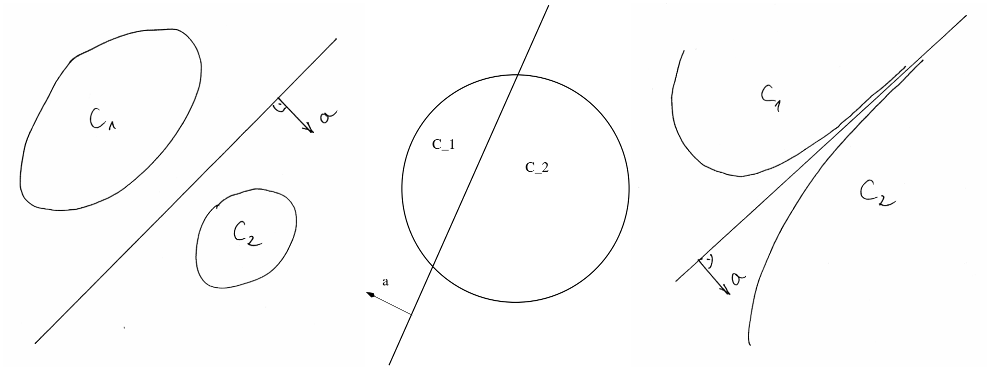


Abbildung 12.4. Trennung konvexer Mengen durch eine Hyperebene mit Normalenvektor a

Dann ist $\operatorname{int}(C)$ nicht leer (warum?) und [als Minkowski-Summe von C_2 und $-C_1$] wieder konvex. Wegen $C_1 \cap C_2 = \emptyset$ gilt $0 \notin C$, also erst recht $0 \notin \operatorname{int}(C)$. Aus Lemma 12.6 bekommen wir ein $a \in \mathbb{R}^n$, $a \neq 0$ mit $0 \le a^\top x$ für alle $x \in C$ [sogar $x \in \overline{C}$].

$$\Rightarrow \quad a^\top x_1 \le a^\top x_2 \quad \text{für alle } x_1 \in C_1,\ x_2 \in C_2.$$

Um (12.5) zu erhalten, setzen wir z. B. $\beta := \inf_{x_2 \in C_2} a^\top x_2$. □

Wir bereiten nun den strikten Trennungssatz vor.

Lemma 12.8 (Abgeschlossenheit der Minkowskisumme[28]).

Es seien $M_1, M_2 \subset \mathbb{R}^n$, $M_1, M_2 \neq \emptyset$ und M_1 abgeschlossen, M_2 kompakt. Dann ist die Minkowski-Summe $M_1 + M_2$ nicht leer und abgeschlossen.

Beachte: Ohne die Voraussetzung der Kompaktheit von M_2 gilt das nicht![29]

Beweis: $M := M_1 + M_2$ ist nicht leer. Es sei $\{z^{(k)}\} \subset M$ eine konvergente Folge mit Grenzwert z. Es existieren $\{x^{(k)}\} \subset M_1$ und $\{y^{(k)}\} \subset M_2$ mit $z^{(k)} = x^{(k)} + y^{(k)}$.

$$M_2 \text{ ist beschränkt} \quad \Rightarrow \quad \{y^{(k)}\} \text{ ist beschränkt} \quad \Rightarrow \quad \{x^{(k)}\} \text{ ist beschränkt.}$$

Es existieren also konvergente Teilfolgen $x^{(k_\ell)} \to x$ und $y^{(k_\ell)} \to y$, sodass auch $z^{(k_\ell)} = x^{(k_\ell)} + y^{(k_\ell)} \to x + y$ für $\ell \to \infty$. M_1 und M_2 sind abgeschlossen, also liegt $x + y \in M_1 + M_2 = M$. Andererseits konvergiert $z^{(k_\ell)}$ auch gegen z, also gilt $z = x + y \in M$, d. h., M ist abgeschlossen. □ Ende 17. V

Satz 12.9 (Strikter Trennungssatz[30]).

Es seien $C_1, C_2 \subset \mathbb{R}^n$ konvex, $C_1, C_2 \neq \emptyset$ und C_1 abgeschlossen, C_2 kompakt und zusätzlich $C_1 \cap C_2 = \emptyset$. Dann existieren $a \in \mathbb{R}^n$, $a \neq 0$, und $\beta \in \mathbb{R}$ mit

$$a^\top x_1 < \beta < a^\top x_2 \quad \text{für alle } x_1 \in C_1,\ x_2 \in C_2. \tag{12.6}$$

[28] Konvexität ist hier nicht entscheidend.

[29] Übung 10, Hausaufgabe 34

[30] Damit beweist man z. B., dass sich abgeschlossene konvexe Mengen als Schnitt abgeschlossener Halbräume darstellen lassen, siehe Übung 10, Aufgabe 33. Die Annahme der Existenz innerer Punkte, was wir in Lemma 12.6 und damit auch in Satz 12.7 zur Vereinfachung gemacht haben, bringt hier keine Vereinfachung.

Beachte: (12.6) bedeutet, dass es eine Hyperebene mit Normalenvektor a gibt, sodass C_1 ganz im einen und C_2 ganz im anderen *offenen* Halbraum enthalten ist.[31]

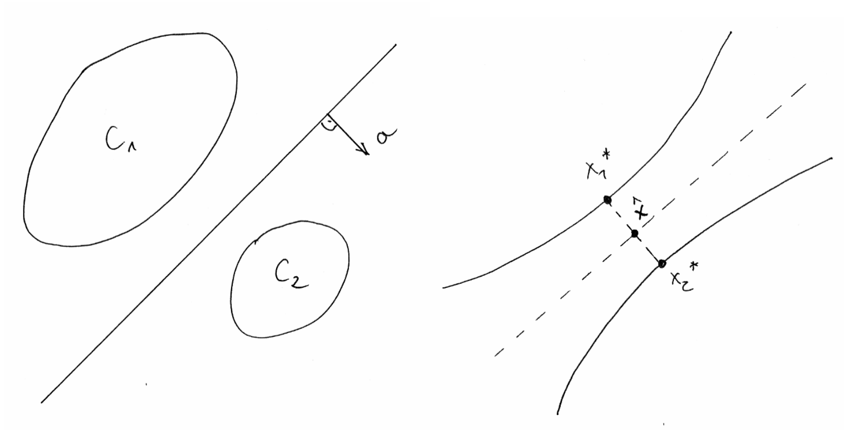


ABBILDUNG 12.5. Strikte Trennung konvexer Mengen durch eine Hyperebene und Lage von x_1^*, x_2^* und $\widehat{x}$

Beweis: Betrachte

$$\text{Minimiere} \quad \|(x_1 - x_2) - 0\|, \quad x_1 \in C_1,\ x_2 \in C_2. \tag{$*$}$$

Die Menge $C := C_1 - C_2$ ist nach Lemma 12.8 nicht leer, abgeschlossen und konvex. $(*)$ ist eine Projektionsaufgabe (12.1) mit $y = 0$ (Projektion der Null auf C). Es sei $(x_1^*, x_2^*) \in C_1 \times C_2$ die eindeutige Lösung von $(*)$ (Lemma 12.1). Setze

$$a := \frac{x_2^* - x_1^*}{2} \neq 0, \quad \widehat{x} := \frac{x_1^* + x_2^*}{2}, \quad \beta := a^\top \widehat{x}.$$

Behauptung: Es gilt

$$x_1^* = \operatorname{proj}_{C_1}(\widehat{x}), \quad x_2^* = \operatorname{proj}_{C_2}(\widehat{x}). \tag{$**$}$$

Beweis: Für beliebige $x_1 \in C_1$ und $x_2 \in C_2$ gilt

$$\begin{aligned}
\|x_1^* - \widehat{x}\| + \|\widehat{x} - x_2^*\| &= \|x_1^* - \widehat{x} + (\widehat{x} - x_2^*)\| && \text{denn } x_1^* - \widehat{x} = \widehat{x} - x_2^* \\
&= \|x_1^* - x_2^*\| \\
&\le \|x_1 - x_2\| && \text{denn } (x_1^*, x_2^*) \text{ ist optimal für } (*) \\
&\le \|x_1 - \widehat{x}\| + \|\widehat{x} - x_2\| && \text{Dreiecksungleichung.}
\end{aligned}$$

$$\begin{aligned}
&\text{Setze } x_2 = x_2^* && \Rightarrow && \|x_1^* - \widehat{x}\| \le \|x_1 - \widehat{x}\| && \text{für alle } x_1 \in C_1, \\
&\text{Setze } x_1 = x_1^* && \Rightarrow && \|x_2^* - \widehat{x}\| \le \|x_2 - \widehat{x}\| && \text{für alle } x_2 \in C_2,
\end{aligned}$$

d. h., es gilt $(**)$.

Aus dem Projektionssatz 12.2 folgt daher

$$\underbrace{(x_1^* - \widehat{x})}_{=-a}{}^\top (x_1 - x_1^*) \ge 0 \quad \text{für alle } x_1 \in C_1.$$

$$\Rightarrow \quad a^\top x_1 \le a^\top x_1^* = a^\top \widehat{x} + a^\top (x_1^* - \widehat{x}) = \beta - \|a\|^2 < \beta \quad \text{für alle } x_1 \in C_1.$$

Analog: $a^\top x_2 \ge \beta + \|a\|^2 > \beta$ für alle $x_2 \in C_2$. □

[31] In Übung 10, Aufgabe 32 soll eine strikt trennende Hyperebene gefunden werden.

§ 13 Das Subdifferential und die Richtungsableitung konvexer Funktionen

Literatur: [Geiger and Kanzow, 2002, Kapitel 6.3], [Alt, 2004, Kapitel 2.4–2.8]

Ziel: Verallgemeinerung der Ableitung für nicht-glatte konvexe Funktionen

In diesem Abschnitt sei $C \subset \mathbb{R}^n$ konvex und $f : C \to \mathbb{R}$ eine konvexe Funktion.

Definition 13.1 (Subdifferential)**.**
Es sei $x_0 \in C$.

(a) Ein Vektor $s \in \mathbb{R}^n$ heißt ein **Subgradient**[32] von f in x_0, wenn gilt:

$$f(x) \ge f(x_0) + s^\top (x - x_0) \quad \text{für alle } x \in C. \tag{13.1}$$

Man sagt: Die rechte Seite in (13.1) ist eine **lineare Minorante**[33] (mit Gradient s), die f in x_0 **stützt**, kurz: eine **lineare Stützfunktion**.[34]

(b) Die Menge $\partial f(x_0)$ *aller* Subgradienten im Punkt x_0 heißt das **Subdifferential** von f in x_0.[35]

Beachte: (13.1) verallgemeinert die Ungleichung (11.2), die für konvexe *diffbare* Funktionen gilt.

Beachte: Das Subdifferential hängt vom Definitionsgebiet C der Funktion ab![36]

Satz 13.2 (Existenz und Eigenschaften des Subdifferentials)**.**
Es sei $x_0 \in \operatorname{int}(C)$. Dann ist $\partial f(x_0) \neq \emptyset$.

Beweis: Betrachte den Epigraphen

$$\operatorname{epi}(f) = \{(x, \alpha) \in C \times \mathbb{R} : \alpha \ge f(x)\}.$$

---

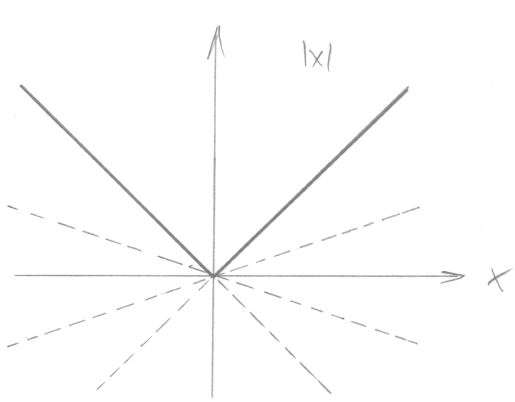
[32]Ein besserer Begriff wäre eigentlich **Subableitung**.

[33]In Aufgabe (noch nicht gestellt): `Opt_aufgabe105v5` wird die Menge aller linearen Minoranten an einem Beispiel betrachtet.

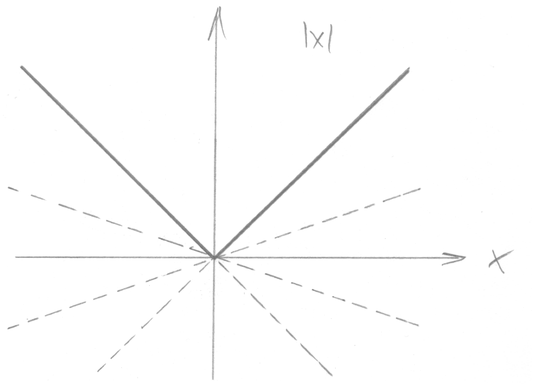
[34]Eine andere Anschauung: $(s, -1)$ ist Normalenvektor einer Hyperebene durch den Punkt $(x_0, f(x_0))$, die den (Epi-)Graphen in diesem Punkt stützt.

[35]**Beachte:** Das Subdifferential hängt vom Definitionsbereich der Funktion ab.
In Aufgabe (noch nicht gestellt): `Opt_aufgabe235v2` wird gezeigt, dass das Subdifferential einer konvexer Funktionen ein monotoner Operator ist. (Es ist sogar unter bestimmten Voraussetzungen maximal monoton.) Subdifferentiale einiger Normen werden in Aufgabe (noch nicht gestellt): `Opt_aufgabe082v5` berechnet. In Aufgabe (noch nicht gestellt): `Opt_aufgabe082v6` wird eine Kettenregel für das Subdifferential bewiesen.

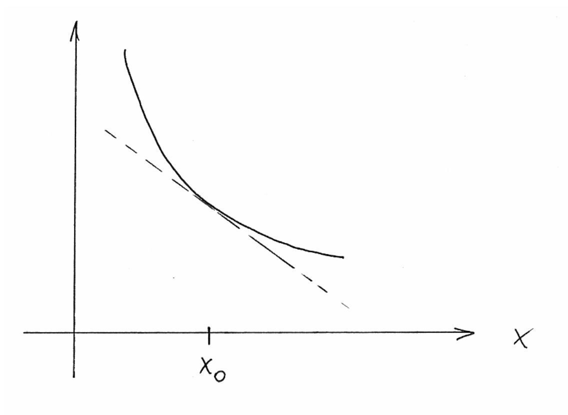
[36]Bei erweiterten konvexen Funktionen fordert man dann natürlich die Ungleichung (13.1) auf ganz $\mathbb{R}^n$.



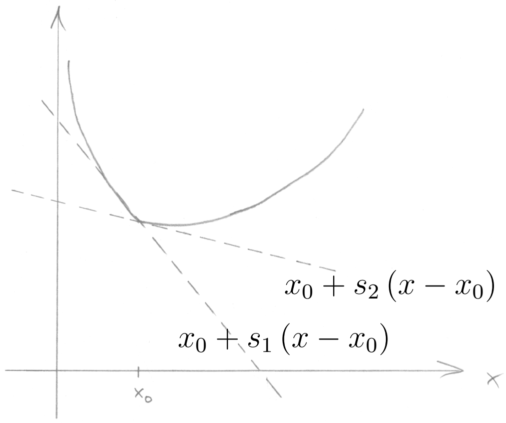
$f(x) = |x|$, $C = \mathbb{R}$, $\quad \partial f(0) = [-1, 1]$



$f(x) = |x|$, $C = [0, \infty)$, $\quad \partial f(0) = (-\infty, 1]$



$C = \mathbb{R}$, $\quad \partial f(x_0) = \{\nabla f(x_0)\} \overset{1D}{=} \{f'(x_0)\}$



$C = \mathbb{R}$, $\quad \partial f(x_0) = [s_1, s_2]$

ABBILDUNG 13.1. Das Subdifferential besteht aus den *Steigungen* aller linearen Minoranten an f im Punkt x_0.

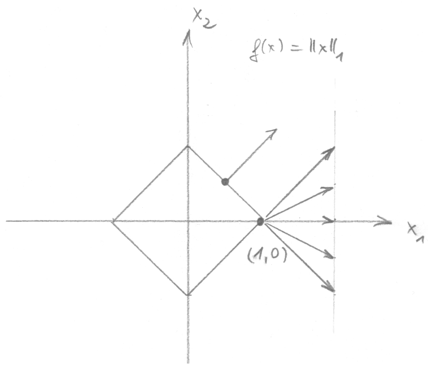
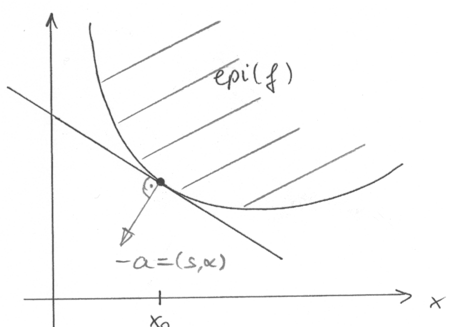


ABBILDUNG 13.2. Das Subdifferential der 1-Norm $\|x\|_1 = |x_1| + |x_2|$



Nach Satz 11.10 ist $\operatorname{epi}(f) \subset \mathbb{R}^n \times \mathbb{R}$ konvex. Es ist $(x_0, f(x_0)) \notin \operatorname{int}(\operatorname{epi}(f))$, sonst gäbe es einen Punkt $(x_0, f(x_0) - \varepsilon) \in \operatorname{epi}(f)$, Widerspruch zur Definition von $\operatorname{epi}(f)$. Nach Lemma 12.6 [Trennung von Punkt und konvexer Menge] existiert $a = -(s, \alpha) \in \mathbb{R}^n \times \mathbb{R}$, $(s, \alpha) \neq 0$ mit

$$\begin{pmatrix} s \\ \alpha \end{pmatrix}^\top \begin{pmatrix} x \\ y \end{pmatrix} \leq \begin{pmatrix} s \\ \alpha \end{pmatrix}^\top \begin{pmatrix} x_0 \\ f(x_0) \end{pmatrix} \quad \text{für alle } (x, y) \in \operatorname{epi}(f). \tag{$*$}$$

<u>Behauptung:</u> Es gilt $\alpha \leq 0$, denn:

Fall (i): Ist $f(x_0) = 0$, wähle $x = x_0$, $y = 1$

$$\overset{(*)}{\Rightarrow} \quad s^\top x_0 + \alpha \leq s^\top x_0 + 0 \quad \Rightarrow \quad \alpha \leq 0.$$

Fall (ii): Ist $f(x_0) > 0$, wähle $x = x_0$, $y = 2\, f(x_0)$

$$\overset{(*)}{\Rightarrow} \quad s^\top x_0 + 2\,\alpha\, f(x_0) \leq s^\top x_0 + \alpha\, f(x_0) \quad \Rightarrow \quad \alpha \leq 0.$$

Fall (iii): ist $f(x_0) < 0$, wähle $x = x_0$, $y = 0$

$$\overset{(*)}{\Rightarrow} \quad s^\top x_0 + 0 \leq s^\top x_0 + \alpha\, f(x_0) \quad \Rightarrow \quad \alpha \leq 0.$$

<u>Behauptung:</u> Es gilt sogar $\alpha < 0$, denn:

$$x_0 \in \operatorname{int}(C) \quad \Rightarrow \quad x_0 + \varepsilon\, s \in C \quad \text{für ein } \varepsilon > 0.$$

Setze in $(*)$ $x = x_0 + \varepsilon\, s$ und $y = f(x)$.

$$\begin{aligned} &\Rightarrow \quad s^\top x + \alpha\, y = s^\top x_0 + \varepsilon\, \|s\|^2 + \alpha f(x_0 + \varepsilon\, s) \overset{(*)}{\leq} s^\top x_0 + \alpha\, f(x_0) \\ &\Rightarrow \quad \varepsilon\, \|s\|^2 \leq \alpha\, [f(x_0) - f(x_0 + \varepsilon\, s)]. \end{aligned}$$

Aus $\alpha = 0$ würde $s = 0$ folgen, Widerspruch zu $(s, \alpha) \neq 0$.

Wir können O. B. d. A. $\alpha = -1$ annehmen, dann ergibt $(*)$ mit $y = f(x)$

$$\begin{aligned} & s^\top x - f(x) \leq s^\top x_0 - f(x_0) \quad \text{für alle } x \in C \\ \Rightarrow \quad & f(x) \geq f(x_0) + s^\top (x - x_0) \quad \text{für alle } x \in C, \end{aligned}$$

d. h., $s \in \partial f(x_0)$. □

Beispiel 13.3 (Leeres Subdifferential)**.**

Am Rand der Menge C kann das Subdifferential leer sein. Betrachte z. B. die konvexe (unstetige) Funktion

$$f(x) := \begin{cases} 1 & x = 0 \\ x & x > 0 \end{cases}$$

auf $C = [0, 1]$.

Frage: Gibt es einen Zusammenhang mit der Richtungsableitung? (Satz 13.10)

Definition 13.4 ((Einseitige) Richtungsableitung)**.**

Es sei $M \subset \mathbb{R}^n$, $g : M \to \mathbb{R}$ (hier nicht notwendig konvex), $x_0 \in M$ und $d \in \mathbb{R}^n$. Weiter gebe es ein $\delta > 0$, sodass $x_0 + t\, d \in M$ liegt für alle $t \in (0, \delta)$. Dann heißt der Grenzwert (sofern er in $\mathbb{R}$ existiert)

$$\delta g(x_0; d) := \lim_{t \searrow 0} \frac{g(x_0 + t\, d) - g(x_0)}{t} \tag{13.2}$$

die **(einseitige) Richtungsableitung** der Funktion g im Punkt x_0 in Richtung d.[37]

Wir führen die Abkürzung

$$q(t) := \frac{f(x_0 + t\,d) - f(x_0)}{t}, \quad t > 0. \tag{13.3}$$

für den Differenzenquotienten unserer konvexen Funktion $f : C \to \mathbb{R}$ ein.

Lemma 13.5 (Monotonie des Differenzenquotienten)**.**
Es seien $x_0 \in C$ und $x_0 + d \in C$. Dann ist $q(t)$ monoton wachsend auf $(0, 1]$.

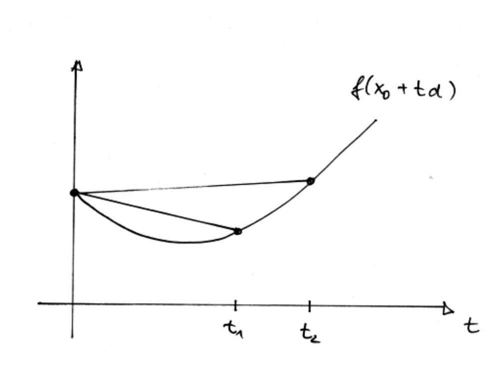
Ende 18. V

Beweis: Es sei $0 < t_1 < t_2 \leq 1$. Es gilt $x_0 + t_1\,d \in C$, $x_0 + t_2\,d \in C$ und

$$x_0 + t_1\,d = \tfrac{t_1}{t_2}(x_0 + t_2\,d) + (1 - \tfrac{t_1}{t_2})\,x_0,$$

und C ist konvex, also ist $x_0 + t_1\,d \in C$. Weiter gilt

$$\begin{aligned} f(x_0 + t_1\,d) &= f\big(\tfrac{t_1}{t_2}(x_0 + t_2\,d) + (1 - \tfrac{t_1}{t_2})\,x_0\big) \\ &\leq \frac{t_1}{t_2} f(x_0 + t_2\,d) + (1 - \frac{t_1}{t_2}) f(x_0) \\ \Rightarrow \quad f(x_0 + t_1\,d) - f(x_0) &\leq \frac{t_1}{t_2} f(x_0 + t_2\,d) - \frac{t_1}{t_2} f(x_0) \\ \Rightarrow \quad q(t_1) &\leq q(t_2). \end{aligned}$$



$\square$

Lemma 13.6 (Lokale Beschränktheit und lokale Lipschitz-Stetigkeit)**.**
Es sei $x_0 \in \operatorname{int}(C)$. Dann gibt es ein $\delta > 0$ mit $U_\delta(x_0) \subset C$, sodass gilt:

(a) Es existieren $m, M \in \mathbb{R}$ mit $f(x) \in [m, M]$ für alle $x \in U_\delta(x_0)$.

(b) Es existiert $L(x_0) \geq 0$ mit

$$|f(x) - f(y)| \leq L(x_0)\,\|x - y\| \quad \text{für alle } x, y \in U_{\delta/2}(x_0).$$

Beachte: Konvexe Funktionen sind im Inneren ihres Definitionsbereiches also **lokal beschränkt** und **lokal Lipschitz-stetig**.

Beweis: Es sei $x_0 \in \operatorname{int}(C)$.
(a): Nach Lemma 12.5 existieren Vektoren $v_0, \ldots, v_n \in C$ und $\delta > 0$, sodass mit $\Delta = \operatorname{conv}(\{v_0, \ldots, v_n\})$ gilt: $U_\delta(x_0) \subset \operatorname{int}(\Delta) \subset C$ gilt.

[37] Die Richtungsableitung wird manchmal auch mit $g'(x_0; d)$ oder $Dg(x_0; d)$ bezeichnet.