5 Nichtlineare Optimierungsaufgaben

5.1 Konvexe Programmierungsaufgaben

Wir betrachten im Folgenden allgemeine Optimierungsaufgaben in der „Standardform“

$$(P) \qquad x \in \mathbb{R}^n : \quad f(x) \to \min!, \quad g(x) \le 0,$$

mit einer Zielfunktion $f : \mathbb{R}^n \to \mathbb{R}$ und Restriktionen $g = (g_1, \cdot, g_m) : \mathbb{R}^n \to \mathbb{R}^m$. Der zugehörige „zulässige“ Bereich ist wieder definiert durch $M := \{x \in \mathbb{R}^n \,|\, g(x) \le 0\}$.

Im Allgemeinen Fall einer *nichtlinearen* Aufgabe sind die Fragen nach Charakterisierung und Berechnung von Lösungen natürlich sehr viel schwieriger als im *linearen* Fall.

Beispiel 5.1: Wir betrachten die Situation $n = 2, m = 3$ mit den Restriktionen

$$g_1(x) = -x_1 \le 0, \;\; g_2(x) = -x_2 \le 0, \;\; g_3(x) = x_1 + x_2 - 1 \le 0.$$

a) Linearer Fall $f(x) = x_1 - x_2$ mit Eckenlösung (Niveaulinien $f(x) = c$):

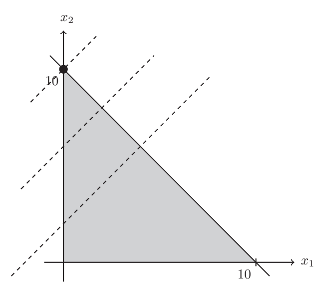

Abbildung 5.1: *Problem mit Eckenlösung*

b) Linear-quadratischer Fall $f(x) = (x_1 - 7)^2 + (x_2 - 7)^2$ mit Randlösung und $f(x) = (x_1 - 3)^2 + (x_2 - 3)^2$ mit innerer Lösung (Niveaulinien $f(x) = c$)

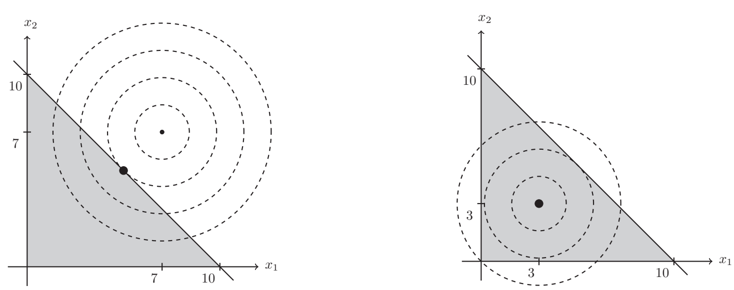

Abbildung 5.2: *Problem mit Randlösung (links) und innerer Lösung (rechts)*

Diese Beispiele zeigen, dass bei nichtlinearen Optimierungsaufgaben die Lösung, falls eine solche überhaupt existiert, sicher nicht durch eine spezielle Lage im zulässigen Bereich $M = \{x \in \mathbb{R}^n \,|\, g(x) \le 0\}$ ausgezeichnet ist. Um zu verwendbaren Aussagen zu kommen, müssen zusätzliche Annahmen über die Eigenschaften der Funktionen f und g_i getroffen werden. Die Aufgabe (P) wird „linear" genannt, wenn sowohl die Zielfunktion f als auch die Restriktion g affin-linear sind:

$$f(x) = c^T \cdot x + d, \quad g(x) = Ax - b,$$

mit Vektoren $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$ und einer Matrix $A \in \mathbb{R}^{m\times n}$, andernfalls ist sie „nichtlinear". Eine Funktion $f : \mathbb{R}^n \to \mathbb{R}$ ist genau dann affin-linear, wenn gilt:

$$f(\lambda x + (1-\lambda)x') = \lambda f(x) + (1-\lambda)f(x'), \quad x, x' \in \mathbb{R}^n, \quad \lambda \in (0,1). \tag{5.1.1}$$

Man nennt f „konvex", wenn gilt:

$$f(\lambda x + (1-\lambda)x') \le \lambda f(x) + (1-\lambda)f(x'), \quad x, x' \in \mathbb{R}^n, \quad \lambda \in (0,1), \tag{5.1.2}$$

(„konkav" für „$\ge$") und „streng konvex", wenn „$<$" für $x \ne x'$ gilt. Sind die Funktionen f und g_i konvex, so spricht man von einer „konvexen Optimierungsaufgabe". Viele Eigenschaften der linearen Aufgabe lassen sich sinngemäß auch für den konvexen Fall übertragen.

Ein zulässiger Punkt $x \in M$ der Aufgabe (P) wird „relatives (lokales) Minimum" genannt, wenn für eine ε-Umgebung $U_\varepsilon(x) = \{y \in \mathbb{R}^n \,|\, \|x-y\| < \varepsilon\}$ gilt:

$$f(x) \le f(y), \quad y \in U_\varepsilon(x) \cap M. \tag{5.1.3}$$

Ein Minimum von f auf ganz M heißt dann „absolut" oder auch „global".

Für spätere Zwecke listen wir einige Aussagen über konvexe Funktionen:

- Ist f konvex, so ist $-f$ konkav.
- Sind f_i $(i = 1,\dots,m)$ konvex, so ist $\sum_{i=1}^m f_i$ konvex.
- Für eine konvexe Funktion ist jedes relative Minimum auch absolut.
- Die Menge der Minima einer konvexen Funktion ist konvex.
- Eine streng konvexe Funktion hat höchstens ein Minimum.
- Eine auf einer offenen (konvexe) Menge definierte konvexe Funktion ist dort notwendig stetig.
- Ist f konvex, so ist die Menge $\{x \in \mathbb{R}^n \,|\, f(x) \le a\}$ für jedes $a \in \mathbb{R}$ konvex.

Als Durchschnitt von konvexen Mengen ist der zulässige Bereich einer konvexen Programmierungsaufgabe stets konvex; ist er beschränkt, so hat die Aufgabe (P) eine eindeutig bestimmte Lösung. Wir werden im Folgenden die Minimalstellen durch notwendige und hinreichende Bedingungen charakterisieren und dann Methoden zu ihrer numerischen Berechnung diskutieren. Dazu werden gewiss Glattheitseigenschaften von f und g erdorderlich sein. In der Regel werden diese Funktionen als einmal oder auch zweimal stetig (partiell) differenzierbar angenommen.

Lemma 5.1: *Eine Funktion $f \in C^1(M)$, M konvex, ist genau dann konvex, wenn*

$$f(y) \geq f(x) + (y-x)^T \cdot \nabla f(x), \quad x, y \in M, \tag{5.1.4}$$

und genau dann streng konvex, wenn für $x \neq y$ in (5.1.4) die strikte Ungleichung gilt.

Beweis: Aus der Ungleichung (5.1.4) folgt für $y, z \in M$ und $x = \lambda y + (1-\lambda)z$ mit $0 \leq \lambda \leq 1$:

$$\begin{aligned}
\lambda f(y) + (1-\lambda) f(z) &\geq \lambda\big(f(x) + (y-x)^T \cdot \nabla f(x)\big) + (1-\lambda)\big(f(x) + (z-x)^T \cdot \nabla f(x)\big) \\
&= f(x) + \big(\lambda(y-x) + (1-\lambda)(z-x)\big)^T \cdot \nabla f(x) \\
&= f(x) + \big(\lambda y + (1-\lambda)z - x\big)^T \cdot \nabla f(x) = f(x),
\end{aligned}$$

d. h.: f ist konvex. Ist umgekehrt f konvex, so gilt für die auf $(0,1)$ differenzierbare Funktion

$$\varphi(\lambda) := \lambda f(y) + (1-\lambda) f(x) - f(\lambda y + (1-\lambda)x) \geq 0$$

und $\varphi(0) = 0$. Also ist notwendig $\varphi'(0) \geq 0$. Dies impliziert

$$0 \leq \varphi'(0) = f(y) - f(x) - \nabla f(x)^T \cdot (y-x).$$

Für die strenge Konvexität argumentiert man analog. Q.E.D.

Lemma 5.2: *Eine Funktion $f \in C^2(M)$, M konvex mit nicht leerem Inneren $M^o \neq \emptyset$, ist konvex bzw. streng konvex, genau dann wenn die zugehörige Hesse[1]-Matrix*

$$\nabla^2 f(x) = \Big(\frac{\partial^2 f}{\partial x_i \partial x_k}(x)\Big)_{i,k=1}^n$$

für alle $x \in M$ positiv semi-definit bzw. positiv definit ist.

Beweis: Die Behauptung ergibt sich aus der Taylorschen Formel

$$f(y) = f(x) + (y-x)^T \cdot \nabla f(x) + \tfrac{1}{2}(y-x)^T \cdot \nabla^2 f(\xi)(y-x), \quad \xi \in \overline{(x,y)},$$

mit Hilfe von Lemma 5.1. Q.E.D.

Als Folgerung aus Lemma 5.2 finden wir, dass quadratische Funktionen

$$f(x) = c^T \cdot x + x^T \cdot Cx$$

mit (symmetrischer) positiv semi-definiter Matrix $C \in \mathbb{R}^{n \times n}$ konvex sind.

Für konvexe Optimierungsaufgaben haben wir den folgenden Satz über die Existenz von Minima.

[1] Ludwig Otto Hesse (1811–1874): Deutscher Mathematiker; wirkte in Königsberg, Heidelberg und München; Beiträge zur Theorie der algebraischen Funktionen und Invarianten

Satz 5.1 (Existenzsatz für konvexe Aufgaben): *In der Optimierungaufgabe* (P) *mit konvexen Funktionen* f *und* g_i, $i = 1, \dots, n$, *sei die zulässige Menge* M *beschränkt oder die Zielfunktion habe die Eigenschaft*

$$f(x) \to \infty \quad (x \in M,\ \|x\|_2 \to \infty). \tag{5.1.5}$$

Dann besitzt f *auf* M *ein Minimum* $\hat{x}$. *Im Fall der strikten Konvexität von* f *ist das Minimum eindeutig.*

Beweis: i) Auf der beschränkten (und damit kompakten) zulässigen Menge M nimmt die konvexe (und damit stetige) Funktion f ihr Minimum an.

ii) Sei nun M unbeschränkt und erfülle (5.1.5). Wir verwenden das sog. „Minimalfolgenargument". Wegen der Wachstumsbedingung (5.1.5) kann es keine Folge von Punkten $x^t \in M$ geben mit $f(x^t) \to -\infty$ $(t \to \infty)$. Also ist f nach unten beschränkt:

$$-\infty < \inf_{x \in M} f(x) =: d.$$

Sei nun $(x^t)_{t \in \mathbb{N}} \subset M$ eine zugehörige „Minimalfolge", d. h.: $f(x^t) \to d$ $(t \to \infty)$. Diese Minimalfolge muss, wieder wegen (5.1.5), beschränkt sein und besitzt wegen der Abgeschlossenheit von M folglich mindestens einen Häufungspunkt $\hat{x} \in M$. Für diesen gilt dann mit der zugehörigen Teilfolge $(x^{t_s})_{s \in \mathbb{N}}$ wegen der Stetigkeit von f:

$$f(\hat{x}) = \lim_{s \to \infty} f(x^{t_s}) = d,$$

d. h.: $\hat{x}$ ist Minimum von f auf M. Dieses Minimum ist dann für *strikt* konvexes f eindeutig bestimmt. Q.E.D.

5.2 Dualitätstheorie (globale Extremalbedingungen)

Im Folgenden untersuchen wir die „globalen" Eigenschaften der Optimierungsaufgabe (P), d. h.: charakterisierende Eigenschaften ihrer „globalen" Lösungen. Der „primalen" Programmierungsaufgabe

$$(P) \qquad x \in \mathbb{R}^n: \quad f(x) \to \min! \quad g(x) \le 0,$$

mit stetigen Funktionen $f: \mathbb{R}^n \to \mathbb{R}$ und $g \in \mathbb{R}^n \to \mathbb{R}^m$ wird die folgende sog. „Lagrange-Funktion" zugeordnet:

$$L(x, y) := f(x) + y^T \cdot g(x), \quad x \in \mathbb{R}^n,\ y \in \mathbb{R}^m.$$

Wir definieren

$$L^*(x) := \sup_{y \in \mathbb{R}^m_+} L(x, y) \in \mathbb{R} \cup \{\infty\}, \quad x \in \mathbb{R}^n,$$
$$L_*(y) := \inf_{x \in \mathbb{R}^n} L(x, y) \in \mathbb{R} \cup \{-\infty\}, \quad y \in \mathbb{R}^m_+,$$

sowie das zugehörige Paar zueinander „dualer“ Aufgaben

$$(\tilde{P}) \qquad x \in \mathbb{R}^n : \quad L^*(x) \to \min!$$

$$(P^*) \qquad y \in \mathbb{R}^m_+ : \quad L_*(y) \to \max!$$

Im Fall, dass $M := \{x \in \mathbb{R}^n \,|\, g(x) \le 0\} \ne \emptyset$, ist $(\tilde{P})$ eine äquivalente Formulierung zu (P), denn es gilt:

$$\begin{aligned} x \in M : &\qquad L^*(x) = \sup_{y \in \mathbb{R}^m_+} \{f(x) + y^T \cdot g(x)\} = f(x), \\ x \notin M : &\qquad L^*(x) = +\infty. \end{aligned}$$

Wir bezeichnen (P^*) als die zu (P) „duale“ Aufgabe. Wie in der linearen Programmierung ist man dann an „Alternativsätzen“ für die Probleme (P) und (P^*) interessiert. Im Folgenden setzen wir stets $M \ne \emptyset$ voraus.

Lemma 5.3: *Es gilt allgemein (im Zahlbereich $\mathbb{R} \cup \{\pm\infty\}$)*

$$L^*(x) \ge L_*(y), \quad x \in \mathbb{R}^n,\ y \in \mathbb{R}^m_+. \tag{5.2.6}$$

Ist für ein Paar $(\hat{x}, \hat{y}) \in \mathbb{R}^n \times \mathbb{R}^m_+$ das folgende sog. „Sattelpunktkriterium“ erfüllt,

$$L^*(\hat{x}) = L_*(\hat{y}), \tag{5.2.7}$$

so sind $\hat{x}$ und $\hat{y}$ Lösungen von (P) bzw. (P^*).

Beweis: i) Für $x \notin M$ ist definitionsgemäß $L^*(x) = \infty$ und folglich $L^*(x) \ge L_*(y)$ für alle $y \in \mathbb{R}^m_+$. Für $x \in M$ ist

$$\begin{aligned} L^*(x) = f(x) \ge f(x) + \underbrace{y^T}_{\ge 0} \cdot \underbrace{g(x)}_{\le 0} = L(x,y) \\ \ge \inf_{x \in \mathbb{R}^n} L(x,y) = L_*(y), \quad y \in \mathbb{R}^m_+. \end{aligned}$$

ii) Im Fall $L^*(\hat{x}) = L_*(\hat{y})$ ist also

$$\begin{aligned} L_*(\hat{y}) = L^*(\hat{x}) \ge L_*(y), \quad y \in \mathbb{R}^m_+, \\ L^*(\hat{x}) = L_*(\hat{y}) \le L^*(x), \quad x \in \mathbb{R}^n, \end{aligned}$$

d. h.: $\hat{x}$ and $\hat{y}$ sind Lösungen der Aufgaben $(\tilde{P})$ bzw. (P^*). Wegen $M \ne \emptyset$ muss $\hat{x} \in M$ sein, d. h.: $\hat{x}$ ist auch Lösung von Aufgabe (P). Q.E.D.

Lemma 5.4: *Die folgenden Aussagen sind äquivalent:*

1. $L^*(\hat{x}) = L_*(\hat{y})$.
2. $L^*(\hat{x}) = L(\hat{x}, \hat{y}) = L_*(\hat{y})$.

3. $L(\hat{x},y) \le L(\hat{x},\hat{y}) \le L(x,\hat{y}), \quad x \in \mathbb{R}^n,\ y \in \mathbb{R}^m_+ .$

4. $\min_{x\in\mathbb{R}^n} \sup_{y\in\mathbb{R}^m_+} L(x,y) = \max_{y\in\mathbb{R}^m_+} \inf_{x\in\mathbb{R}^n} L(x,y) .$

Beweis: (1) $\to$ (2) : Gilt (1), so folgt auch (2) wegen:

$$L(\hat{x},\hat{y}) \le \sup_{z\in\mathbb{R}^m_+} L(\hat{x},z) = \underbrace{L^*(\hat{x}) = L_*(\hat{y})}_{(1)} = \inf_{x\in\mathbb{R}^n} L(x,\hat{y}) \le L(\hat{x},\hat{y}).$$

(2) $\to$ (3) : Gilt (2), so folgt (3) wegen der folgenden Beziehungen für $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m_+$:

$$L(\hat{x},y) \le \sup_{z\in\mathbb{R}^m_+} L(\hat{x},z) = \underbrace{L^*(\hat{x}) = L(\hat{x},\hat{y}) = L_*(\hat{y})}_{(2)} = \inf_{z\in\mathbb{R}^n} L(z,\hat{y}) \le L(x,\hat{y}).$$

(3) $\to$ (2) : Dies ergibt sich trivialerweise.

(2) $\to$ (1) : Dies ergibt sich trivialerweise.

(1) $\to$ (4) : Gilt (1), so folgt nach Lemma 5.3 auch (4) gemäß

$$\min_{x\in\mathbb{R}^n} L^*(x) = \underbrace{L^*(\hat{x}) = L_*(\hat{y})}_{(1)} = \max_{y\in\mathbb{R}^m_+} L_*(y).$$

(4) $\to$ (1) : Gilt (4), so gibt es $\hat{x} \in \mathbb{R}^n$ und $\hat{y} \in \mathbb{R}^m_+$, mit denen (1) gilt:

$$L^*(\hat{x}) = L_*(\hat{y}).$$

Dies vervollständigt den Beweis. Q.E.D.

Die Beziehung (3) in Lemma 5.4 besagt, dass das Paar $(\hat{x},\hat{y}) \in \mathbb{R}^n \times \mathbb{R}^m_+$ ein „Sattelpunkt" der Lagrange-Funktion $L(x,y)$ ist. Lemma 5.3 besagt also, dass jeder Sattelpunkt $(\hat{x},\hat{y})$ von $L(x,y)$ mit $\hat{x}$ und $\hat{y}$ Lösungen von (P) bzw. (P^*) liefert. Die Umkehrung ist i. Allg. nicht richtig; dazu sind weitere Voraussetzungen erforderlich. An den zulässigen Bereich M stellen wir im Folgenden die sog. „Slater[2]-Bedingung":

$$M^o := \{x \in M \mid g(x) < 0\} \ne \emptyset. \tag{5.2.8}$$

Diese Bedingung macht offenbar bei Gleichungsrestriktionen keinen Sinn. Damit haben wir den folgenden zentralen Satz der „konvexen Optimierung".

Satz 5.2 (Globaler Karush-Kuhn-Tucker-Satz): *Es seien f, g_i, $i = 1,\dots,m$, konvex und $M^o \ne \emptyset$. Dann ist ein $\hat{x} \in M$ genau dann Lösung von (P), wenn es ein $\hat{y} \in \mathbb{R}^m_+$ gibt, so dass $(\hat{x},\hat{y})$ Sattelpunkt von $L(x,y)$ ist.*

Zum Beweis dieses Satzes benötigen wir den folgenden sog. „Trennunssatz" für konvexe Mengen.

[2]Morton L. Slater (1921–2002): US-amerikanischer Mathematiker; arbeitete an den Sandia National Laboratories (USA); bekannt durch die nach ihm benannte „Slater-Bedingung" in der nichtlinearen Optimierung.

Lemma 5.5 (Satz von der trennenden Hyperebene): *Es seien $B_1, B_2 \subset \mathbb{R}^m$ zwei nichtleere, disjunkte Teilmengen, wobei B_1 als konvex und B_2 als konvex und offen angenommen sind. Dann existiert eine Hyperebene, welche die beiden Mengen trennt, d. h.: Es existieren ein $a \in \mathbb{R}^m \setminus \{0\}$ und ein $\beta \in \mathbb{R}$, so dass*

$$a^T \cdot y - \beta > 0 \geq a^T \cdot x - \beta, \quad x \in B_1, \ y \in B_2.$$

Beweis: Anschaulich ist die Aussage klar. Ihr rigoroser Beweis ist dennoch sehr technisch. i) Wir zeigen die Behauptung zunächst für den Spezielfall $B_1 = \{x\}$. Da $\bar{B}_2$ konvex und abgeschlossen ist, gibt es ein $\hat{y} \in \bar{B}_2$ mit der Eigenschaft (Übungsaufgabe)

$$\|x - \hat{y}\|_2 = \inf_{y \in B_2} \|x - y\|_2.$$

Für $y \in B_2$ ist $\lambda y + (1-\lambda)\hat{y} \in \bar{B}_2$ für $\lambda \in (0,1)$. Damit erhalten wir

$$\|x - \hat{y}\|_2^2 \leq \|x - \lambda y - (1-\lambda)\hat{y}\|_2^2 = \|x - \hat{y} + \lambda(\hat{y} - y)\|_2^2$$

$$= \|x - \hat{y}\|_2^2 + 2\lambda(x - \hat{y}, \hat{y} - y) + \lambda^2 \|\hat{y} - y\|_2^2 .$$

Für $\lambda \to 0$ folgern wir $(x - \hat{y}, \hat{y} - y)_2 \geq 0$ und damit

$$(\hat{y} - x, y)_2 \geq (\hat{y} - x, \hat{y})_2 = (\hat{y} - x, x)_2 + \|\hat{y} - x\|_2^2 .$$

Wir unterscheiden nun zwei Fälle.

a) Fall $x \notin \partial B_2$: In diesem Fall ist $\|\hat{y} - x\|_2 > 0$ und mit der Setzung

$$a := \frac{\hat{y} - x}{\|\hat{y} - x\|_2}, \qquad \beta := (a, x)_2 ,$$

ergibt sich

$$(a, y)_2 \geq (a, x)_2 + \underbrace{\|\hat{y} - x\|_2}_{>0} \quad \Rightarrow \quad (a, y)_2 - \beta > 0 = (a, x)_2 - \beta, \quad y \in B_2 .$$

b) Fall $x \in \partial B_2$: Sei $(x^t)_{t \in \mathbb{N}} \subset \bar{B}_2^c$ (Komplement von $\bar{B}_2$) eine Folge mit $x^t \to x$ $(t \to \infty)$ und a^t die in (a) zugeordneten Vektoren mit $\|a^t\|_2 = 1$ und Zahlen $\beta_t := (a^t, x^t)_2$. Da die Folge der a^t beschränkt ist, besitzt sie eine Teilfolge $(a^{t_s})_{s \in \mathbb{N}}$, welche gegen ein a konvergiert: $a^{t_s} \to a$ $(s \to \infty)$. Damit folgt für $y \in B_2$:

$$(a, y)_2 = \lim_{s \to \infty} (a^{t_s}, y)_2 \geq \lim_{s \to \infty} \left\{ (a^{t_s}, x^{t_s})_2 + \|\hat{y} - x^{t_s}\|_2 \right\} \geq (a, x)_2 ,$$

d. h.:

$$(a, y)_2 - \beta \geq 0 = (a, x)_2 - \beta, \quad y \in B_2 .$$

Die volle Behauptung mit dem „>“-Zeichen ergibt sich mit Hilfe eines Widerspruchargu-ments analog zu dem weiter unten verwendeten.

ii) Wir behandeln nun den Fall einer allgemeinen konvexen Menge B_1. Die Menge

$$B := \{y - x \,|\, y \in B_2,\ x \in B_1\}$$

ist konvex und wegen $B = \cup_{x\in B_1}\{y-x \,|\, y \in B_2\}$ offen. Wegen $B_1 \cap B_2 = \emptyset$ ist außerdem $0 \notin B$. Teil (i) angewendet für die Mengen $B_1 = \{0\}$ und $B_2 = B$ liefert also die Existenz eines $a \in \mathbb{R}^n \setminus \{0\}$, so dass

$$(a, y-x)_2 > 0, \qquad (a,y)_2 > (a,x)_2, \quad x \in B_1,\ y \in B_2\,.$$

Mit der Setzung $\beta := \inf_{y\in B_2}(a,y)_2$ gilt dann

$$(a,x)_2 \le \beta \le (a,y)_2, \quad x \in B_1,\ y \in B_2\,.$$

Angenommen, es gibt ein $y \in B_2$ mit $(a,y)_2 = \beta$. Dann ist auch $(a, y \pm \varepsilon a)_2 \ge \beta$, für kleines $\varepsilon \in \mathbb{R}_+$, da B_2 offen ist. Dies ergibt den Widerspruch $\pm\varepsilon\|a\|_2\| \ge 0$ und somit $(a,x)_2 \le \beta < (a,y)_2$. Dies vervollständigt den Beweis. Q.E.D.

Beweis: (Beweis von Satz 5.2)

i) Für einen Sattelpunkt $(\hat{x},\hat{y}) \in \mathbb{R}^n \times \mathbb{R}^m_+$ von $L(x,y)$ ist nach Lemma 5.3 $\hat{x}$ Lösung von Aufgabe (P).

ii) Sei nun umgekehrt $\hat{x}$ eine Lösung von (P). Wir definieren die Mengen

$$\begin{aligned} B_1 &:= \big\{y^T = (y_0, y_1, \dots, y_m)^T \in \mathbb{R}^{1+m} \,|\, y_0 \ge f(x),\ y_i \ge g_i(x) \\ &\qquad\qquad\qquad\qquad \text{für mindestens ein } x \in \mathbb{R}^n\ (i = 1, \dots, n)\big\} \\ B_2 &:= \big\{z^T = (z_0, z_1, \dots, z_m)^T \in \mathbb{R}^{1+m} \,|\, z_0 < f(\hat{x}),\ z_i < 0\ (i = 1, \dots, m)\big\}. \end{aligned}$$

Die Konvexität von f und der g_i impliziert die von B_1; B_2 ist trivialerweise konvex. Ferner ist $B_1 \neq \emptyset$ und $B_2 \neq \emptyset$ und offen sowie $B_1 \cap B_2 = \emptyset$, da $\hat{x}$ Lösung von (P) ist. Also ist der Trennungssatz (Lemma 5.5) auf die Mengen B_1, B_2 anwendbar und liefert die Exsitenz eines Vektors $a^T = (a_0, a_1, \dots, a_m)^T \in \mathbb{R}^{1+m}$, $a \neq 0$, mit dem gilt:

$$a^T \cdot y > a^T \cdot z, \quad y \in B_1,\ z \in B_2.$$

Es ist $a \ge 0$, da sonst für eine geeignete Folge $(z^t)_{t\in\mathbb{N}} \subset B_2$ gälte: $a^T \cdot z^t \to \infty\ (t \to \infty)$. Weiter gilt

$$a^T \cdot y \ge a^T \cdot z, \quad y \in B_1,\ z \in \bar{B}_2.$$

Wir setzen für $x \in \mathbb{R}^n$:

$$y^T := (f(x), g_1(x), \dots, g_m(x))^T \in B_1, \qquad z^T := (f(\hat{x}, 0, \dots, 0)^T \in \bar{B}_2,$$

und erhalten dafür

$$a_0 f(x) + \sum_{i=1}^m a_i g_i(x) = a^T \cdot y \ge a^T \cdot z = a_0 f(\hat{x}), \quad x \in \mathbb{R}^n.$$

Es ist $a_0 > 0$, denn für $a_0 = 0$ wäre

$$\sum_{i=1}^{m} a_i g_i(x) \geq 0, \quad x \in \mathbb{R}^n,$$

im Widerspruch zu $M^o \neq \emptyset$. Wir setzen nun

$$\hat{y}^T := \left(\frac{a_1}{a_0}, \ldots, \frac{a_m}{a_0}\right)^T \in \mathbb{R}^m_+.$$

Damit gilt dann:

$$f(x) + \hat{y}^T \cdot g(x) = f(x) + \sum_{i=1}^{m} \hat{y}_i g_i(x) \geq f(\hat{x}), \quad x \in \mathbb{R}^n. \tag{5.2.9}$$

Für $x = \hat{x}$ folgt speziell

$$\underbrace{\hat{y}^T}_{\geq 0} \cdot \underbrace{g(\hat{x})}_{\leq 0} \geq 0 \quad \Rightarrow \quad \hat{y}^T \cdot g(\hat{x}) = 0.$$

Ausgehend von (5.2.9) ergibt sich also für $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m_+$:

$$L(x, \hat{y}) \geq f(\hat{x}) = f(\hat{x}) + \underbrace{\hat{y}^T \cdot g(\hat{x})}_{= 0} \geq f(\hat{x}) + \underbrace{y^T \cdot g(\hat{x})}_{\leq 0} = L(\hat{x}, y),$$

d. h.: $(\hat{x}, \hat{y})$ ist Sattelpunkt von $L(x, y)$. Q.E.D.

Korollar 5.1: *Es seien f, $g_i, i = 1, \ldots, m$, konvex, $M^o \neq \emptyset$ und die Aufgabe (P) sei lösbar. Dann ist auch die duale Aufgabe (P^*) lösbar. Gilt für eine Lösung $\hat{y} \in \mathbb{R}^m_+$ von (P^*) mit einem $\hat{x} \in M$*

$$L_*(\hat{y}) = f(\hat{x}), \tag{5.2.10}$$

so ist $\hat{x}$ Lösung von (P).

Beweis: Ist (P) lösbar, so gibt es nach Satz 5.2 einen Sattelpunkt $(\hat{x}, \hat{y}) \in \mathbb{R}^n \times \mathbb{R}^m_+$ von $L(x, y)$. Nach Lemma 5.4 genügt dieser dem Sattelpunktkriterium von Lemma 5.3, und $\hat{y}$ ist folglich Lösung von (P^*). Für $\hat{x} \in M$ ist $L^*(\hat{x}) = f(\hat{x})$. Im Fall $L_*(\hat{y}) = f(\hat{x})$ ist das Sattelpunktkriterium erfüllt und nach Lemma 5.3 ist $\hat{x}$ Lösung von (P). Q.E.D.

Beispiel 5.2: Wir betrachten als Beispiel eine quadratische Optimierungsaufgabe im $\mathbb{R}^2$ mit $n = 2$, $m = 1$:

$$f(x) := x_1^2 + x_2^2 - x_1 - x_2, \quad g(x) := 1 + x_1;$$

der zulässige Bereich ist $M := \{x \in \mathbb{R}^2 \,|\, 1 + x_1 \leq 0\}$. Diese Aufgabe ist offensichtlich konvex und es ist $M^o \neq \emptyset$. Die Hesse-Matrix von f ist positiv-definit,

$$\nabla^2 f(x) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix},$$

so dass f *strikt* konvex ist. Da $f(x) \to \infty$ $(\|x\|_2 \to \infty)$ existiert (nach Satz 5.1) ein eindeutig bestimmtes (globales) Minimum $\hat{x} \in M$ von f. Die zugehörige Lagrange-Funktion ist

$$L(x, y) = x_1^2 + x_2^2 - x_1 - x_2 + y(1 + x_1).$$

Nach Korollar 5.1 besitzt dann die zugehörige *duale* Aufgabe

$$(P^*) \qquad y \in \mathbb{R}_+ : \quad L_*(y) := \inf_{x \in \mathbb{R}^2} L(x, y) \to \max!$$

eine Lösung $\hat{y} \in \mathbb{R}^1_+$. Zur Berechnung von $\hat{y}$ dienen die folgenden Gleichungen:

$$0 = \nabla_x L(x, y)|_{x=\hat{x}} = \begin{pmatrix} 2\hat{x}_1 - 1 + y \\ 2\hat{x}_2 - 1 \end{pmatrix} \quad \Rightarrow \quad \begin{pmatrix} \hat{x}_1 \\ \hat{x}_2 \end{pmatrix} = \begin{pmatrix} \frac{1-y}{2} \\ \frac{1}{2} \end{pmatrix}.$$

Es ergibt sich damit

$$L_*(y) = \Big(\frac{1-y}{2}\Big)^2 + \frac{1}{4} - \frac{1-y}{2} - \frac{1}{2} + y\Big(1 + \frac{1-y}{2}\Big) = -\frac{y^2}{4} + \frac{3}{2}y - \frac{1}{2}.$$

Das Maximum $\hat{y}$ dieser Funktion erhalten wir aus der Gleichung

$$\frac{d}{dy} L_*(y)|_{y=\hat{y}} = -\frac{1}{2}\hat{y} + \frac{3}{2} = 0$$

zu $\hat{y} = 3$. Damit folgt weiter $\hat{x}_1 = -1$. Es gilt

$$L_*(\hat{y}) = \max_{y \in \mathbb{R}_+} L_*(y) = \frac{7}{4} = f(\hat{x}).$$

Da $\hat{x} \in M$, ist es nach Korollar 5.1 auch Lösung von (P).

5.3 Lokale Extremalbedingungen

In diesem Abschnitt werden Bedingungen hergeleitet, welche von *lokalen* Minima der Optimierungsaufgabe (P) notwendig erfüllt sind. Die Funktionen f und g_i, $i = 1, \dots, m$, sind im Folgenden stets als mindestens einmal stetig (partiell) differenzierbar angenommen. Jedes unrestringierte, lokale Minimum $\hat{x} \in \mathbb{R}^n$ der Funktion f genügt notwendig der Bedingung $\nabla f(\hat{x}) = 0$, bzw.

$$d^T \cdot \nabla f(\hat{x}) = 0 \quad \forall d \in \mathbb{R}^n. \tag{5.3.11}$$

Definition 5.1: *Ein* $d \in \mathbb{R}^n$ *heißt „zulässige Richtung" für* $x \in M = \{y \in \mathbb{R}^n \,|\, g(y) \le 0\}$*, wenn mit einem* $\varepsilon(x,d) > 0$ *gilt:*

$$g(x + \varepsilon d) \le 0 \quad \forall \varepsilon \in [0, \varepsilon(x,d)]. \tag{5.3.12}$$

Die Menge aller zulässigen Richtungen für x *sei mit* $D(x)$ *bezeichnet.*

Lemma 5.6: *Für ein lokales Minimum* $\hat{x} \in M$ *der Optimierungsaufgabe* (P) *gilt notwendig*

$$d^T \cdot \nabla f(\hat{x}) \ge 0 \quad \forall d \in \overline{D(x)}. \tag{5.3.13}$$

Beweis: Für $d \in D(\hat{x})$ folgt aus $f(\hat{x} + \varepsilon d) \ge f(\hat{x})$:

$$d^T \cdot \nabla f(\hat{x}) = \lim_{\varepsilon \to 0} \varepsilon^{-1}\big(f(\hat{x} + \varepsilon d) - f(\hat{x})\big) \ge 0.$$

Dies impliziert wegen der Stetigkeit von ∇f die Behauptung. Q.E.D.

Für $x \in M$ sei mit $I(x)$ die Indexmenge der „aktiven" Restriktionen bezeichnet:

$$I(x) := \big\{ i \in \{1, \dots, n\} \,|\, g_i(x) = 0 \big\}.$$

Lemma 5.7: *Sei* $x \in M$*. Dann gilt für jede zulässige Richtung* $d \in D(x)$ *:*

$$d^T \cdot \nabla g_i(x) \le 0, \quad i \in I(x). \tag{5.3.14}$$

Beweis: Sei $d \in D(x)$ und $i \in I(x)$. Dann ist

$$d^T \cdot \nabla g_i(x) = \lim_{\varepsilon \to 0} \varepsilon^{-1}\big(\underbrace{g_i(x + \varepsilon d)}_{\le 0} - \underbrace{g_i(x)}_{=0}\big) \le 0,$$

was gerade die Behauptung ist. Q.E.D.

Für $x \in M$ setzen wir

$$D_1(x) := \{ d \in \mathbb{R}^n \,|\, d^T \cdot \nabla g_i(x) \le 0,\, i \in I(x) \}.$$

Aufgrund von Lemma 5.7 ist sicher $\overline{D(x)} \subset D_1(x)$. Die Umkehrung ist aber i. Allg. nicht richtig, was durch folgende Beispiele belegt wird.

Beispiel 5.3: Wir betrachten den zweidimensionalen Fall $\mathbb{R}^2$:

a) $g(x) = x_1^3$, $M = \{x \in \mathbb{R}^3 \,|\, x_1 \in \mathbb{R}_+,\, x_2 \in \mathbb{R}\}$,

$$\nabla g(x) = (3x_1^2, 0)^T \quad \Rightarrow \quad D_1(0) = \mathbb{R}^2 \not\subset \overline{D(0)} = M.$$

b) $g(x) = x_1$ $M = \{x \in \mathbb{R}^3 \,|\, x_1 \in \mathbb{R}_+,\, x_2 \in \mathbb{R}\}$,

$$\nabla g(x) = (1, 0)^T \quad \Rightarrow \quad D_1(0) = M \subset D(0) = M.$$

Die Beispiele illustrieren, dass die Menge $D(x)$ durch die geometrischen Eigenschaften der zulässigen Menge M bestimmt ist, wogegen $D_1(x)$ von der speziellen Darstellung von M durch die Funktionen g_i abhängt. Dies gibt Anlass zu der folgenden Regularitätsbedingung der Kuhn-Tucker-Theorie:

Definition 5.2: *Ein Punkt $x \in M$ wird „regulär“ genannt, wenn $\overline{D(x)} = D_1(x)$ ist.*

Für Minimierungsaufgaben mit Gleichungsnebenbedingungen

$$f(x) \to \min! \quad g(x) = 0,$$

gilt die bekannte „Euler-Lagrangesche Multiplikatorenregel“ (Satz 4.3): Sei $\hat{x} \in D$ ein lokales Minimum von f unter der Nebenbedingung $g(\hat{x}) = 0$, und habe die Jacobi-Matrix $\nabla g(\hat{x}) \in \mathbb{R}^{m\times m}$ maximalen Rang m. Dann gibt es einen Vektor $\hat{y} \in \mathbb{R}^m$ („Multiplikator“), so dass

$$\nabla f(\hat{x}) + \hat{y}^T \cdot \nabla g(\hat{x}) = 0. \tag{5.3.15}$$

Wir werden jetzt diess Resultat auf den Fall von Ungleichungsnebenbedingungen übertragen (Satz von Karush-Kuhn-Tucker).

Satz 5.3 (Lokaler Karush-Kuhn-Tucker-Satz): *Sei $\hat{x}$ ein lokales Minimum der Aufgabe (P), und sei $\hat{x}$ regulär. Dann gibt es einen Vektor $\hat{y} \in \mathbb{R}^m_+$ so dass die folgenden „Karush-Kuhn-Tucker-Bedingungen“ erfüllt sind:*

$$\hat{y}^T \cdot g(\hat{x}) = 0, \tag{5.3.16}$$

$$\nabla f(\hat{x}) + \hat{y}^T \cdot \nabla g(\hat{x}) = 0. \tag{5.3.17}$$

Beweis: Wir setzen $b := \nabla f(\hat{x}) \in \mathbb{R}^n$ und

$$A^T := -\big(\nabla g_i(\hat{x})^T\big)_{i\in I(\hat{x})} \in \mathbb{R}^{k\times n}, \quad k = \#I(\hat{x}).$$

Wegen $\overline{D(\hat{x})} = D_1(\hat{x})$ gilt dann nach Lemma 5.6 notwendig $b^T \cdot d \geq 0$ für alle Richtungsvektoren $d \in \mathbb{R}^n$ mit $A^T d \geq 0$. Dies bedeutet, dass die Aufgabe

$$d \in \mathbb{R}^n : \qquad b^T \cdot d < 0, \quad A^T d \geq 0, \tag{5.3.18}$$

unlösbar ist. Nach dem Alternativsatz 1.2 muss dann die Aufgabe

$$y \in \mathbb{R}^k : \qquad Ay = b, \quad y \geq 0, \tag{5.3.19}$$

lösbar sein. Mit einer solchen Lösung y^* setzen wir $\hat{y}_i := y^*_i$, $i \in I(\hat{x})$, $\hat{y}_i := 0$ sonst, und erhalten somit ein $\hat{y} \in \mathbb{R}^m_+$, mit dem gilt:

$$\sum_{i=1}^m \hat{y}_i g_i(\hat{x}) = \sum_{i\in I(\hat{x})} y^*_i g_i(\hat{x}) + \sum_{i\notin I(\hat{x})} \hat{y}_i g_i(\hat{x}) = 0, \tag{5.3.20}$$

$$\nabla f(\hat{x}) + \sum_{i=1}^m \hat{y}_i \nabla g_i(\hat{x}) = b - Ay = 0, \tag{5.3.21}$$

was den Beweis vervollständigt. Q.E.D.

Mit Hilfe der Lagrange-Funktion

$$L(x,y) := f(x) + y^T \cdot g(x)$$

lassen sich die Karush-Kuhn-Tucker-Bedingungen in der folgenden abgekürzten Form schreiben:

$$\hat{y}^T \cdot \nabla_y L(\hat{x},\hat{y}) = 0, \tag{5.3.22}$$

$$\nabla_x L(\hat{x},\hat{y}) = 0. \tag{5.3.23}$$

Die Karush-Kuhn-Tucker-Bedingungen sind zwar notwendig aber i. Allg. nicht hinreichend für reguläre Punkte $x \in M$, ein lokales Minimum zu sein; z. B. für „innere“ Sattelpunkte von f. Dazu müssen wesentlich einschränkendere Bedingungen gestellt werden.

Lemma 5.8: *Die Funktionen g_i, $i = 1,\dots,m$, seien konvex, und es sei $M^o \neq \emptyset$ (sog. „Slater-Bedingung“). Dann ist jeder Punkt $x \in M$ regulär.*

Beweis: Aufgrund von Lemma 5.7 bleibt, $D_1(x) \subset \overline{D(x)}$ zu zeigen. Sei $d \in D_1(x)$, d. h.: $d^T \cdot \nabla g_i(x) \le 0$, $i \in I(x)$. Wegen $M^o \neq \emptyset$ existiert ein $\hat{x} \in M$ mit $g(\hat{x}) < 0$. Für $i \in I(x)$ folgt damit nach Lemma 5.1:

$$0 > g_i(\hat{x}) \ge \underbrace{g_i(x)}_{=0} + (\hat{x} - x)^T \cdot \nabla g_i(x).$$

Für $t > 0$ ist dann auch

$$\big(d + t(\hat{x} - x)\big)^T \cdot \nabla g_i(x) < 0, \quad i \in I(x).$$

Hieraus schließen wir, dass $d_t := d + t(\hat{x} - x) \in D(x)$ ist. Andernfalls gäbe es eine Folge von $\varepsilon_k > 0$, $\varepsilon_k \to 0$ $(k \to \infty)$, mit $x + \varepsilon_k d_t \notin M$, d. h.: $g_i(x + \varepsilon_k d_t) > 0$ für ein i. Im Falle $i \notin I(x)$ gilt $g_i(x) < 0$, d. h.: Es folgt ein Widerspruch. Im Falle $i \in I(x)$ gilt $g_i(x) = 0$, d. h.: Es folgt

$$d_t^T \cdot \nabla g_i(x) = \lim_{\varepsilon_k \to 0} \varepsilon_k^{-1}\big(g_i(x + \varepsilon_k d_t) - g_i(x)\big) \ge 0,$$

im Widerspruch zu $d_t^T \cdot \nabla g_i(x) < 0$. Für $t \to 0$, ergibt sich somit $d \in \overline{D(x)}$. Q.E.D.

Satz 5.4: *Die Optimierungsaufgabe (P) sei konvex, und es sei $M^o \neq \emptyset$. Dann sind die Karush-Kuhn-Tucker-Bedingungen notwendig und hinreichend dafür, dass ein Punkt $\hat{x} \in M$ Lösung von (P) ist.*

Beweis: Es bleibt zu zeigen,dass die Karush-Kuhn-Tucker-Bedingungen hinreichend sind. Seien $\hat{x} \in M$ und $\hat{y} \in \mathbb{R}^m_+$ mit

$$\begin{aligned} \hat{y}^T \cdot \nabla_y L(\hat{x}, \hat{y}) &= 0, \\ \nabla_x L(\hat{x}, \hat{y}) &= 0. \end{aligned}$$

Wir wollen zeigen, dass $(\hat{x}, \hat{y})$ ein Sattelpunkt der Euler-Lagrange-Funktion $L(x, y)$ ist. Lemma 5.3 (Sattelpunktkriterium) besagt dann, dass $\hat{x}$ globale Lösung von (P) ist. Die Funktion $\varphi(x) := L(x, \hat{y})$ ist wegen $\hat{y} \geq 0$ konvex und erfüllt nach Lemma 5.1 die Beziehung

$$\varphi(x) \geq \varphi(\hat{x}) + (x - \hat{x})^T \cdot \nabla\varphi(\hat{x}), \quad x \in \mathbb{R}^n.$$

Die Funktion $\psi(y) := L(\hat{x}, y)$ ist affin-linear, so dass

$$\psi(y) = \psi(\hat{y}) + (y - \hat{y})^T \cdot \nabla\psi(\hat{y}), \quad y \in \mathbb{R}^m.$$

Dies zusammengenommen impliziert für $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m_+$:

$$\begin{aligned} L(\hat{x}, y) &= L(\hat{x}, \hat{y}) + (y - \hat{y})^T \cdot \nabla_y L(\hat{x}, \hat{y}) \\ &= L(\hat{x}, \hat{y}) + \underbrace{y^T \cdot g(\hat{x})}_{\leq 0} - \underbrace{\hat{y}^T \cdot g(\hat{x})}_{=0} \\ &\leq L(\hat{x}, \hat{y}) \leq L(x, \hat{y}) - (x - \hat{x})^T \cdot \underbrace{\nabla_x L(\hat{x}, \hat{y})}_{=0} = L(x, \hat{y}), \end{aligned}$$

d. h.: $(\hat{x}, \hat{y})$ ist ein Sattelpunkt von $L(x, y)$. Q.E.D.

Bemerkung 5.1: Unter den Voraussetzungen von Satz 5.4 sind die Karush-Kuhn-Tucker-Bedingungen sogar äquivalent zur Sattelpunktbedingung, denn $\varphi(\hat{x}) = \min_{x \in \mathbb{R}^n} \varphi(x)$ und $\psi(\hat{y}) = \max_{y \in \mathbb{R}^m} \psi(y)$ impliziert notwendig

$$\nabla_x L(\hat{x}, \hat{y}) = \nabla\varphi(\hat{x}) = 0, \quad \hat{y}^T \cdot \nabla_y L(\hat{x}, \hat{y}) = \frac{d}{dr}\psi(\hat{y} + r\hat{y})_{|r=0} = 0$$

Beispiel 5.4: Wir betrachten noch einmal die folgende Optimierungsaufgabe im $\mathbb{R}^2$:

$$x \in \mathbb{R}^2 : \quad f(x) = x_1^2 + x_2^2 - x_1 - x_2 \to \min!, \quad g(x) = 1 + x_1 \leq 0.$$

Beide Funktionen f, g sind konvex und der zulässige Bereich erfüllt $M^o \neq \emptyset$. Die Lagrange-Funktion ist

$$L(x, y) = x_1^2 + x_2^2 - x_1 - x_2 + y(1 + x_1).$$

Wir konstruieren eine Lösung mit Hilfe der Karush-Kuhn-Tucker-Bedingungen:

$$\begin{array}{rll} i) & x \in M : & 1 + x_1 \leq 0, \\ ii) & yg(x) = 0 : & y(1 + x_1) = 0, \;\; y \geq 0, \\ iii) & \nabla_x L(x, y) = 0 : & 2x_1 - 1 + y = 0, \\ & & 2x_2 - 1 = 0. \end{array}$$

Aus diesen Gleichungen erhalten wir die Lösung $\hat{x} = (-1, 1/2)^T$, $\hat{y} = 3$. Satz 5.4 liefert dann $\hat{x} = (-1, 1/2)^T$ als globales Minimum der Optimierungsaufgabe.

Für konvexe Optimierungsaufgaben ist nach Lemma 5.8 im Fall $M^o \neq \emptyset$ jeder zulässige Punkt $x \in M$ regulär. Sind die Restriktionen affin-linear,

$$g(x) = Ax - b,$$

so kann auf die Slater-Bedingung $M^o \neq \emptyset$ verzichtet werden.

Lemma 5.9: *Mit einer Matrix $A \in \mathbb{R}^{m \times n}$ und einem Vektor $b \in \mathbb{R}^m$ sei die zulässige Menge gegeben durch*

$$M = \{x \in \mathbb{R}^n \mid Ax \leq b\}.$$

Dann ist jeder Punkt $x \in M$ regulär.

Beweis: Zu zeigen ist wieder $D_1(x) \subset \overline{D(x)}$. Sei $d \in D_1(x)$. Mit den Zeilenvektoren a^i, $i = 1, \ldots, m$, von A ist $g_i(x) = a^{iT} \cdot x - b_i$ und $\nabla g_i(x) = a^{iT}$. Es gilt

$$a^{iT} \cdot d \leq 0, \;\; i \in I(x), \qquad a^{iT} \cdot x < b_i, \;\; i \notin I(x).$$

Also gibt es $\varepsilon(d,x) > 0$, so dass für $0 < \varepsilon \leq \varepsilon(d,x)$ gilt:

$$\begin{aligned} \varepsilon a^{iT} \cdot d &\leq 0, \quad i \in I(x), \\ a^{iT} \cdot x + \varepsilon a^{iT} \cdot d &\leq b_i, \quad i \notin I(x). \end{aligned}$$

Dies impliziert

$$a^{iT} \cdot x + \varepsilon a^{iT} \cdot d \leq b_i, \quad i = 1, \ldots, m, \; 0 < \varepsilon \leq \varepsilon(d,x),$$

bzw. $A(x + \varepsilon d) \leq b$. Also ist $x + \varepsilon d \in M$ für $0 < \varepsilon \leq \varepsilon(d,x)$ und somit $d \in D(x)$.
Q.E.D.

Korollar 5.2: *Es sei f konvex und g affin-linear. Dann sind die Karush-Kuhn-Tucker-Bedingungen notwendig und hinreichend dafür, dass ein Punkt $\hat{x} \in M$ Lösung des Optimierungsproblems (P) ist.*

Beweis: Nach Lemma 5.9 ist jedes $x \in M$ regulär. Der Beweis verläuft nun analog zu dem von Satz 5.3, da dort die Slater-Bedingung $M^o \neq \emptyset$ lediglich für die Anwendbarkeit von Lemma 5.7 benötigt wurde. Q.E.D.

Bemerkung 5.2: Für Aufgabe (P) mit affin-linearer Restriktion lässt sich auch der *globale* Karush-Kuhn-Tucker-Satz (Satz 4.3) ohne Verwendung der Slater-Bedingung beweisen.

Eine für Anwendungen wichtige Unterklasse der konvexen Optimierungsaufgaben sind die sog. „quadratischen Programmierungsaufgaben“.

Definition 5.3: *Sei $C \in \mathbb{R}^{n\times n}$ eine (symmetrische) positiv semi-definite Matrix und $c \in \mathbb{R}^n$. Eine Optimierungsaufgabe mit der Zielfunktion*

$$f(x) = x^T \cdot Cx + c^T \cdot x \to \min!$$

und affin-linearer Restriktion

$$g(x) = Ax - b \leq 0$$

heißt „quadratisch“ (genauer „linear-quadratisch“). In diesem Fall hat die Lagrange-Funktion die Gestalt

$$L(x, y) = x^T \cdot Cx + c^T \cdot x + y(Ay - b).$$

Korollar 5.3: *Ein Punkt $\hat{x} \in \mathbb{R}^n$ ist genau dann Minimallösung der quadratischen Programmierungsaufgabe, wenn es einen Vektor $\hat{y} \in \mathbb{R}^m_+$ gibt, so dass gilt:*

$$A\hat{x} \leq b, \tag{5.3.24}$$

$$\hat{y}^T \cdot (A\hat{x} - b) = 0, \tag{5.3.25}$$

$$2C\hat{x} + c + A^T\hat{y} = 0. \tag{5.3.26}$$

Beweis: Konsequenz von Korollar 5.2 (Übungsaufgabe). Q.E.D.

5.4 Übungsaufgaben

Aufgabe 5.1: a) Man rekapituliere für eine Funktion $f : D \subset \mathbb{R}^n \to \mathbb{R}$ auf einer offenen, konvexen Menge $D \subset \mathbb{R}^n$ die Eigenschaften „konvex“, „strikt konvex“, „konkav“, „strikt Konkav“ und gebe entsprechende Beispiele an.

b) Man zeige, dass eine konvexe Funktion notwendig stetig ist.

c) Man zeige, dass für konvexe Funktionen f_i $(i = 1, \dots, m)$ auch $\sum_{i=1}^m f_i$ konvex ist.

d) Man zeige, dass die Menge der Minima einer konvexen Funktion konvex ist.

e) Man zeige, dass für eine konvexe Funktion f die Niveaumenge $\{x \in D \,|\, f(x) \leq a\}$ für jedes $a \in \mathbb{R}$ konvex ist.

Aufgabe 5.2: Man verifiziere, dass die quadratische Funktion

$$f(x) = c^T \cdot x + x^T \cdot Cx, \quad x \in \mathbb{R}^n,$$

mit (symmetrischer) positiv semi-definiter Matrix $C \in \mathbb{R}^{n\times n}$ und Vektor $c \in \mathbb{R}^n$ konvex ist. Unter welcher Zusatzbedingung ist die Funktion sogar *strikt* konvex?

Aufgabe 5.3: Was besagt der lokale Karush-Kuhn-Tucker-Satz der konvexen Optimierung im Fall der *linearen* Aufgabe

$$(P) \qquad x \in \mathbb{R}^n: \quad c^T \cdot x \to \min \quad Ax \geq b\,?$$

Man zeige, dass die Langrange-Multiplikatoren Lösungen der dualen Aufgabe

$$(P^*) \qquad y \in \mathbb{R}^m: \quad b^T \cdot y \to \max! \quad A^T y = c, \;\; y \geq 0,$$

sind, und dass die erste KKT-Bedingung (i) mit dem Gleichgewichtssatz der linearen Optimierung äquivalent ist.

Aufgabe 5.4: Sei $f: \mathbb{R}^n \to \mathbb{R}$ eine konvexe Funktion. Man zeige, dass zu jedem $x \in \mathbb{R}^n$ ein $p \in \mathbb{R}^n$ existiert, mit dem gilt:

$$f(y) \geq f(x) - (y-x)^T \cdot p, \quad y \in \mathbb{R}^b.$$

Im Falle $f \in C^1(\mathbb{R}^n)$ ist dieses p eindeutig bestimmt. (Hinweis: Man veranschauliche sich die Aussage grafisch und wende dann in naheliegender Weise den Satz von der trennenden Hyperebene für konvexe Mengen an.)

Aufgabe 5.5: Man zeige, dass das Sattelpunktkriterium

$$L^*(\hat{x}) = L_*(\hat{y})$$

für ein Paar $(\hat{x}, \hat{y}) \in \mathbb{R}^n \times \mathbb{R}^m_+$ äquivalent ist zu der Beziehung

$$\max_{y \in \mathbb{R}^m_+} \min_{x \in \mathbb{R}^n} L(x,y) = \min_{x \in \mathbb{R}^n} \max_{y \in \mathbb{R}^m_+} L(x,y).$$

Aufgabe 5.6: Man löse die nichtlineare Programmierungsaufgabe

$$x \in \mathbb{R}^2: \quad x_1^2 + x_1 x_2 + x_2^2 - x_1 - x_2 \to \min! \quad 1 + x_1 \geq 0, \;\; 1 + x_2 \leq 0,$$

mit Hilfe der Dualitätstheorie („indirekte Methode“). Wie lautet die zugehörige *duale* Aufgabe?

Aufgabe 5.7: Gegeben sei die Standardaufgabe der linearen Programmierung

$$(I) \qquad x \in \mathbb{R}^n: \quad c^T \cdot x \to \max! \quad Ax \leq b, \;\; x \geq 0.$$

a) Man schreibe die *lineare* Programmierungsaufgabe (I) in der Standardform einer allgemeinen *nichtlinearen* Programmierungsaufgabe (P).

b) Wie lautet die dazu *duale* Aufgabe (P^*)?

c) Man verifiziere, dass die zu (P^*) duale Aufgabe (P^{**}) in diesem Fall wieder äquivalent zu (P) ist.

d) Was besagt der globale Karush-Kuhn-Tucker-Satz für die Aufgabe (I) (im regulären Fall $M^o \neq \emptyset$)?

Aufgabe 5.8: Was besagt der lokale Karush-Kuhn-Tucker-Satz der konvexen Optimierung im Fall der *linearen* Aufgabe

$$(P) \qquad x \in \mathbb{R}^n : \quad c^T \cdot x \to \min \quad Ax \geq b\,?$$

Man zeige, dass die Langrange-Multiplikatoren Lösungen der dualen Aufgabe

$$(P^*) \qquad y \in \mathbb{R}^m : \quad b^T \cdot y \to \max! \quad A^T y = c, \;\; y \geq 0,$$

sind, und dass die erste KKT-Bedingung (i) mit dem Gleichgewichtssatz der linearen Optimierung äquivalent ist.

Aufgabe 5.9: Man löse die quadratische Optimierungsaufgabe

$$x \in \mathbb{R}^2 : \quad x_1^2 + x_1 x_2 + x_2^2 - x_1 - x_2 \to \min! \quad 1 + x_1 \geq 0, \;\; 1 + x_2 \leq 0,$$

mit Hilfe der Karush-Kuhn-Tucker-Theorie.

Aufgabe 5.10: Man beweise die folgende Modifikation des lokalen Karush-Kuhn-Tucker-Satzes für Optimierungsaufgaben mit Vorzeichenrestriktionen:

$$(P) \qquad x \in \mathbb{R}^n : \quad f(x) \to \min \quad g(x) \leq 0, \;\; x \geq 0.$$

Sei $\hat{x}$ ein lokales Minimum von (P), und sei $\hat{x}$ regulär bzgl. der Restriktion $g(x) \leq 0$. Dann existiert ein $\hat{y} \in \mathbb{R}^m_+$, mit dem gilt

$$\begin{array}{ll} i) & \hat{y}^T \cdot g(\hat{x}) = 0, \\ ii) & \nabla_x L(\hat{x}, \hat{y}) \geq 0\,. \end{array}$$

Aufgabe 5.11: Sei $\hat{x}$ ein zulässiger Punkt der Optimierungsaufgabe

$$(P) \qquad x \in \mathbb{R}^n : \quad f(x) \to \min \quad g(x) \leq 0,$$

mit stetig differenzierbaren Funktionen $f : \mathbb{R}^n \to \mathbb{R}$ und $g : \mathbb{R}^n \to \mathbb{R}^m$. Man zeige, dass $\hat{x}$ regulär bzgl. der Restriktion $g(x) \leq 0$ ist, wenn die Gradienten $\nabla g_i(x)$, $i \in I(x)$, linear unabhängig sind. Ist diese Bedingung auch notwendig für die Regularität von $\hat{x}$? (Hinweis: Man versuche, den Beweis von Lemma 5.8 des Textes geeignet zu modifizieren.)

Aufgabe 5.12: Gegeben sei die quadratische Optimierungsaufgabe

$$x \in \mathbb{R}^n : \qquad x^T \cdot Cx + c^T \cdot x \to \min! \quad Ax = b, \;\; x \geq 0,$$

mit $C \in \mathbb{R}^{n \times n}$ symmetrisch, positiv definit, $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$ und $b \in \mathbb{R}^m$. Man zeige, mit Hilfe des Karush.Kuhn-Tucker-Satzes, dass ein $\hat{x} \in \mathbb{R}^n_+$ genau dann Minimallösung ist, wenn es Vektoren $\hat{u} \in \mathbb{R}^n_+$ und $\hat{y} \in \mathbb{R}^m$ gibt, mit denen gilt:

$$\begin{array}{ll} i) & A\hat{x} = b, \\ ii) & -2C\hat{x} + \hat{u} - A^T \hat{y} = 0, \\ iii) & \hat{u}^T \cdot \hat{x} = 0. \end{array}$$

Aufgabe 5.13: Man zeige, dass der Punkt $\hat{x} = (4,4)^T$ eine Lösung der folgenden Optimierungsaufgabe ist:

$$\begin{aligned} x \in \mathbb{R}^2: \quad & 2x_1^2 + x_2^2 - 48x_1 - 40x_2 \to \min! \\ & x_1 \geq 0, \quad x_2 \geq 0, \quad x_1 + x_2 \leq 8, \quad x_1 \leq 6, \quad x_1 + 3x_2 \leq 18. \end{aligned}$$

Aufgabe 5.14: Gegeben sei die quadratische Optimierungsaufgabe

$$Q(x) = x^T \cdot Cx + c^T \cdot x \to \min! \quad Ax \leq b.$$

Seien $x^0 \in \mathbb{R}^n$ und $y^0 \in \mathbb{R}^m_+$ Punkte mit den Eigenschaften

$$Ax^0 \leq b, \quad 2Cx^0 + c + A^T y^0 = 0.$$

Man zeige, dass dann für den Optimalwert $Q_{\min} = \inf_{x \in M} Q(x)$ die folgende Einschließung gilt:

$$Q(x^0) + y^{0T} \cdot (Ax^0 - b) \leq Q_{\min}(x^0) \leq Q(x^0).$$