Konvexe Mengen und konvexe Funktionen

Teilnehmer:

Moritz Butz	Herder-Gymnasium
Franziska Ihlefeldt	Herder-Gymnasium
Johannes Jendersie	Georg-Forster-Gymnasium
Marie Lambert	Georg-Forster-Gymnasium
Eike Müller	Herder-Gymnasium
Gregor Pasemann	Georg-Forster-Gymnasium
Konstantin Rohde	Heinrich-Hertz-Gymnasium

Gruppenleiter:

Konrad Gröger	Humboldt-Universität zu Berlin

Im Rahmen des Projektes haben wir uns mit konvexen Mengen und konvexen Funktionen im Raum $\mathbb{R}^N$ beschäftigt. Innerhalb dieses Raumes haben wir Begriffe und Sachverhalte erörtert, die mit Konvexität im Zusammenhang stehen. Beispiele hierfür sind die konvexe Hülle einer beliebigen Teilmenge von $\mathbb{R}^N$, konvexe Kombinationen von Punkten aus $\mathbb{R}^N$, Polytope und Polyeder. Für konvexe Mengen haben wir einen Trennungssatz bewiesen und mit dessen Hilfe gezeigt, dass jede abgeschlossene konvexe Teilmenge von $\mathbb{R}^N$ der Durchschnitt von abgeschlossenen Halbräumen ist.
Nach der Betrachtung der konvexen Mengen haben wir uns konvexen Funktionen zugewandt. Mit Hilfe von Epigraphen konnten wir Fragestellungen für konvexe Funktionen auf die konvexen Mengen zurückführen. Außerdem haben wir das Subdifferential von konvexen Funktionen definiert und an konkreten Beispielen veranschaulicht. Das Subdifferential, eine mehrdeutige Abbildung, lässt sich als ein Ersatz für die Ableitung von nicht überall differenzierbaren konvexen Funktionen ansehen. Abschließend haben wir komplexe Optimierungsprobleme, mit denen man im Zusammenhang mit konvexen Mengen und Funktionen konfrontiert werden kann, angedeutet und deren Lösbarkeit nachgewiesen.

1 Konvexe Mengen

Der N-dimensionale Raum $\mathbb{R}^N$ ist definiert als das N-fache kartesische Produkt der reellen Zahlen mit sich selbst. Ein Punkt x dieses Raumes kann durch N Koordinaten x_1 bis x_N dargestellt werden:

$$x := (x_1, x_2, \ldots, x_N) \in \mathbb{R}^N.$$

Folgende Regeln sind für das Rechnen mit Punkten aus $\mathbb{R}^N$ erklärt:
Sind $x = (x_1, x_2, \ldots, x_N)$ und $y = (y_1, y_2, \ldots, y_N)$ zwei Punkte aus $\mathbb{R}^N$ und ist t eine reelle Zahl, so setzt man

$$x + y := (x_1 + y_1, x_2 + y_2, \ldots, x_N + y_N)$$

und

$$tx := (tx_1, tx_2, \ldots, tx_N).$$

Definition 1 *Eine Teilmenge $E \subset \mathbb{R}^N$ [1)] heißt **konvex**, wenn die Verbindungsstrecke von je zwei Punkten $x, y \in E$ zu E gehört.*

Eine Teilmenge $E \subset \mathbb{R}^1$ ist genau dann konvex, wenn E ein Intervall ist, denn die Verbindungsstrecke von zwei Punkten aus dem Intervall liegt auch wieder in dem Intervall. Beispiele für konvexe und nicht konvexe Teilmengen von $\mathbb{R}^2$ zeigt die Abbildung 1.

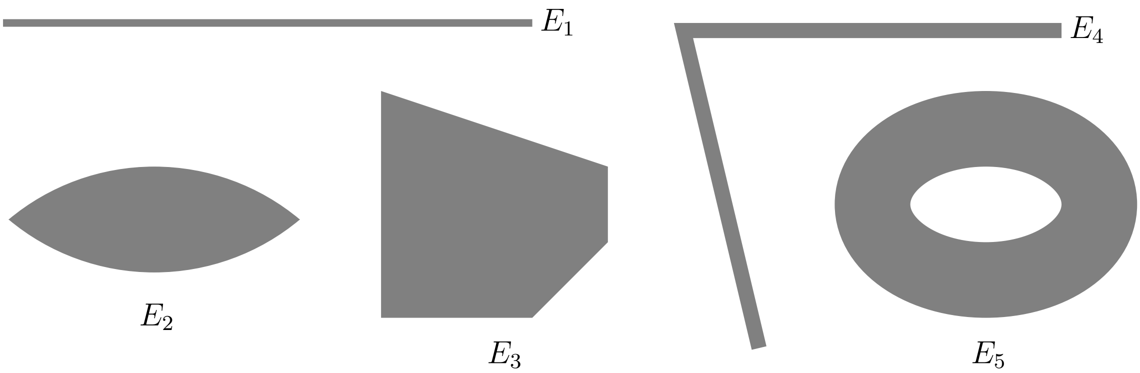

Abbildung 1: E_1, E_2, E_3 sind konvex, E_4, E_5 sind nicht konvex

Die leere Menge und alle einelementigen Mengen sind konvex, denn es existieren keine zwei Punkte in diesen Mengen, somit müssen diese Mengen keine Bedingung erfüllen, um konvex zu sein.

Satz 1 *Der Durchschnitt beliebig vieler konvexer Teilmengen aus $\mathbb{R}^N$ ist konvex.*

Beweis:
Liegen x und y in allen beteiligten konvexen Teilmengen, so liegt die Verbindungsstrecke auch in allen beteiligten konvexen Teilmengen und somit im Durchschnitt. □

[1)] $E \subset \mathbb{R}^N$ bedeutet, dass E eine Teilmenge von $\mathbb{R}^N$ ist. E kann dabei auch gleich $\mathbb{R}^N$ sein.

Satz 2 *Zu jeder Menge $E \subset \mathbb{R}^N$ existiert eine kleinste konvexe Menge, die E enthält.*

Beweis:
Nach Satz 1 ist der Durchschnitt aller konvexen Mengen, die E enthalten, wieder konvex. Dieser Durchschnitt ist gleich der kleinsten konvexen Menge, die E enthält. Gäbe es noch eine kleinere konvexe Menge, die E enthält, so wird mit ihr auch die Schnittmenge gebildet und somit kann sie nicht kleiner sein als der Durchschnitt. $\square$

Definition 2 *Die kleinste konvexe Menge, die E enthält, wird **konvexe Hülle** von E genannt und mit $\operatorname{conv}(E)$ bezeichnet. Eine Teilmenge von $\mathbb{R}^N$ wird **Polytop** genannt, wenn sie die konvexe Hülle einer endliche Teilmenge von $\mathbb{R}^N$ ist.*

Beispielsweise sind konvexe n-Ecke in $\mathbb{R}^2$ Polytope, ebenso Quader, Tetraeder und Oktaeder in $\mathbb{R}^3$. Dagegen sind Kreisscheiben und Kreiskegel zwar konvex, aber keine Polytope.

Definition 3 *Man nennt einen Punkt $x \in \mathbb{R}^N$* **konvexe Kombination** *der Punkte $v_1, \dots, v_n$ aus $\mathbb{R}^N$, wenn nicht negative Zahlen $t_1, \dots, t_n$ mit der Eigenschaft $\sum_{i=1}^{n} t_i = 1$ existieren, sodass $x = \sum_{i=1}^{n} t_i v_i$ gilt.*

Satz 3 *Ist E eine konvexe Teilmenge von $\mathbb{R}^N$, so gehören zu E alle konvexen Kombinationen von je endlich vielen Punkten $v_1, \dots, v_n$ aus E.*

Beweis:
Der Beweis wird mit Hilfe der vollständigen Induktion geführt. Für $n = 1$ ist die Behauptung trivial. Nehmen wir nun an, dass die Behauptung für alle konvexen Kombinationen von $n-1$ Punkten gilt. Es sei

$$x = \sum_{i=1}^{n} t_i v_i, \quad t_i > 0, \quad \sum_{i=1}^{n} t_i = 1, \quad v_1, \dots, v_n \in E.$$

Dann gilt

$$x = (1 - t_n) \sum_{i=0}^{n-1} \frac{t_i v_i}{1 - t_n} + t_n v_n.$$

Es ist $\sum_{i=0}^{n-1} \frac{t_i}{1 - t_n} = 1$ und $\frac{t_i}{1 - t_n} > 0$. Nach der Induktionsannahme gehört $y := \sum_{i=1}^{n-1} \frac{t_i}{1 - t_n} v_i$ zu E. Da auch v_n zu E gehört und $x = (1 - t_n)y + t_n v_n$ Punkt der Verbindungsstrecke von y und v_n ist, gilt $x \in E$. Somit ist die Induktion abgeschlossen. $\square$

Satz 4 *Ist $E \subset \mathbb{R}^N$ beliebig, so ist die Menge K aller konvexen Kombinationen von je endlich vielen Punkten von E konvex.*

Beweis:
Es sei $x, y \in K$. Es gelte etwa

$$x = \sum_{i=1}^{n} r_i v_i, \quad r_i \geq 0, \quad \sum_{i=1}^{n} r_i = 1, \qquad y = \sum_{i=1}^{n} s_i v_i, \quad s_i \geq 0, \quad \sum_{i=1}^{n} s_i = 1;$$

dabei seien $v_1, \ldots, v_n$ Punkte von E. Wir können bei beiden Punkten x und y deshalb mit den gleichen v_i arbeiten, weil man auch "überflüssige" Punkte mit dem Koeffizienten 0 mitnehmen kann. Dann ist für $t \in [0,1]$

$$(1-t)x + ty = \sum_{i=1}^{n} ((1-t)r_i + ts_i)v_i = \sum_{i=1}^{n} t_i v_i \text{ für } t_i := (1-t)r_i + ts_i.$$

Offensichtlich ist $t_i \geq 0$ und $\sum\limits_{i=1}^{n} t_i = (1-t) \sum\limits_{i=1}^{n} r_i + t \sum\limits_{i=1}^{n} s_i = 1 - t + t = 1$. Folglich gehören die Punkte $(1-t)x + ty$ für $t \in [0,1]$ zu K, d. h., K ist konvex. $\square$

Folgerung 1 *Die konvexe Hülle* $\operatorname{conv}(E)$ *einer beliebigen Menge* $E \subset \mathbb{R}^N$ *ist gleich der Menge alle konvexen Kombinationen von Punkten aus* E.

Beweis:
Nach Satz 3 muss $\operatorname{conv}(E)$ als konvexe Menge die Menge K der konvexen Kombinationen von Punkten aus E enthalten. Nach Satz 4 ist K konvex und daher $\operatorname{conv}(E)$ als kleinste konvexe Menge, die E enthält, in K enthalten.

Für $x = (x_1, x_2, \ldots, x_N)$, $y = (y_1, y_2, \ldots, y_N) \in \mathbb{R}^N$ definiert man das **Skalarprodukt** $x \cdot y$ und den **Betrag** $|x|$ durch die Vorschriften

$$x \cdot y := \sum_{i=1}^{N} x_i y_i, \quad |x| := \sqrt{x \cdot x} = \left(\sum_{i=1}^{N} x_i^2 \right)^{\frac{1}{2}}. \tag{1}$$

Der Betrag $|x|$ ist der Abstand des Punktes x vom Nullpunkt bzw. die Länge des Vektors x. Für $N = 1, 2, 3$ ist das mit dem Satz des Pythagoras zu begründen, für größere N ist das als *Definition* des Abstands anzusehen.

Das Skalarprodukt $x \cdot y$ ist symmetrisch in x und y, d. h., es gilt $y \cdot x = x \cdot y$. Der Wert $x \cdot y$ hängt sowohl von x als auch von y linear ab, d. h., es ist

$$(x + \xi) \cdot y = x \cdot y + \xi \cdot y, \quad x \cdot (y + \eta) = x \cdot y + x \cdot \eta, \quad (tx) \cdot y = x \cdot (ty) = t(x \cdot y) \tag{2}$$

für beliebige $x, y, \xi, \eta \in \mathbb{R}^N$ und $t \in \mathbb{R}$. Den Nullpunkt von $\mathbb{R}^N$ bezeichnen wir wie die Zahl Null mit 0. Welche Bedeutung das Symbol 0 jeweils hat, wird aus dem Zusammenhang hervorgehen. Gilt für $x, y \in \mathbb{R}^N$ die Beziehung $x \cdot y = 0$, so ist

$$|y - x|^2 = (y - x) \cdot (y - x) = y \cdot y - y \cdot x - x \cdot y + x \cdot x = |y|^2 + |x|^2,$$

d. h., in dem Dreieck, das von x, $(y - x)$ und y gebildet wird, gilt ein Analogon des Satzes von Pythagoras. Deshalb werden Vektoren $x, y \in \mathbb{R}^N$, für die $x \cdot y = 0$ ist, zueinander senkrecht oder zueinander orthogonal genannt.

Hilfssatz *Für beliebige $x, y \in \mathbb{R}^N$ gilt*

$$x \cdot y \leq |x||y|. \tag{3}$$

Beweis:
Für $y = 0$ ist die Behauptung (3) trivial. Es sei nun $y \neq 0$ und $t := \dfrac{x \cdot y}{|y|^2}$. Dann ist

$$0 \leq |x - ty|^2 = (x - ty) \cdot (x - ty) = x \cdot x - 2t\, x \cdot y + t^2 y \cdot y = |x|^2 - 2\frac{(x \cdot y)^2}{|y|^2} + \frac{(x \cdot y)^2}{|y|^2}.$$

Folglich ist $0 \leq |x|^2|y|^2 - (x \cdot y)^2$, d. h., es gilt die Ungleichung (3). $\square$

Die Beziehung (3) nennt man **Cauchy-Schwarzsche Ungleichung**.

Für beliebige $x, y \in \mathbb{R}^N$ ergibt sich unter Benutzung der Cauchy-Schwarzschen Ungleichung $|x + y|^2 = (x + y) \cdot (x + y) = |x|^2 + 2x \cdot y + |y|^2 \leq (|x| + |y|)^2$. Folglich ist

$$|x + y| \leq |x| + |y| \quad \text{für beliebige } x, y \in \mathbb{R}^N. \tag{4}$$

Die Ungleichung (4) bezeichnet man als **Dreiecksungleichung**.

Die Menge

$$B_r(\xi) = \left\{a \in \mathbb{R}^N; |a - \xi| < r\right\} \tag{5}$$

wird **Kugel** in $\mathbb{R}^N$ genannt mit dem Mittelpunkt ξ und dem Radius r. Existiert zu einem Punkt $\xi \in E$ ein $r > 0$, sodass $B_r(\xi) \subset E$ gilt, so wird ξ **innerer Punkt** von E genannt.

Es sei $x_n = (x_{n1}, x_{n2}, \ldots, x_{nN})$ für $n \in \mathbb{N}$ eine Folge von Punkten im $\mathbb{R}^N$. Die Folge (x_n) heißt **konvergent** gegen $\xi = (\xi_1, \ldots, \xi_N)$, falls $\lim\limits_{n\to\infty} |x_n - \xi| = 0$ gilt. Dies ist der Fall, wenn die Folge koordinatenweise konvergiert. Es muss somit für alle $i = 1, \ldots, N$ gelten $\lim\limits_{n\to\infty} |x_{ni} - \xi_i| = 0$.

Die Folge (x_n) heißt **Cauchyfolge**, wenn $\lim\limits_{m,n\to\infty} |x_m - x_n| = 0$ ist. Jede Cauchyfolge in $\mathbb{R}^N$ besitzt einen Grenzwert. Das folgt daraus, dass eine Folge (x_n) im $\mathbb{R}^N$ genau dann Cauchyfolge ist, wenn die Koordinatenfolgen (x_{ni}) $(i = 1, \ldots, N)$ Cauchyfolgen reeller Zahlen sind.

Definition 4 *Eine Menge $E \subset \mathbb{R}^N$ heißt* ***abgeschlossen****, wenn für jede konvergente Folge (x_n) aus E deren Grenzwert $\lim\limits_{n\to\infty} x_n$ auch in E liegt. Eine Menge $E \subset \mathbb{R}^N$ heißt* ***offen****, wenn jeder Punkt $\xi \in E$ innerer Punkt von E ist.*

Der Durchschnitt beliebig vieler abgeschlossener Mengen in $\mathbb{R}^N$ ist abgeschlossen. Zu jeder Menge $E \subset \mathbb{R}^N$ gibt es eine kleinste abgeschlossene Teilmenge und eine kleinste abgeschlossene, konvexe Teilmenge von $\mathbb{R}^N$, die E enthält. Die erste nennt man die **abgeschlossene Hülle** von E, und man bezeichnet sie meist mit $\overline{E}$. Die zweite nennt man die **abgeschlossene konvexe Hülle** von E, und man bezeichnet sie meist mit $\overline{\text{conv}}(E)$. Zur Begründung kann man wie bei der konvexen Hülle argumentieren (vgl. die Aussagen und Beweise von Satz 1 und Satz 2).

Man kann beweisen, dass die abgeschlossene Hülle einer konvexen Menge $E \subset \mathbb{R}^N$ ebenfalls konvex ist. Polytope sind stets abgeschlossen. Das kann man mit Hilfe von Folgerung 1 beweisen.

Wir kommen zu weiteren Beispielen für konvexe Mengen. Ist $K \subset \mathbb{R}^N$ eine konvexe, nicht leere Menge und $r > 0$, so sind auch folgende Mengen konvex:

$$\begin{aligned} U_r(K) &:= \left\{\xi \in \mathbb{R}^N; |\xi - x| < r \text{ für ein } x \in \mathbb{R}^N\right\} \\ \overline{U_r}(K) &:= \left\{\xi \in \mathbb{R}^N; |\xi - x| \leq r \text{ für ein } x \in \mathbb{R}^N\right\} \end{aligned}$$

Begründung:
Es sei $\xi \in U_r(K)$ und $\eta \in U_r(K)$. Wir wählen x und y in K so, dass $|\xi - x| < r$ und $|\eta - y| < r$ gilt. Solche Punkte x und y müssen nach der Definition von $U_r(K)$ existieren. Für $t \in [0, 1]$ ist dann $z := (1 - t)x + ty \in K$ und für $\zeta := (1 - t)\xi + t\eta$ gilt

$$|\zeta - z| = |(1-t)(\xi - x) + t(\eta - y)| \leq (1-t)|\xi - x| + t|\eta - y| < (1-t)r + tr = r. \quad (6)$$

Folglich gehört ζ zu $U_r(K)$. Mit zwei Punkten ξ, η liegt also auch deren Verbindungsstrecke in $U_r(K)$, d. h., $U_r(K)$ ist konvex. Für $\overline{U}_r(K)$ argumentiert man analog. □

Diese Mengen werden **Umgebungen** von K genannt.

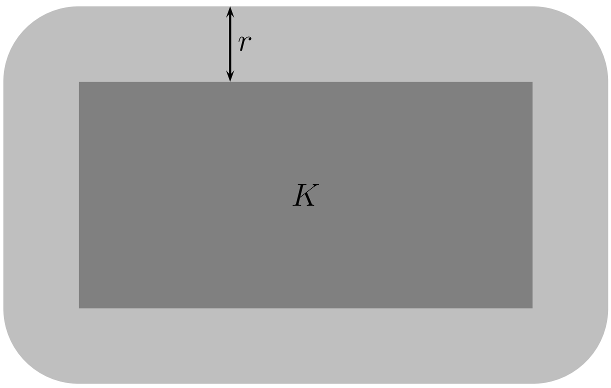

Abbildung 2: Umgebung einer konvexen Menge K

Weil eine aus einem Punkt bestehende Menge K konvex ist, zeigt die Feststellung insbesondere, dass offene und abgeschlossene Kugeln konvex sind.

Es gibt noch weitere spezielle konvexe Mengen. Seien $a \in \mathbb{R}^N \setminus \{0\}$ und $z \in \mathbb{R}^N$ beliebig aber fest. Die Menge

$$H_{a,z} := \left\{x \in \mathbb{R}^N; a \cdot (x - z) = 0\right\}$$

besteht aus den Punkten x, für die der Vektor $(x - z)$ senkrecht zu a steht. Im $\mathbb{R}^2$ ist dies somit eine Gerade und im $\mathbb{R}^3$ eine Ebene. Im $\mathbb{R}^N$ wird eine solche Menge **Hyperebene** genannt.

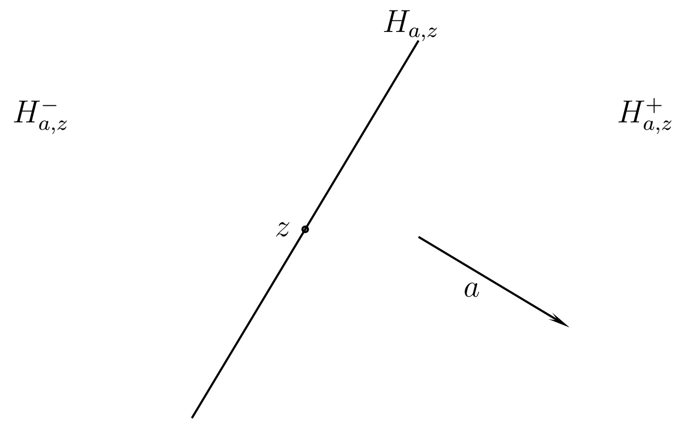

Abbildung 3: Hyperebene und Halbräume für $N = 2$

Die Hyperebene $H_{a,z}$ teilt den $\mathbb{R}^N$ in die Mengen

$$\begin{aligned} H_{a,z}^+ &:= \left\{x \in \mathbb{R}^N ; a \cdot (x - z) > 0\right\} \\ H_{a,z}^- &:= \left\{x \in \mathbb{R}^N ; a \cdot (x - z) < 0\right\} \end{aligned}$$

auf, die **Halbräume** genannt werden.

Für beliebige $x, y \in \mathbb{R}^N$ und $t \in [0, 1]$ gilt:

$$a \cdot (tx + (1-t)\, y - z) = a \cdot (t\,(x - z) + (1-t)\,(y - z)) = ta \cdot (x - z) + (1-t)\, a \cdot (y - z)$$

Daher sind $H_{a,z}$, $H_{a,z}^+$ und $H_{a,z}^-$ konvexe Mengen. Die Hyperebene $H_{a,z}$ ist abgeschlossen und sowohl $H_{a,z}^+$ als auch $H_{a,z}^-$ sind offen. Die abgeschlossene Hülle von $H_{a,z}^+$ und $H_{a,z}^-$ erhält man, indem man die Hyperebene $H_{a,z}$ hinzunimmt.

Satz 5 *Es sei $K \subset \mathbb{R}^N$ konvex, abgeschlossen und nicht leer. Dann gibt es zu jedem Punkt $\xi \in \mathbb{R}^N$ genau einen Punkt $x \in K$, der ξ am nächsten liegt, für den also die Beziehung*

$$|\xi - x| < |\xi - y| \quad \text{für alle } y \in K \backslash \{x\}$$

gilt. Dieser Punkt x wird auch durch die Beziehung

$$x \in K, \quad (\xi - x) \cdot (y - x) \leq 0 \quad \text{für alle } y \in K \tag{7}$$

charakterisiert.

Beweis:
1. Für beliebige Punkte $x, y \in \mathbb{R}^N$ gilt

$$\begin{aligned} |x + y|^2 + |x - y|^2 &= (x + y) \cdot (x + y) + (x - y) \cdot (x - y) \\ &= x \cdot x + 2x \cdot y + y \cdot y + x \cdot x - 2x \cdot y + y \cdot y \\ &= 2|x|^2 + 2|y|^2. \end{aligned} \tag{8}$$

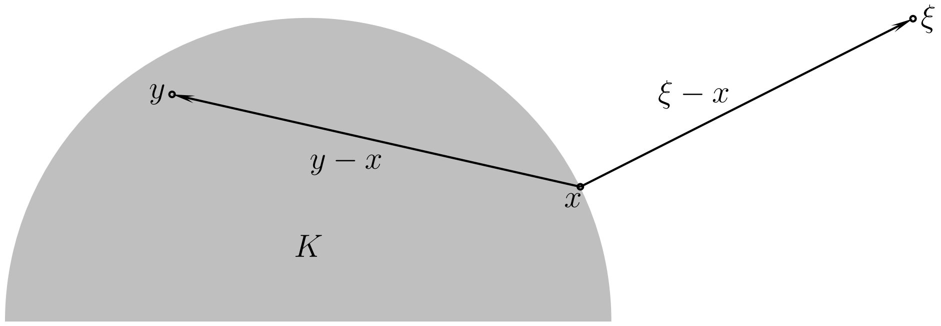

Abbildung 4: Konvexe Projektion

Wir setzen $d := \inf_{x \in K} |\xi - x|$ und wählen eine Folge (x_n) in K so, dass $\lim_{n\to\infty} |\xi - x_n| = d$ ist. Unter Verwendung der Formel (8) für $(\xi - x_n)$ und $(\xi - x_m)$ anstatt von x und y ergibt sich

$$\begin{aligned}
|x_m - x_n|^2 &= |\,(\xi - x_n) - (\xi - x_m)\,|^2 \\
&= 2|\xi - x_n|^2 + 2|\xi - x_m|^2 - |2\xi - (x_n + x_m)|^2 \\
&= 2|\xi - x_n|^2 + 2|\xi - x_m|^2 - 4\left|\xi - \frac{x_n + x_m}{2}\right|^2 \\
&\leq 2|\xi - x_n|^2 + 2|\xi - x_m|^2 - 4d^2.
\end{aligned}$$

Daher gilt

$$\lim_{m,n\to\infty} |x_m - x_n|^2 \leq 2d^2 + 2d^2 - 4d^2 = 0.$$

Somit gilt auch $\lim_{m,n\to\infty} |x_m - x_n| = 0$, d. h., die Folge (x_n) ist eine Cauchyfolge in $\mathbb{R}^N$. Der Grenzwert dieser Folge sei x. Dann gilt

$$|\xi - x| = \lim_{n\to\infty} |\xi - x_n| = d.$$

2. Nun muss noch gezeigt werden, dass für den Punkt x die Beziehung (7) gilt. Ist $y \in K$, so gehören auch die Punkte $z_t := (1-t)x + ty,\ 0 \leq t \leq 1$, zu K. Wir setzen

$$\begin{aligned}
\varphi(t) &:= |\xi - z_t|^2 \\
&= |\xi - (1-t)x - ty|^2 \\
&= (\xi - (1-t)x - ty) \cdot (\xi - (1-t)x - ty) \\
&= (\xi - x + tx - ty) \cdot (\xi - x + tx - ty) \\
&= |\xi - x|^2 - 2t(\xi - x) \cdot (y - x) + t^2|y - x|^2.
\end{aligned}$$

Dann ist

$$\begin{aligned}
\varphi'(t) &= -2(\xi - x) \cdot (y - x) + 2t|y - x|^2, \\
\varphi'(0) &= -2(\xi - x) \cdot (y - x).
\end{aligned}$$

Wäre $(\xi - x) \cdot (y - x) > 0$, so wäre $\varphi'(0) < 0$, und $\varphi(t) = |\xi - z_t|^2$ wäre für kleine $t > 0$ echt kleiner als $\varphi(0) = |\xi - x|^2 = d^2$. Das kann nach Definition von d nicht sein. Der Widerspruch beweist die Beziehung (7).

3. Es gelte die Beziehung (7). Für beliebiges $y \in K$ ist dann

$$|\xi - y|^2 = |\xi - x + x - y|^2 = |\xi - x|^2 + 2(\xi - x) \cdot (x - y) + |x - y|^2 \geq |\xi - x|^2 + |x - y|^2.$$

Folglich ist $|\xi - x| < |\xi - y|$ für alle $y \in K \setminus \{x\}$. Der Punkt x liegt also echt näher an ξ als alle anderen Punkte $y \in K$. $\square$

Wir definieren die Abbildung $P_K : \mathbb{R}^N \longrightarrow K$ durch die Vorschrift

$$P_K(\xi) := x, \quad \text{falls } x \in K,\ |\xi - x| < |\xi - y| \text{ für alle } y \in K \setminus \{x\}.$$

Die Abbildung P_K ordnet also jedem Punkt $\xi \in \mathbb{R}^N$ den Punkt x der konvexen Menge K zu, der am nächsten zu ξ liegt. Man nennt P_K die **konvexe Projektion** von $\mathbb{R}^N$ auf K. Die konvexe Projektion verallgemeinert die senkrechte Projektion auf eine Gerade bzw. Ebene.

Satz 6 *Es sei $K \subset \mathbb{R}^N$ konvex, abgeschlossen und nicht leer. Ferner sei ξ ein Punkt von $\mathbb{R}^N$, der nicht zu K gehört. Dann gibt es eine Hyperebene $H_{a,z}$ in $\mathbb{R}^N$, so dass $K \subset H^-_{a,z}$ und $\xi \in H^+_{a,z}$ gilt.*

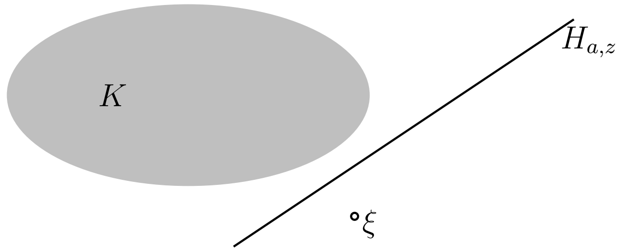

Abbildung 5: Trennung von konvexer Menge und Punkt

Beweis:
Es sei $x = P_K(\xi)$. Es ist $x \neq \xi$, weil ξ nicht zu K gehört. Wir setzen $a := \xi - x$ und $z := \dfrac{\xi + x}{2}$. Nach Wahl von a und z ist

$$a \cdot (\xi - z) = (\xi - x) \cdot \left(\xi - \frac{\xi + x}{2}\right) = \frac{1}{2}|\xi - x|^2 > 0.$$

Folglich gilt $\xi \in H^+_{a,z}$. Für $y \in K$ ergibt sich mit Hilfe von (7)

$$\begin{aligned} a \cdot (y - z) &= (\xi - x) \cdot \left(y - \frac{\xi + x}{2}\right) = (\xi - x) \cdot \left(y - x - \frac{\xi - x}{2}\right) \\ &= (\xi - x) \cdot (y - x) - \frac{1}{2}|\xi - x|^2 \leq -\frac{1}{2}|\xi - x|^2 < 0. \end{aligned}$$

Damit ist gezeigt, dass K in $H^-_{a,z}$ liegt. $\square$

Man sagt, die Hyperebene $H_{a,z}$ **trenne** den Punkt ξ und die konvexe Menge K.

Man kann den Satz verallgemeinern und statt ξ eine konvexe, abgeschlossene und beschränkte Menge L betrachten. Dann lässt sich eine Hyperebene finden, die K und L trennt. Der Beweis wird hier jedoch nicht angeführt.

Folgerung 2 *Jede konvexe abgeschlossene Teilmenge K von $\mathbb{R}^N$ ist der Durchschnitt von Halbräumen, die K enthalten.*

Beweis:
Ist $K = \mathbb{R}^N$, so kann man diese Menge als Durchschnitt von null Halbräumen auffassen. Ist $K \subsetneqq \mathbb{R}^N$, so sei D der Durchschnitt aller K enthaltenden abgeschlossenen Halbräume. Dann gilt $K \subset D$. Die Menge D kann aber nicht mehr Punkte enthalten als K. Ist nämlich $\xi \notin K$, so gibt es nach Satz 6 einen abgeschlossenen Halbraum, der K, aber nicht ξ enthält. $\square$

Als Anwendung stellen wir eine Variante des Mittelwertsatzes der Differentialrechnung für Funktionen auf $\mathbb{R}$ mit Werten in $\mathbb{R}^N$ vor. Es sei

$$f(t) = (f_1(t), \ldots, f_N(t)) \text{ für } t \in \mathbb{R}.$$

Man nennt f stetig differenzierbar, wenn $f_1, \ldots, f_N$ stetig differenzierbar sind, und setzt $f'(t) := (f_1'(t), \ldots, f_N'(t))$ für $t \in \mathbb{R}$.

Satz 7 *Es sei $f : \mathbb{R} \longrightarrow \mathbb{R}^N$ stetig differenzierbar und $t_0, t_1 \in \mathbb{R}$, $t_0 < t_1$. Dann gilt*

$$\frac{f(t_1) - f(t_0)}{t_1 - t_0} \in \overline{\operatorname{conv}}(E) \text{ für } E := \{f'(t);\ t_0 < t < t_1\}.$$

Beweis:
Wegen der Stetigkeit von f' ist E in $\mathbb{R}^N$ beschränkt. Es sei $\overline{H_{a,z}^-}$ ein abgeschlossener Halbraum in $\mathbb{R}^N$, der E enthält. Dann gilt

$$a \cdot (f'(t) - z) = \sum_{i=1}^{N} a_i(f_i'(t) - z_i) \leq 0 \text{ für alle } t \in]t_0, t_1[.$$

Integriert man diese Ungleichung von t_0 bis t_1 und dividiert durch $t_1 - t_0$, so findet man

$$a \cdot \Big(\frac{f(t_1) - f(t_0)}{t_1 - t_0} - z\Big) = \sum_{i=1}^{N} a_i \Big(\frac{f_i(t_1) - f_i(t_0)}{t_1 - z_0} - z_i\Big) \leq 0.$$

Diese Ungleichung zeigt, dass $\dfrac{f(t_1) - f(t_0)}{t_1 - t_0}$ ebenfalls zu $\overline{H_{a,z}^-}$ gehört. Aufgrund der Folgerung 2 ist $\overline{\operatorname{conv}}(E)$ der Durchschnitt aller E enthaltenden abgeschlossenen Halbräume. Folglich gilt $\dfrac{f(t_1) - f(t_0)}{t_1 - t_0} \in \overline{\operatorname{conv}}(E)$. $\square$

Man kann durch Beispiele für $N = 2$ zeigen, dass $\dfrac{f(t_1) - f(t_0)}{t_1 - t_0}$ im Allgemeinen nicht gleich einem der Werte $f'(t)$, $t_0 < t < t_1$, ist. Man muss deshalb mit weniger zufrieden sein. Ist $|f'(t)| \leq r$ für $t_0 < t < t_1$, so liegt E und damit auch $\overline{\text{conv}}(E)$ in der abgeschlossenen Kugel $\overline{B_r(0)}$. Es ist also $\left|\dfrac{f(t_1) - f(t_0)}{t_1 - t_0}\right| \leq r$, d. h., der Betrag des Differenzenquotienten lässt sich durch die Beträge der Ableitungen in den Zwischenpunkten abschätzen.

Zum Abschluss des Abschnitts über konvexe Mengen werden wir den Begriff Polyeder klären. Eine Teilmenge von $\mathbb{R}^N$ nennt man **polyedrisch**, wenn sie der Durchschnitt von *endlich* vielen abgeschlossenen Halbräumen ist. Unter einem **Polyeder** in $\mathbb{R}^N$ versteht man eine *beschränkte* polyedrische Teilmenge von $\mathbb{R}^N$.

2 Konvexe Funktionen

Nach der Betrachtung der konvexen Mengen folgt die Beschreibung von konvexen Funktionen im N-dimensionalen Raum.

In diesem Abschnitt bezeichnet D eine nicht leere konvexe Menge in $\mathbb{R}^N$.

Definition 5 *Für eine Funktion $f : D \longrightarrow \mathbb{R}$ nennt man die Menge*

$$\text{epi}(f) := \{(x, s) \in \mathbb{R}^N \times \mathbb{R};\ x \in D,\ f(x) \leq s\}$$

den **Epigraphen** *von f. Man nennt die Funktion f* **konvex***, wenn ihr Epigraph eine konvexe Teilmenge von $\mathbb{R}^N \times \mathbb{R} = \mathbb{R}^{N+1}$ ist.*

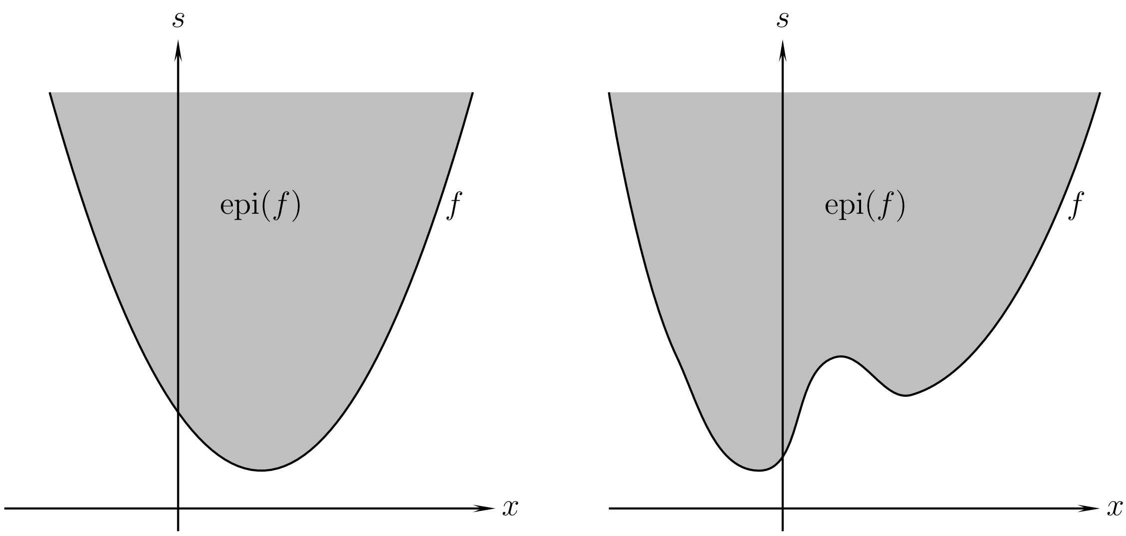

Abbildung 6: Epigraphen einer konvexen und einer nicht konvexen Funktion

Bemerkung 1 Eine Funktion $f : D \longrightarrow \mathbb{R}$ ist genau dann konvex, wenn sie die *Jensensche Ungleichung* erfüllt:

$$f((1-t)x + ty) \leq (1-t)f(x) + tf(y) \text{ für alle } x, y \in D,\ t \in [0, 1]. \tag{9}$$

Begründung:
1. Es sei $f : D \longrightarrow \mathbb{R}$ konvex. Für $x, y \in D$ gehören die Punkte $(x, f(x))$ und $(y, f(y))$ zur Menge epi(f). Da epi(f) konvex ist, gehört auch die Verbindungsstrecke dieser Punkte zu epi(f). Die Punkte der Verbindungsstrecke sind

$$(1-t)(x, f(x)) + t(y, f(y)) = ((1-t)x + ty, (1-t)f(x) + tf(y)),\ 0 \le t \le 1.$$

Die Zugehörigkeit dieser Punkte zu epi(f) besagt, dass die Beziehung (9) gilt.
2. Die Funktion $f : D \longrightarrow \mathbb{R}$ erfülle die Bedingung (9). Ferner seien (x, r) und (y, s) beliebige Punkte von epi(f). Für $t \in [0, 1]$ gilt dann auf Grund von (9)

$$f((1-t)x + ty) \le (1-t)f(x) + tf(y) \le (1-t)r + ts.$$

Daher gehören die Punkte

$$((1-t)x + ty, (1-t)r + ts) = (1-t)(x, r) + t(y, s),\ 0 \le t \le 1,$$

zu epi(f). Da dies die Punkte der Verbindungsstrecke von (x, r) und (y, s) sind, ist epi(f) konvex, also f eine konvexe Funktion.

Beispiel 1. Es sei $D = \mathbb{R}^N$ und $g(x) := a \cdot x + b$ für $a \in \mathbb{R}^N$, $b \in \mathbb{R}$. Solche Funktionen heißen **affin**. Für $t \in [0, 1]$ und $x, y \in D$ gilt

$$g((1-t)x+ty) = a\cdot((1-t)x+ty)+b = (1-t)(a\cdot x+b)+t(a\cdot y+b) = (1-t)g(x)+tg(y).$$

Folglich genügt g der Bedingung (9), d. h., g ist konvex. Die Graphen von affinen Funktionen bilden Hyperebenen in $\mathbb{R}^{N+1}$. Dies bedeutet, dass die Epigraphen abgeschlossene Halbräume sind.

Beispiel 2. Ist $D = \mathbb{R}^N$ und $f(x) := |x|$, so ist f konvex, denn es ist

$$f((1-t)x+ty) = |(1-t)x+ty| \le (1-t)|x|+t|y| = (1-t)f(x)+tf(y) \text{ für } x, y \in \mathbb{R}^N,\ t \in [0, 1].$$

Definition 6 *Ein Element $a \in \mathbb{R}^N$ heißt* **Subgradient** *einer konvexen Funktion $f : D \to \mathbb{R}$ im Punkt $x^* \in D$, wenn die durch*

$$g(x) := a \cdot (x - x^*) + f(x^*) \quad \text{für} \quad x \in \mathbb{R}^N \tag{10}$$

definierte affine Funktion g die Eigenschaft $g(x) \le f(x)$ für alle $x \in D$ besitzt. Ist a Subgradient von f in x^*, so nennt man diese affine Funktion g **Stützfunktion** *von f in x^* (siehe Abbildung 7 unten). Die Menge aller Subgradienten von f im Punkt x^* bezeichnet man mit $\partial f(x^*)$, und man nennt diese Menge das* **Subdifferential** *von f in x^*. Dabei ist es möglich, dass $\partial f(x^*) = \varnothing$ ist.*

Bemerkung 2 Für eine konvexe Funktion f ist $a \in \partial f(x^*)$ genau dann, wenn die Beziehung

$$a \cdot (x - x^*) \le f(x) - f(x^*) \quad \text{für alle} \quad x \in D \tag{11}$$

gilt. Die Formel (11) ist für $g(x) := a \cdot (x - x^*) + f(x^*)$ gleichwertig zur Ungleichung $g(x) \le f(x)$ für alle $x \in D$.

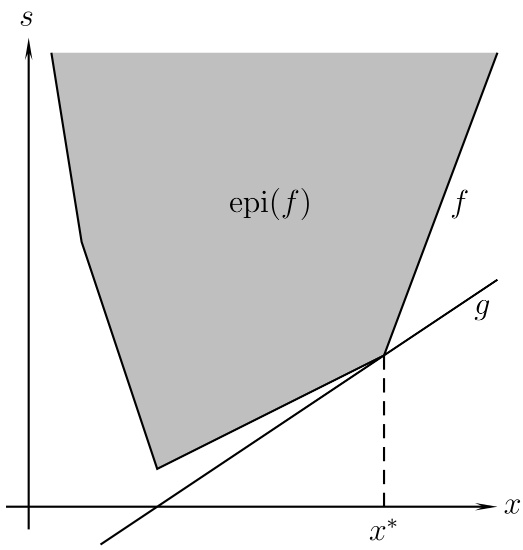

Abbildung 7: Stützfunktion g einer konvexen Funktion f im Punkt x^*

Satz 8 *Es sei $N = 1$, d. h., D sei ein Intervall in $\mathbb{R}$. Ferner sei $f : D \longrightarrow \mathbb{R}$ eine konvexe Funktion, die in dem inneren Punkt x^* von D differenzierbar ist. Dann ist $f'(x^*)$ der einzige Subgradient von f im Punkt x^*.*

Beweis:

1. Nach Bemerkung 2 ist $a \in \partial f(x^*)$, wenn die Beziehung

$$a(x - x^*) \leq f(x) - f(x^*) \quad \text{für alle} \quad x \in D$$

gilt. Da x^* innerer Punkt von D ist, gehört $x = x^* + h$ für h mit hinreichend kleinen Beträgen zu D. Folglich ist $ah \leq f(x^* + h) - f(x^*)$ für hinreichend kleine $|h|$. Für $h > 0$ ergibt sich nach Division durch h und Ausführung des Grenzübergangs $h \to 0$ die Ungleichung $a \leq f'(x^*)$. Für $h < 0$ ergibt sich analog $a \geq f'(x^*)$. Folglich kann nur $f'(x^*)$ Subgradient von f in x^* sein.

2. Wir zeigen, dass $f'(x^*)$ tatsächlich Subgradient ist. Es sei $x^* + h \in D$. Aufgrund der Konvexität von f folgt aus der Beziehung $x^* + th = (1 - t)x^* + t(x^* + h)$ für $t \in [0, 1]$ die Ungleichung

$$f(x^* + th) \leq (1 - t)f(x^*) + tf(x^* + h).$$

Für $t \in]0, 1]$ ist daher

$$\frac{1}{t}(f(x^* + th) - f(x^*)) \leq f(x^* + h) - f(x^*).$$

Das ist gleichwertig mit

$$\frac{1}{th}(f(x^* + th) - f(x^*))h \leq f(x^* + h) - f(x^*).$$

Für $t \to 0$ und damit auch $th \to 0$ ergibt sich

$$f'(x^*)h \leq f(x^* + h) - f(x^*). \tag{12}$$

Nach Bemerkung 2 zeigt diese Ungleichung, dass $f'(x^*) \in \partial f(x^*)$ ist. □

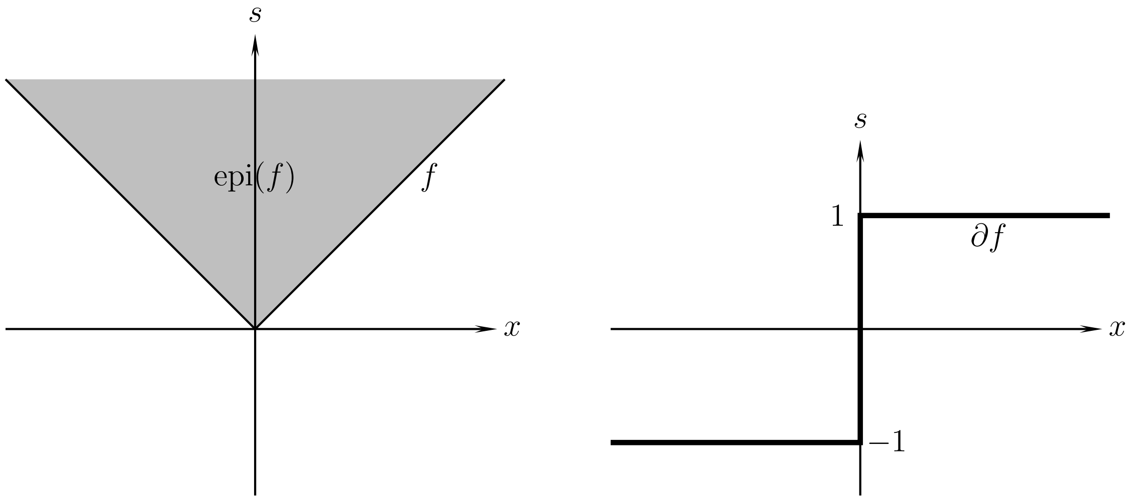

Abbildung 8: Epigraph und Subdifferential der Betragsfunktion

Satz 8 macht deutlich, dass man bei konvexen Funktionen das Subdifferential als eine Verallgemeinerung der Ableitung ansehen kann. Als weiteres Beispiel für $N = 1$ ist in Abbildung 8 Epigraph und Subdifferential der durch $f(x) := |x|$ für $x \in \mathbb{R}$ definierten Betragsfunktion f dargestellt. Es ist $\partial f(0) = [-1, 1]$, weil genau diejenigen Geraden durch den Nullpunkt Stützfunktionen für f liefern, deren Anstieg im Intervall $[-1, 1]$ liegt.

Wir kehren nun wieder zum Fall beliebiger Dimensionen N zurück.

Für eine konvexe Funktion $f : D \longrightarrow \mathbb{R}$ interpretiert man ∂f als eine „mehrdeutige Abbildung“ von D in $\mathbb{R}^N$, d. h., als eine Vorschrift A, die jedem Punkt $x \in D$ eine (möglicherweise leere) Teilmenge $A(x)$ von $\mathbb{R}^N$ zuordnet. Mehrdeutige Abbildungen sind gewöhnungsbedürftig. Das Arbeiten mit mehrdeutigen Abbildungen gewährt mehr Freiheit, als man sie beim Arbeiten allein mit eindeutigen Abbildungen hat.

Satz 9 *Es sei $f : D \longrightarrow \mathbb{R}$ konvex. Dann und nur dann, wenn $0 \in \partial f(x^*)$ ist, nimmt die Funktion f im Punkt x^* ihr Minimum an.*

Beweis:
Nach Bemerkung 2 ist $0 \in \partial f(x^*)$ genau dann, wenn $0 \cdot (x - x^*) \leq f(x) - f(x^*)$ ist für alle $x \in D$. Das ist offensichtlich genau dann der Fall, wenn $f(x^*) = \min\limits_{x \in D} f(x)$ ist. □

Beispiel 3. Es sei $f(x) := p \cdot x + q$ für $x \in D$ und Parameter $p \in \mathbb{R}^N$, $q \in \mathbb{R}$. Für $x^* \in D$ ist

$$\begin{aligned} a \in \partial f(x^*) \quad &\Longleftrightarrow \quad a \cdot (x - x^*) \leq f(x) - f(x^*) = p \cdot (x - x^*) \text{ für alle } x \in D \\ &\Longleftrightarrow \quad (a - p) \cdot (x - x^*) \leq 0 \quad \text{für alle } x \in D. \end{aligned}$$

Insbesondere ist $0 \in \partial f(x^*)$ und damit x^* Minimalstelle von f auf D genau dann, wenn

$$-p \cdot (x - x^*) \leq 0 \quad \text{für alle } x \in D \tag{13}$$

gilt.

Ist D zusätzlich beschränkt und abgeschlossen, so existiert (mindestens) ein $x^* \in D$ derart, dass $f(x^*) = d := \inf\limits_{x \in D} f(x)$ ist. Das zeigt man wie folgt: Man wählt eine Folge (x_n)

aus D mit der Eigenschaft $\lim_{n\to\infty} f(x_n) = d$. Dann besitzt die Folge (x_n) eine gegen einen Punkt $x^* \in D$ konvergente Teilfolge (x_{n_j}) (Satz von Bolzano-Weierstraß), und es gilt $f(x^*) = \lim_{j\to\infty} f(x_{n_j}) = d$.

Wir zeigen nun, dass die Minimierung affiner Funktionen auf konvexen Mengen, wie sie in Beispiel 3 beschrieben wird, auch für Optimierungsaufgaben der Ökonomie von Interesse ist.

Es seien $E_1, \ldots, E_m$ Erzeuger einer Ware (beispielsweise Molkereien, die Milch erzeugen) und $G_1, \ldots, G_n$ Geschäfte einer Ladenkette, in denen die Ware zum Verkauf angeboten wird. Die Kapazitäten der Erzeuger seien $c_1, \ldots, c_m$ Wareneinheiten pro Zeiteinheit (beispielsweise Liter/Tag). Der Bedarf von Geschäft G_j betrage b_j Wareneinheiten pro Zeiteinheit. Die Kosten für den Transport einer Wareneinheit vom Erzeuger E_i zum Geschäft G_j seien p_{ij}. Werden x_{ij} Wareneinheiten von E_i nach G_j transportiert, so sind die Gesamtkosten also $\sum_{i=1}^{m}\sum_{j=1}^{n} p_{ij}x_{ij}$. Es liegt im Interesse der Ladenkette, diesen Wert zu minimieren.
Dabei sind folgende Nebenbedingungen zu erfüllen:
1. $x_{ij} \geq 0$ für $i = 1, \ldots, m$, $j = 1, \ldots, n$: Es gibt keine negativen Warenmengen.
2. $\sum_{i=1}^{m} x_{ij} = b_j$, $j = 1, \ldots, n$: Der Bedarf der Geschäfte soll gedeckt werden.
3. $\sum_{j=1}^{n} x_{ij} \leq c_i$, $i = 1, \ldots, m$: Es kann nicht mehr abtransportiert als erzeugt werden.

Man fasst $x = (x_{ij})$ als Element von $\mathbb{R}^{mn}$ auf. Durch die Nebenbedingungen 1 - 3 wird in $\mathbb{R}^{mn}$ ein Polyeder D definiert. Es handelt sich nämlich um den Durchschnitt endlich vieler abgeschlossener Halbräume. (Jede lineare Gleichung ist zu zwei Ungleichungen äquivalent, und jede lineare Ungleichung definiert einen abgeschlossenen Halbraum.) Für $x \in D$ und für $i = 1, \ldots, m$, $j = 1, \ldots, n$ ist $0 \leq x_{ij} \leq b_i$. Daher ist D in $\mathbb{R}^{mn}$ beschränkt. Es kann sein, dass D leer ist. Das ist der Fall, wenn die Gesamtkapazität nicht ausreicht, um den Gesamtbedarf zu befriedigen. Wir nehmen an, dass $D \neq \varnothing$ ist. Die Aufgabe besteht darin, die durch die Vorschrift $f(x) := p\cdot x := \sum_{i=1}^{m}\sum_{j=1}^{n} p_{ij}x_{ij}$ für $x \in \mathbb{R}^{mn}$ definierte lineare Funktion f auf dem Polyeder D zu minimieren. Wie wir im Zusammenhang mit Beispiel 3 gesehen haben, ist dieses Problem lösbar. Auf die praktisch wichtige Frage nach einem Algorithmus zur Bestimmung einer Lösung sind wir aus Zeitgründen nicht eingegangen.

Schlusswort

Die Theorie der konvexen Mengen und der konvexen Funktionen ist umfangreicher, als unser Text zeigen kann. Obwohl die räumliche Vorstellungskraft für den N-dimensionalen Raum nicht ausreicht, ist es möglich Rechenregeln und Definitionen vom dreidimensionalen Raum auf höherdimensionale Räume zu übertragen. Dadurch sind wir in die Lage versetzt worden, hoch komplexe Optimierungsaufgaben zu erfassen und als lösbar nachzuweisen.

Mehr zu dem Thema findet man in Literatur über *Konvexe Analysis* [1)].

---

1) z. B.: “Convex Analysis” von R. Tyrrell Rockafellar, Princeton University Press, 1970