\to\infty} x^k$.

(a) Wir sagen, daß $x^k \to x^*$ mit wenigstens linearer Konvergenzgeschwindigkeit, wenn ein $q \in [0,1)$ und ein k_0 existieren, so daß

$$\left\|x^{k+1} - x^*\right\| \leq q \left\|x^k - x^*\right\|, \; k \geq k_0.$$

(b) Wenn die Existenz einer Nullfolge $\{\varepsilon_k\}_{k=1}^{\infty}$ gesichert ist, so daß

$$\left\|x^{k+1} - x^*\right\| \leq \varepsilon_k \left\|x^k - x^*\right\|, \; k \geq 1,$$

so spricht man von überlinearer Konvergenzgeschwindigkeit.

(c) Wenn existieren $c > 0$ und k_0, so daß

$$\left\|x^{k+1} - x^*\right\| \leq c \left\|x^k - x^*\right\|^2, \; k \geq k_0$$

so spricht man von wenigstens quadratischer Konvergenzgeschwindigkeit.

(d) Schließlich spricht man von geometrischer Konvergenzgeschwindigkeit, wenn ein $c > 0$ und ein $q \in [0,1)$ sowie ein k_0 existieren, so daß

$$\left\|x^{k+1} - x^*\right\| \leq c q^{k-k_0}, \; k \geq k_0$$

Bemerkung. 1. Man kann zeigen, daß (a) äquivalent mit

$$\limsup_{k\to\infty} \frac{\left\|x^{k+1} - x^*\right\|}{\left\|x^k - x^*\right\|} < 1,$$

(b) äquivalent mit

$$\lim_{k\to\infty} \frac{\left\|x^{k+1} - x^*\right\|}{\left\|x^k - x^*\right\|} = 0$$

und (c) äquivalent mit

$$\limsup_{k\to\infty} \frac{\left\|x^{k+1} - x^*\right\|}{\left\|x^k - x^*\right\|^2} < \infty$$

ist.

2. (b) und (c), (d) sind unäbhangig von der gewählten Norm. Ob (a) vorliegt oder nicht, kann von der gew"hlten Norm abhängen.

3. Generall gilt: quadratischer Konvergenz⇒ überlinearer Konvergenz⇒ linearer Konvergenz⇒geometrischer Konvergenz.

4. Für die Aufgabe des freien Optimums einer differenzierbaren Funktion sind für die Verfahren vom Gradienten-Typ. Aussagen über lineare Konvergenzgeschwindigkeit bekannt, für die gleiche Aufgabe sind für Verfahren vom Newton-Typ Aussagen über quadratische Konvergenzgeschwindigkeit bekannt (siehe Grundvorlesung "Nichtdifferenzierbare Optimierung").

5. Für das Verfahren(SGV) sind die Aussagen zur Konvergenzgeschwindigkeit eher schwach. Für eine spezielle Klasse von konvexen, nichtdifferenzierbaren Funktionen(vgl. z.B. N.Z. Shor: Nondifferentiable Optimization and Polynomial Problems, Kluwer Academic Publishers 1998) zeigte Shor geometrische Konvergenzgeschwindigkeit der erzeugten Folge $\{x^k\}$. Für die Wahl der Schrittweite nach B.T. Poljak (vgl. Satz 4.8)

Satz 4.6. *Im Verfahren (SGV) aus Abschnitt 4.2.1 trete nie "Stop" ein, aber die Ausgangsfolge $\{x^k\}_{k=1}^{\infty}$ konvergiere gegen eine optimale Lösung x^*. Dann ist die Konvergenzgeschwindigkeit i.a. schwächer als geometrisch.*

Beweis. Entgegen dem zu Beweisenden existiere ein k_0 sowie ein $c > 0$ und ein $q \in [0, 1)$, so daß

$$\left\|x^{k+1} - x^*\right\| \leq cq^k, \forall k \geq k_0.$$

Da im Verfahren

$$x^{k+1} = x^k - \lambda_k \frac{g(x^k)}{\|g(x^k)\|}, \forall k,$$

so gilt

$$\lambda_k = \left\|x^{k+1} - x^k\right\| \leq \left\|(x^{k+1} - x^*) - (x^k - x^*)\right\| \leq \left\|x^{k+1} - x^*\right\| - \left\|x^k - x^*\right\| \leq cq^k + cq^{k-1}.$$

Daraus ergibt sich

$$\sum_{k=k_0+1}^{\infty} \lambda_k \leq c \sum_{k=k_0+1}^{\infty} (q^k + q^{k-1}) = c\frac{q^{k_0+1} + q^{k_0}}{1 - q} < +\infty. \tag{4.2.11}$$

Im Satz 4.4 hatten wir aber $\sum_{k=k_0+1}^{\infty} \lambda_k = +\infty$ als hinreichend für die Konvergenz von (SGV) und in Bemerkung 3 nach Hilfssatz 4.1 mehr oder weniger als notwendig für die Konvergenz erkannt. Nach (4.2.11) widerspricht $\sum_{k=k_0+1}^{\infty} \lambda_k = +\infty$ aber der geometrischen Konvergenzgeschwindigkeit. □

4.2.4 Schrittweitenwahl nach B.T. Poljak

Wir betrachten weiterhin die Aufgabe

$$f^* = \inf\{f(x) : x \in \mathbb{R}^n\} \tag{4.2.12}$$

mit $f : \mathbb{R}^n \to \mathbb{R}$, f - konvex im $\mathbb{R}^n$.

Das Shor'sche Verfahren SGV krankt daran, daß die Schrittweiten λ_k sehr langsam gegen Null gestrebt werden müssen um Konvergenz zu sichern.

Für den Fall, daß der Optimalwert f^* von (4.2.12) bekannt ist vor, die Schrittweiten nach

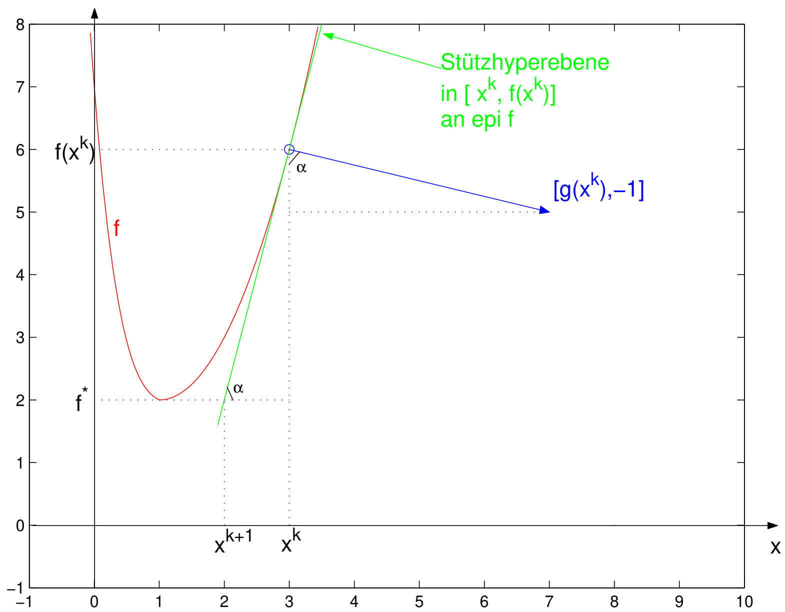
$$\lambda_k = \frac{f(x^k) - f^*}{\|g(x^k)\|} \tag{4.2.13}$$

zu wählen. Es erweist sich, daß dann die Konvergenzeigenschaften (praktisch und theoretisch) des (SGV) günstiger ausfallen.

Bemerkung. 1. Es gibt durch aus Aufgaben, in denen f^* apriori bekannt ist (z.B. beim Lösen eines lösbaren Ungleichungssystems über eine Optimierungsaufgabe, vgl. Abschnitt 1.5).

2. In vielen Fällen ist zwar nicht f^*, aber eine Schaätzung für f^* bekannt (vgl. die Beispiele aus den Abschnitten 1.2-1.8). Wir werden weiter unten wie man dann mit λ_k nach (4.2.13) umgehen kann.

3. Für den Fall $n = 1$ wollen wir eine geometrische Motivation für die Wahl der Schrittweite nach (4.2.13) herleiten.



Wir haben einerseits $\tan\alpha = \frac{f(x^k)-f^*}{x^k-x^{k+1}}$. Andererseits gilt für den mit der Stützhyperebene verbundenen Subgradienten (vgl. geometrische Deutung des Subgradienten aus Abschnitt 2.2.1) $\tan\alpha = \frac{g(x^k)}{1}$.

Aus

$$\frac{f(x^k)-f^*}{x^k-x^{k+1}} = g(x^k)$$

ergibt sich $f(x^k)-f^* = (x^k-x^{k+1})g(x^k)$, woraus wir nach Multiplikation mit $g(x^k)$ erhalten $(f(x^k)-f^*)g(x^k) = (x^k-x^{k+1})g^2(x^k)$, d.h.

$$x^k = x^k - \frac{f(x^k)-f^*}{|g(x^k)|}\frac{g(x^k)}{|g(x^k)|} = x^k - \lambda_k \frac{g(x^k)}{|g(x^k)|}.$$

Die Schrittweitenwahl nach (4.2.13) entspricht somit der Wahl des nächsten Iterationspunktes x^{k+1} als einen Punkt, auf dem die Stüzthyperebene in x^k an $f(x)$ das Niveau f^* erreicht.

Satz 4.7. *Gegeben sei Aufgabe (3.3.1) mit $f : \mathbb{R}^n \to \mathbb{R}$, f - konvex. Es sei die Optimalmenge $X^* = \{x : f(x) = f^*\} \neq \emptyset$. Auf (3.3.1) werde (SGV) aus Abschnitt 4.2.1 mit*

$$\lambda_k = \frac{f(x^k)-f^*}{\|g(x^k)\|}, \forall k,$$

angewendet.

Dann gilt:

(i) Wenn das Verfahren in der k-ten Iteration stoppt, dann gilt

$$0 \le f(x^k) - f^* \le \varepsilon\rho(x^k, X^*).$$

(ii) Wenn das Verfahren nie stoppt, dann existiert $\lim_{k\to\infty} f(x^k)$ und es gilt $\lim_{k\to\infty} f(x^k) = f^*$. Außerdem existiert dann eine Teilfolge der Ausgangsfolge $\left\{x^k\right\}_{k=1}^{\infty}$ die gegen ein $x^* \in X^*$ konvergiert. Wenn X^* ein elementig ist, dann konvergiert die Folge $\left\{x^k\right\}_{k=1}^{\infty}$ selbst gegen die optimale Lösung x^*.

Beweis. 1. Teil (i) der Satzaussage beweist man ebenso wie wir Teil $(i1)$ des Satzes 4.5 bewiesen haben, da die Wahl der Schrittweite dort ohne Bedeutung war und da der Abstand $\rho(x^k, X^*)$ auch dann existiert, wenn X^* nicht beschränkt ist.

2. Wir zeigen, daß $f^* = \lim_{k\to\infty} f(x^k)$, wenn das Verfahren nie stoppt.

   (a) Wir betrachten zuerst den stationären Fall. Genauer: Es sei $\lambda_{k_0} = 0$. Dann ist $f(x^{k_0}) = f^*$, d.h. x^{k_0} ist eine optimale Lösung und (im Gegensatz zur generellen apriori-Wahl der λ_k) merken wir das auch, denn dann ist $x^{k_0+1} = x^{k_0}$, was nach sich zieht, daß

$$\lambda_{k_0+1} = \frac{f(x^{k_0+1})-f^*}{\|g(x^{k_0+1})\|} = 0.$$

Somit ist $\lambda_k = 0$, $x^k = x^{k_0}$, $\forall k \geq k_0$, d.h. die Behauptung $f^* = \lim_{k\to\infty} f(x^k)$ ist erfüllt (vermerke, daß durch unsere Vorgehenweise $g(x^k) \neq 0$, $\forall k \geq k_0$).

(b) Es sei deshalb im weiteren $x^k \notin X^*$, $\forall k$, d.h. $\lambda_k > 0$, $\forall k$.
Wir fixieren ein $x^* \in X^*$ (existiert, da $X^* \neq \emptyset$ vorausgesetzt wurde).
Wir haben dann

$$\begin{aligned}\left\|x^{k+1} - x^*\right\|^2 &= \left\|x^k - \lambda_k \frac{g(x^k)}{\|g(x^k)\|} - x^*\right\|^2 \\ &= \left\|x^k - x^*\right\|^2 - 2\lambda_k < \frac{g(x^k)}{\|g(x^k)\|}, x^k - x^* > +\lambda_k^2.\end{aligned}$$

Aufgrund der Subgradientenungleichung und der speziellen Wahl der Schrittweite haben wir

$$< g(x^k), x^* - x^k > \leq f(x^*) - f(x^k) = f^* - f(x^k) = -\lambda_k \left\|g(x^k)\right\|.$$

Damit ergibt sich

$$\left\|x^{k+1} - x^*\right\|^2 \leq \left\|x^k - x^*\right\|^2 - 2\lambda_k^2 + \lambda_k^2 = \left\|x^k - x^*\right\|^2 - \lambda_k^2. \quad (4.2.14)$$

Folglich ist $\left\{\left\|x^k - x^*\right\|\right\}_{k=1}^{\infty}$ eine monoton fallende Folge. Also existiert $\lim_{k\to\infty} \left\|x^k - x^*\right\|$ und die Folge $\left\{x^k\right\}_{k=1}^{\infty}$ ist beschränkt ($\lim_{k\to\infty} \left\|x^k - x^*\right\| = 0$ läßt sich generell nicht zeigen).
Durch Aufaddieren der Gleichungen (4.2.14) erhalten wir

$$\begin{aligned}\sum_{k=1}^{m} \lambda_k^2 &\leq \sum_{k=1}^{m} (\left\|x^k - x^*\right\|^2 - \left\|x^{k+1} - x^*\right\|^2) \\ &= \left\|x^1 - x^*\right\|^2 - \left\|x^{m+1} - x^*\right\|^2 \leq \left\|x^1 - x^*\right\|^2.\end{aligned}$$

Folglich ist

$$\sum_{k=1}^{m} \lambda_k^2 \leq \left\|x^1 - x^*\right\|^2 < \infty,$$

d.h. $\lim_{k\to\infty} \lambda_k = 0$, was nach Wahl der λ_k bedeutet, daß

$$\lim_{k\to\infty} \frac{f(x^k) - f^*}{\|g(x^k)\|} = 0 \quad (4.2.15)$$

Da aber $x^k \in \operatorname{int} \operatorname{dom} f = \mathbb{R}^n$, so ist nach Folgerung 2.12 die Menge $\partial f(x^k)$ kompakt. Aus der bereits bewiesenen Beschränktheit der Folge $\left\{x^k\right\}_{k=1}^{\infty}$ folgt dann aber, daß $\left\{\left\|g(x^k)\right\|\right\}$ eine Beschränkte Folge ist. Dann folgt aber aus (4.2.15), daß $\lim_{k\to\infty}(f(x^k) - f^*) = 0$. Der erste Teil von Behauptung (*ii*) gilt also auch unter (2.b).

3. Nach dem bereits Bewiesenen ist $\{x^k\}_{k=1}^{\infty}$ eine beschränkte Folge. Somit enthält sie eine konvergente Teilfolge $\{x^{k_i}\}_{k=1}^{\infty}$, d.h. es existiert ein $\bar{x}$, so daß $\bar{x} = \lim_{i\to\infty} x^{k_i}$. Aufgrund der Stetigkeit von f haben wir $\lim_{i\to\infty} f(x^{k_i}) = f(\bar{x})$. Da aber nach (2) $f^* = \lim_{k\to\infty} f(x^k)$, so ist $f(\bar{x}) = f^*$, d.h. die Teilfolge $\{x^{k_i}\}$ konvergiert gegen eine optimale Lösung $\bar{x}$.

4. Wir betrachten noch den Fall $X^* = \{x^*\}$, d.h. X^* sei einelementig. Nach Teil (3) muß $\bar{x} = x^*$ sein, d.h. nach Teil (3) ist $\|x^{k_i} - \bar{x}\| \to 0$ bei $i \to \infty$.

   Da aber nach Teil (2.b) die Folge $\{\|x^k - \bar{x}\|\}$ monoton fallend ist, so muß auch $\|x^k - \bar{x}\| \to 0$ bei $k \to \infty$ gelten, d.h. $\lim_{k\to\infty} x^k$ existiert und wir haben $\lim_{k\to\infty} x^k = \bar{x}$.

□

Bemerkung. 1. Wir wollen den Fall diskutieren, daß uns anstelle des exakten Optimalwertes f^* der Aufgabe (3.3.1) nur eine Schätzung $\bar{f}$ bekannt ist. Es werde (SGV) mit

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

durchgeführt.

Sollte während des Verfahrens ein x^k konstruiert werden, bei dem $f(x^k) < \bar{f}$, so ist offensichtlich $f^* < \bar{f}$. Man setzte deshalb in diesem Fall $\bar{f} := \bar{f} - \varepsilon$. Ist im Verlaufe der Realisierung des Verfahrens zu beobachten, daß vermutlich $f(x^k) \to \bar{f}$ nicht gilt, so setze man $\bar{f} := \bar{f} + \varepsilon$ (da vermutlich $f^* > \bar{f}$).

Sollte zu beobachten sein, daß $\|g(x^k)\| \to \infty$, so liegt vermutlich ebenfalls die Situation $f^* > \bar{f}$ vor (die andere Situation hätte sich durch $f(x^k) < \bar{f}$ für ein x^k äßern müssen). Man setze deshalb auch hier $\bar{f} := \bar{f} + \varepsilon$.

2. Wir wollen abschließend noch zwei Aussagen zur Konvergenzgeschwindigkeit machen, die zeigen, daß $f(x^k) \to f^*$ nicht langsame als $\frac{1}{\sqrt{k}}$ strebt.

Satz 4.8. *Unter den Voraussetzungen von Satz 4.7 gilt*

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

Beweis. Per Definition existiert $\liminf_{k\to\infty}$, denn so ist der Grenzwert der monoton nichtwachsenden Folge $\{g_k\}$ mit $g_k = \min_{1\le i\le k} \{\sqrt{i}(f(x^i) - f^*)\}$ zu betrachten.

Entgegen dem zu Beweisenden sei

$$\liminf_{k\to\infty} \sqrt{k}(f(x^k) - f^*) \neq 0.$$

Da $f(x^k) > f^*$, so ist dann $\liminf_{k\to\infty} \sqrt{k}(f(x^k) - f^*) > 0$. Deshalb existieren dann ein k_0 und ein $a > 0$, so daß

$$\sqrt{k}(f(x^k) - f^*) > a, \forall k \geq k_0.$$

Wir haben dann

$$(f(x^k) - f^*)^2 > \frac{a^2}{k}, \forall k \geq k_0,$$

und damit

$$\sum_{k=k_0}^{\infty} (f(x^k) - f^*)^2 > a \sum_{k=k_0}^{\infty} \frac{1}{k} = \infty.$$

Im Beweis von Satz 4.7 haben wir gezeigt, daß die $\left\|g(x^k)\right\|$ im Verfahren beschränkt bleiben. Folglich ist dann auch

$$\sum_{k=k_0}^{\infty} \left(\frac{f(x^k) - f^*}{\|g(x^k)\|}\right)^2 = \infty.$$

Das aber bedeutet $\sum_{k=k_0}^{\infty} \lambda_k^2 = \infty$, was im Widerspruch zur Aussage $\sum_{k=1}^{\infty} \lambda_k^2 = \infty$ im Beweis von Satz 4.7 steht. □

Bemerkung. Für den Fall des sogenannten spitzen Minimums läßt sich lineare Konvergenzgeschwindigkeit für das SGV mit Poljak'sche Schrittweite beweisen (vgl. Satz ??4.9).

Definition. Wenn ein $\alpha > 0$ existiert, so daß

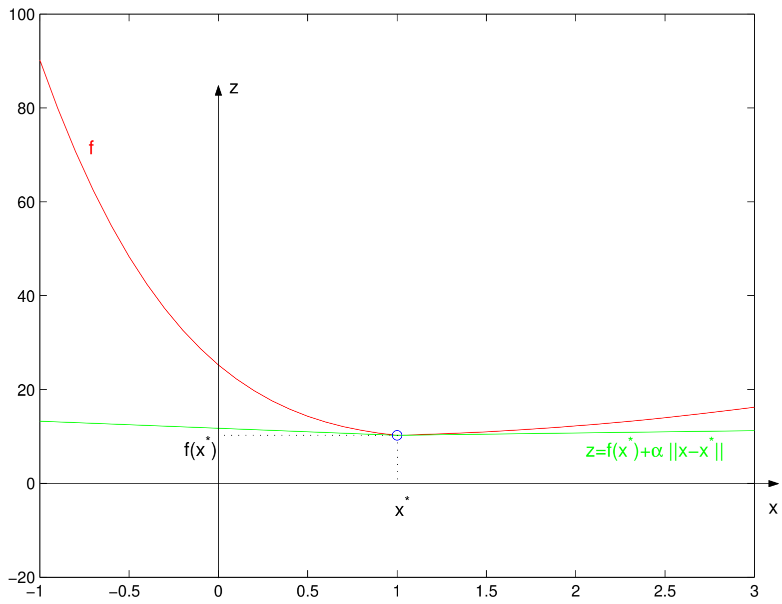
$$f(x) \geq f(x^*) + \alpha \|x - x^*\|, \forall x \in \mathbb{R}^n,$$

dann heißt x^* spitzer Minimumspunkt von f.

Bemerkung.

1. Im betrachteten Fall hat (3.3.1) offensichtlich die einelementige Optimalmenge $X^* = \{x^*\}$.

2. Geometrisch bedeutet die Eigenschaft, daß der Graph von f oberhalb des Kegels mit der Gleichung $z = \alpha \|x - x^*\| + f(x^*)$ liegt.



Satz 4.9. *Unter den Voraussetzungen von Satz 4.7 gelte*

$$f(x) \geq f(x^*) + \alpha \|x - x^*\|, \forall x \in \mathbb{R}^n, \tag{4.2.16}$$

wobei $\alpha > 0$. *Dann existiert zu der nach (SGV) erzeugten Folge* $\{x^k\}$ *ein* k_0 *und ein* $q \in [0, 1)$, *so daß*

$$\left\|x^{k+1} - x^*\right\|^2 \leq q \left\|x^k - x^*\right\|^2. \tag{4.2.17}$$

Beweis. Nach Teil (ii) von Satz 4.7 existiert $\lim_{k\to\infty} x^k$ und es gilt $x^* = \lim_{k\to\infty} x^k$.

Nach Teil (2) des Beweises von Satz 4.7 haben wir

$$\left\|x^{k+1} - x^*\right\|^2 \leq \left\|x^k - x^*\right\|^2 - \lambda_k^2 = \left\|x^k - x^*\right\|^2 - \left(\frac{f(x^k) - f^*}{\|g(x^k)\|}\right)^2.$$

Wie bereits mehrfach vermerkt ist $\left\{\left\|g(x^k)\right\|\right\}_{k=1}^{\infty}$ beschränkt. Folglich existiert ein C, so daß $\left\|g(x^k)\right\| \leq C$, $\forall k$. Somit erhalten wir unter Benutzung der Eigenschaft (4.2.16)

$$\left\|x^{k+1} - x^*\right\|^2 \leq \left\|x^k - x^*\right\|^2 - \left(\frac{f(x^k) - f^*}{C}\right)^2 \leq \left\|x^k - x^*\right\|^2 - \left(\frac{\alpha \left\|x^k - x^*\right\|}{C}\right)^2,$$

d.h. $\left\|x^{k+1} - x^*\right\|^2 \leq (1 - (\frac{\alpha}{C})^2) \left\|x^k - x^*\right\|$, $\forall k$.

Da nach (4.2.16) $c \geq \alpha > 0$ sein muß, so ist $0 \leq q = \sqrt{1 - (\frac{\alpha}{C})^2} < 1$ und die Behauptung (4.2.17) ist für $k_0 = 1$ beweisen. $\square$

4.3 Die gesamte Anzahl der Seiten