Kapitel 4

Optimalitätskriterien

Als Optimalitätskriterien bezeichnet man notwendige oder hinreichende Bedingungen dafür, dass ein $\mathbf{x}_0 \in \Omega \subset \mathbb{R}^n$ Lösung eines Optimierungsproblems ist. Diese Kriterien besitzen sowohl unter theoretischen als auch unter numerischen Aspekten Bedeutung. In diesem Kapitel werden lokale und globale Optimalitätskriterien hergeleitet.

4.1 Einleitung

Ein Gegenstand dieses Kapitels ist die Verallgemeinerung der Methode der Lagrangeschen Multiplikatoren auf Probleme mit Bedingungsungleichungen und beschränkten Variablen. Es wird der Zusammenhang zwischen der Lösung eines Optimierungsproblems und dem Sattelpunkt eines Lagrangeschen Funktionals hergestellt.

Wir betrachten zunächst die Lagrangesche Methode für ein Optimierungsproblem mit Nebenbedingungen in Gleichungsform

$$\begin{aligned} \{f(\mathbf{x}) \ : \mathbf{x} \in \Omega \subset \mathbb{R}^n\} &\rightarrow \min !, \\ \Omega &= \{\mathbf{x} \in \mathbb{R}^n \ : \ h_i(\mathbf{x}) = 0,\ i \in \{1, \ldots, p\},\ p < n\}. \end{aligned} \tag{4.1}$$

Seien $f(\mathbf{x})$ in Ω und $h_i(\mathbf{x})$, $i = 1, \ldots, p$, in $\mathbb{R}^n$ differenzierbar. Für beliebiges festes $\mathbf{x} \in \Omega$ werden die Matrizen

$$H_0(\mathbf{x}) := (\nabla h_1(\mathbf{x}), \ldots, \nabla h_p(\mathbf{x}))^T \in \mathbb{R}^{p \times n}, \quad H(\mathbf{x}) := \begin{pmatrix} H_0(\mathbf{x}) \\ \nabla f(\mathbf{x})^T \end{pmatrix} \in \mathbb{R}^{(p+1) \times n}$$

definiert.

Für das Problem (4.1) wird die Lagrange–Funktion

$$L(\mathbf{x}, \boldsymbol{\lambda}) = \lambda_0 f(\mathbf{x}) + \sum_{i=1}^{p} \lambda_i h_i(\mathbf{x}), \quad \mathbf{x} \in \Omega,\ \boldsymbol{\lambda} \in \mathbb{R}^{p+1},$$

eingeführt. Die Zahlen $\lambda_0, \ldots, \lambda_p$ nennt man Lagrangesche Multiplikatoren. Man kann beispielsweise folgendes Optimalitätskriterium für Problem (4.1) beweisen.

Satz 4.1 *Seien $\mathbf{x}_0 \in \Omega$ und $rg(H_0(\mathbf{x}_0)) = rg(H(\mathbf{x}_0))$. Dann gilt: besitzt $f(\mathbf{x})$ in $\mathbf{x}_0$ ein lokales Minimum bezüglich Ω, so existiert ein $\boldsymbol{\lambda}^{(0)} = \left(\lambda_0^{(0)}, \ldots, \lambda_p^{(0)}\right)^T \in \mathbb{R}^{p+1}$, mit $\lambda_0^{(0)} \neq 0$ und*

$$\nabla_{\mathbf{x}} L(\mathbf{x}_0, \boldsymbol{\lambda}^{(0)}) = \lambda_0^{(0)} \nabla f(\mathbf{x}_0) + \sum_{i=1}^{p} \lambda_i^{(0)} \nabla h_i(\mathbf{x}_0) = \mathbf{0}. \tag{4.2}$$

Verallgemeinerungen dieser Aussage auf den Fall, dass die Matrizen unterschiedlichen Rang besitzen, sind möglich. Mit Hilfe des Gleichungssystems (4.2) hat man ein Kriterium zur Bestimmung von Punkten, in denen $f(\mathbf{x})$ ein lokales Minimum annehmen kann. Sofern dies möglich ist, berechnet man alle Lösungen $(\mathbf{x}, \boldsymbol{\lambda})$ von (4.2). Im allgemeinen sind jedoch nicht alle Lösungen auch lokale Extrempunkte.

Das heißt, man erweitert das gegebene Problem so, dass

- man für die Lösung des erweiterten Problems Standardkriterien, zum Beispiel dass Ableitungen verschwinden, für ein Minimum hat,
- die Lösung des erweiterten Problems Rückschlüsse auf die Lösung des gegebenen Problems zulässt.

4.2 Lokale Minima für Optimierungsprobleme ohne Einschränkungen an das zulässige Gebiet

Wir betrachten das Optimierungsproblem

$$z = \min\{f(\mathbf{x}) \ : \ \mathbf{x} \in \Omega\} \tag{4.3}$$

mit $f \in C^1(\mathbb{R}^n)$ (an sich reicht $f \in C^1(\tilde{\Omega})$ mit $\overline{\Omega} \subset \tilde{\Omega}$, $\tilde{\Omega}$ offen). Für diese Problem sollen im folgenden lokale Optimalitätskriterien hergeleitet werden.

Zunächst werden spezielle konvergente Folgen betrachtet.

Definition 4.2 Konvergent gegen $\mathbf{x}_0$ in Richtung $\mathbf{y}$, gerichtet konvergent. Eine konvergente Folge $\{\mathbf{x}_k\}_{k\in\mathbb{N}}$, $\mathbf{x}_k \in \mathbb{R}^n$, mit $\mathbf{x}_k \to \mathbf{x}_0$ heißt konvergent gegen $\mathbf{x}_0$ in Richtung $\mathbf{y} \in \mathbb{R}^n$ oder gerichtet konvergent, wenn gilt

$$\lim_{k\to\infty} \frac{\mathbf{x}_k - \mathbf{x}_0}{\|\mathbf{x}_k - \mathbf{x}_0\|_2} = \mathbf{y}, \quad \|\mathbf{y}\|_2 = 1.$$

Die Schreibweise ist

$$\mathbf{x}_k \xrightarrow{\mathbf{y}} \mathbf{x}_0.$$

□

Beispiel 4.3 Sei $\{\mathbf{x}_k\}_{k\in\mathbb{N}}$ mit

$$\mathbf{x}_k = \left(\frac{1}{k}, \frac{1}{k}\right)^T.$$

Dann ist $\mathbf{x}_0 = (0,0)^T$ und es gilt

$$\lim_{k\to\infty} \frac{\mathbf{x}_k}{\|\mathbf{x}_k\|_2} = \lim_{k\to\infty} \frac{k}{\sqrt{2}} \begin{pmatrix} \frac{1}{k} \\ \frac{1}{k} \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \mathbf{y}.$$

□

Man kann aus jeder gegen $\mathbf{x}_0$ konvergenten Folge $\{\mathbf{x}_k\}_{k\in\mathbb{N}}$ eine gerichtet konvergente Teilfolge auswählen, sofern unendlich viele Elemente der Folge ungleich $\mathbf{x}_0$ sind. Sei $\mathbf{x}_k \neq \mathbf{x}_0$ für $k \geq k_0$, dann sind alle Glieder der Folge

$$\mathbf{y}_k := \frac{\mathbf{x}_k - \mathbf{x}_0}{\|\mathbf{x}_k - \mathbf{x}_0\|_2}, \quad k \geq k_0,$$

Elemente der kompakten Menge $\{\mathbf{x} \in \mathbb{R}^n \ : \ \|\mathbf{x}\|_2 = 1\}$. Damit besitzt $\{\mathbf{y}_k\}_{k\in\mathbb{N}}$ einen Häufungspunkt in dieser Menge und man findet in $\{\mathbf{y}_k\}_{k\in\mathbb{N}}$ eine konvergente Teilfolge.

Die Aussage des folgenden Lemmas kann in gewisser Weise als Verallgemeinerung der Richtungsableitung aufgefasst werden.

Lemma 4.4 *Die Folge $\{\mathbf{x}_k\}_{k\in\mathbb{N}}$ sei konvergent gegen $\mathbf{x}_0$ in Richtung $\mathbf{y}$. Dann gilt:*

1. *Ist $f:\mathbb{R}^n\to\mathbb{R}$ in $\mathbf{x}_0$ stetig differenzierbar, so folgt*

$$\lim_{k\to\infty}\frac{f(\mathbf{x}_k)-f(\mathbf{x}_0)}{\|\mathbf{x}_k-\mathbf{x}_0\|_2}=\mathbf{y}^T\nabla f(\mathbf{x}_0).$$

2. *Ist $f:\mathbb{R}^n\to\mathbb{R}$ in $\mathbf{x}_0$ zweimal stetig differenzierbar, so gilt*

$$\lim_{k\to\infty}\frac{f(\mathbf{x}_k)-f(\mathbf{x}_0)-(\mathbf{x}_k-\mathbf{x}_0)^T\nabla f(\mathbf{x}_0)}{\|\mathbf{x}_k-\mathbf{x}_0\|_2^2}=\frac{1}{2}\mathbf{y}^T H_f(\mathbf{x}_0)\mathbf{y}.$$

Beweis: Es wird nur die erste Aussage bewiesen, der Beweis der zweiten Aussage ist analog.

Da $f(\mathbf{x})$ in $\mathbf{x}_0$ differenzierbar ist, also insbesondere alle Richtungsableitungen existieren, folgt unter Nutzung der Definition der Richtungsableitung, dass

$$\lim_{\mathbf{x}\to\mathbf{x}_0}\left(\frac{f(\mathbf{x})-f(\mathbf{x}_0)}{\|\mathbf{x}-\mathbf{x}_0\|_2}-\frac{(\mathbf{x}-\mathbf{x}_0)^T}{\|\mathbf{x}-\mathbf{x}_0\|_2}\nabla f(\mathbf{x}_0)\right)=0.$$

Außerdem existiert nach Voraussetzung der Grenzwert

$$\lim_{\mathbf{x}_k\to\mathbf{x}_0}\frac{(\mathbf{x}_k-\mathbf{x}_0)^T}{\|\mathbf{x}_k-\mathbf{x}_0\|_2}\nabla f(\mathbf{x}_0)=\mathbf{y}^T\nabla f(\mathbf{x}_0).$$

Damit folgt

$$\lim_{\mathbf{x}_k\to\mathbf{x}_0}\frac{f(\mathbf{x}_k)-f(\mathbf{x}_0)}{\|\mathbf{x}_k-\mathbf{x}_0\|_2}=\lim_{\mathbf{x}_k\to\mathbf{x}_0}\frac{(\mathbf{x}_k-\mathbf{x}_0)^T}{\|\mathbf{x}_k-\mathbf{x}_0\|_2}\nabla f(\mathbf{x}_0)=\mathbf{y}^T\nabla f(\mathbf{x}_0).$$

■

Definition 4.5 Tangentenkegel. Der Tangentenkegel $T(\mathbf{x}_0)$ an die Menge Ω im Punkt $\mathbf{x}_0\in\Omega$ ist gegeben durch

$$T(\mathbf{x}_0):=\left\{\lambda\mathbf{y}\ :\ \|\mathbf{y}\|_2=1,\ \exists\{\mathbf{x}_k\}_{k\in\mathbb{N}}\in\Omega,\ \mathbf{x}_k\xrightarrow{\mathbf{y}}\mathbf{x}_0,\ \lambda\ge 0\right\}.$$

□

Der Tangentenkegel $T(\mathbf{x}_0)$ hat nur etwas mit dem zulässigen Gebiet Ω zu tun und nicht mit der zu minimierenden Funktion $f(\mathbf{x})$. Er beschreibt die Richtungen, aus denen man sich $\mathbf{x}_0$ mit einer Folge annähern kann, deren Glieder alle in Ω liegen. Der Begriff des Tangentenkegels ist eine Verallgemeinerung der Tangetialhyperebene. Die Menge $T(\mathbf{x}_0)$ ist ein Kegel. Für jeden inneren Punkt $\mathbf{x}_0\in\Omega$ gilt $T(\mathbf{x}_0)=\mathbb{R}^n$. Für isolierte Punkte setzen wir $T(\mathbf{x}_0)=\{\mathbf{0}\}$.

Beispiel 4.6 Sei

$$\Omega=\{\mathbf{x}\in\mathbb{R}^2\ :\ x_1^2+x_2^2=1,\ x_1\ge 0,\ 0<x_2<2\}.$$

Für den Punkt $\mathbf{x}_0=(0,1)^T\in\Omega$ erhält man, direkt aus der geometrischen Anschauung,

$$T(\mathbf{x}_0)=\left\{\mathbf{y}\in\mathbb{R}^2\ :\ \mathbf{y}=\lambda\begin{pmatrix}1\\0\end{pmatrix},\ \lambda\ge 0\right\}.$$

Analytisches Nachrechnen: Übungsaufgabe □

Lemma 4.7 *Für jeden Punkt $\mathbf{x}_0\in\Omega$ ist der Tangentenkegel $T(\mathbf{x}_0)$ abgeschlossen.*

Beweis: Siehe Literatur. ■

Der folgende Satz gibt ein notwendiges Kriterium für die Existenz eines lokalen Minimums von $f(\mathbf{x})$ bezüglich Ω an. Das lokale Minimum kann auch auf dem Rand von Ω liegen. Beachte, dass $f(\mathbf{x})$ nach Voraussetzung auf dem Rand von Ω stetig differenzierbar ist.

Satz 4.8 *Sei $f : \mathbb{R}^n \to \mathbb{R}$ in $\mathbf{x}_0 \in \Omega$ stetig differenzierbar. Besitzt $f(\mathbf{x})$ in $\mathbf{x}_0$ ein lokales Minimum bezüglich Ω, dann gilt*

$$\mathbf{y}^T \nabla f(\mathbf{x}_0) \geq 0 \quad \forall\, \mathbf{y} \in T(\mathbf{x}_0).$$

Beweis: Sei $\mathbf{y} \in T(\mathbf{x}_0)$. Für $\mathbf{y} = \mathbf{0}$ ist $\mathbf{y}^T \nabla f(\mathbf{x}_0) = 0$. Ansonsten existieren ein $\lambda \in \mathbb{R}, \lambda > 0$, und ein $\mathbf{y}^* \in T(\mathbf{x}_0)$, $\|\mathbf{y}^*\|_2 = 1$, mit $\mathbf{y} = \lambda \mathbf{y}^*$. Ferner sei $\{\mathbf{x}_k\}_{k\in\mathbb{N}} \subset \Omega$ eine Folge mit $\mathbf{x}_k \xrightarrow{\mathbf{y}^*} \mathbf{x}_0$.

Da $f(\mathbf{x})$ in $\mathbf{x}_0$ ein lokales Minimum bezüglich Ω annimmt, gilt für hinreichend große k die Ungleichung $f(\mathbf{x}_k) \geq f(\mathbf{x}_0)$. Daraus folgt unter Beachtung von Lemma 4.4, 1)

$$\lim_{k\to\infty} \frac{f(\mathbf{x}_k) - f(\mathbf{x}_0)}{\|\mathbf{x}_k - \mathbf{x}_0\|_2} = (\mathbf{y}^*)^T \nabla f(\mathbf{x}_0) \geq 0$$

und damit auch $\mathbf{y}^T \nabla f(\mathbf{x}_0) \geq 0$. ■

Aus diesem Satz folgt nicht, dass $f(\mathbf{x})$ in $\mathbf{x}_0$ ein lokales Minimum nur annehmen kann, wenn $\nabla f(\mathbf{x}_0) = \mathbf{0}$ gilt. Gilt jedoch $\nabla f(\mathbf{x}_0) = \mathbf{0}$, dann hat man ein zweites notwendiges Kriterium.

Satz 4.9 *Sei $f : \mathbb{R}^n \to \mathbb{R}$ in $\mathbf{x}_0 \in \Omega$ zweimal stetig differenzierbar. Besitzt $f(\mathbf{x})$ in $\mathbf{x}_0$ ein lokales Minimum bezüglich Ω und gilt $\nabla f(\mathbf{x}_0) = \mathbf{0}$, dann folgt*

$$\mathbf{y}^T H_f(\mathbf{x}_0)\mathbf{y} \geq 0 \quad \forall\, \mathbf{y} \in T(\mathbf{x}_0).$$

Beweis: Der Beweis ist analog zum Beweis von Satz 4.8, wobei jedoch jetzt Lemma 4.4, 2) verwendet wird. Aus $(\mathbf{x}_k - \mathbf{x}_0)^T \nabla f(\mathbf{x}_0) = 0$ und $f(\mathbf{x}_k) \geq f(\mathbf{x}_0)$ für hinreichend große Indizes k folgt dann

$$\lim_{k\to\infty} \frac{f(\mathbf{x}_k) - f(\mathbf{x}_0) - (\mathbf{x}_k - \mathbf{x}_0)^T \nabla f(\mathbf{x}_0)}{\|\mathbf{x}_k - \mathbf{x}_0\|_2^2} = \frac{1}{2} (\mathbf{y}^*)^T H_f(\mathbf{x}_0)\mathbf{y}^* \geq 0.$$

■

Der folgende Satz enthält eine hinreichende Bedingung für die Existenz eines eigentlichen lokalen Minimums von $f(\mathbf{x})$ bezüglich Ω.

Satz 4.10 *Sei $f : \mathbb{R}^n \to \mathbb{R}$ in $\mathbf{x}_0 \in \Omega$ zweimal stetig differenzierbar. Gelten $\nabla f(\mathbf{x}_0) = \mathbf{0}$ und $\mathbf{y}^T H_f(\mathbf{x}_0)\mathbf{y} > 0$ für alle $\mathbf{y} \in T(\mathbf{x}_0), \mathbf{y} \neq \mathbf{0}$, dann besitzt $f(\mathbf{x})$ in $\mathbf{x}_0$ ein isoliertes lokales Minimum bezüglich Ω.*

Beweis: Indirekter Beweis. Wir nehmen an, dass $f(\mathbf{x})$ in $\mathbf{x}_0$ kein isoliertes lokales Minimum bezüglich Ω besitzt. Dann existiert eine Folge $\{\mathbf{x}_k\}_{k\in\mathbb{N}} \subset \Omega$ mit $\mathbf{x}_k \to \mathbf{x}_0$, mit unendlich vielen Elementen, die ungleich $\mathbf{x}_0$ sind, und $f(\mathbf{x}_k) \leq f(\mathbf{x}_0), k \geq 1$. Es existiert eine Teilfolge von $\{\mathbf{x}_k\}$, die gegen ein $\tilde{\mathbf{y}}$ gerichtet konvergent ist. Ohne Beschränkung der Allgemeinheit kann auch $\mathbf{x}_k \xrightarrow{\tilde{\mathbf{y}}} \mathbf{x}_0$ angenommen werden. Dann folgt aus Lemma 4.4

$$\lim_{k\to\infty} \frac{f(\mathbf{x}_k) - f(\mathbf{x}_0)}{\|\mathbf{x}_k - \mathbf{x}_0\|_2^2} = \frac{1}{2}\tilde{\mathbf{y}}^T H_f(\mathbf{x}_0)\tilde{\mathbf{y}} \leq 0$$

mit $\tilde{\mathbf{y}} \in T(\mathbf{x}_0)$, im Widerspruch zur Voraussetzung. ■

Für jeden inneren Punkt $\mathbf{x}_0 \in \Omega$ gilt $T(\mathbf{x}_0) = \mathbb{R}^n$ und aus den Aussagen der Sätze 4.8 – 4.10 folgen die bekannten notwendigen und hinreichenden Bedingungen für die Existenz lokaler Minima einer Funktion $f : \mathbb{R}^n \to \mathbb{R}$.

4.3 Lokale Minima für Optimierungsprobleme, bei denen das zulässige Gebiet durch Ungleichungen gegeben ist

In diesem Abschnitt wird das Optimierungsproblem

$$z = \min\{f(\mathbf{x}) \ : \ \mathbf{x} \in \Omega\} \quad \text{mit } \Omega = \{\mathbf{x} \in \mathbb{R}^n \ : \ \mathbf{g}(\mathbf{x}) \le \mathbf{0}\} \tag{4.4}$$

mit $f \in C^1(\mathbb{R}^n)$ und $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^m$, $\mathbf{g} \in C^1(\mathbb{R}^n)$ untersucht. Unter Verwendung der Resultate aus Abschnitt 4.2 werden nun lokale Optimierungskriterien für (4.4) hergeleitet. Dazu wird eine lokale Theorie von Lagrange–Multiplikatoren entwickelt.

Definition 4.11 Aktive Nebenbedingung. Eine Nebenbedingung $g_i(\mathbf{x}) \le 0$, $i = 1, \ldots, m$, wird im Punkt $\mathbf{x}_0 \in \Omega$ aktiv genannt, wenn gilt $g_i(\mathbf{x}_0) = 0$. Bezeichnung:

$$I_0 := \{i \in \{1, \ldots, m\} \ : \ g_i(\mathbf{x}_0) = 0\}\,.$$

□

Sei $I_0 \ne \emptyset$. Die in $\mathbf{x}_0$ aktiven Nebenbedingungen werden nun durch affine Funktionen ersetzt

$$(\nabla g_i(\mathbf{x}_0))^T (\mathbf{x} - \mathbf{x}_0)\,, \ \mathbf{x} \in \mathbb{R}^n, \ i \in I_0,$$

beziehungsweise in Matrixnotation

$$(\nabla \mathbf{g}_{I_0}(\mathbf{x}_0))^T (\mathbf{x} - \mathbf{x}_0)\,, \quad \nabla \mathbf{g}_{I_0}(\mathbf{x}_0) \in \mathbb{R}^{n \times |I_0|}.$$

Der Gradient einer vektorwertigen Funktion $\mathbf{g}(\mathbf{x})$ ist wie folgt definiert:

$$\nabla \mathbf{g}(\mathbf{x}) = \begin{pmatrix} (g_1(\mathbf{x}))_{x_1} & \cdots & (g_m(\mathbf{x}))_{x_1} \\ \cdots & \cdots & \cdots \\ (g_1(\mathbf{x}))_{x_n} & \cdots & (g_m(\mathbf{x}))_{x_n} \end{pmatrix} \in \mathbb{R}^{n \times m}.$$

Ausgehend von der Anschauung, könnte man die Menge

$$\left\{\mathbf{x} \ : \ (\nabla g_{I_0}(\mathbf{x}_0))^T (\mathbf{x} - \mathbf{x}_0) \le \mathbf{0}, \ \mathbf{g}_{I \setminus I_0}(\mathbf{x}_0) + \left(\nabla \mathbf{g}_{I \setminus I_0}(\mathbf{x}_0)\right)^T (\mathbf{x} - \mathbf{x}_0) \le \mathbf{0}\right\}$$

als eine lineare Approximation der Menge Ω im Punkt $\mathbf{x}_0$ ansehen. Dies ist jedoch in gewissen ausgearteten Punkten nicht zutreffend, siehe Beispiel 4.15

Definition 4.12 Linearisierter Kegel. Die Menge

$$K(\mathbf{x}_0) := \{\mathbf{y} \in \mathbb{R}^n \ : \ (\nabla \mathbf{g}_{I_0}(\mathbf{x}_0))^T \mathbf{y} \le \mathbf{0}\}$$

heißt linearisierter Kegel von Ω im Punkt $\mathbf{x}_0 \in \Omega$. Für $I_0 = \emptyset$ setzen wir $K(\mathbf{x}_0) = \mathbb{R}^n$.

□

Lemma 4.13 *Es gilt* $T(\mathbf{x}_0) \subseteq K(\mathbf{x}_0)$.

Beweis: Für $I_0 = \emptyset$ gilt die Behauptung trivialerweise. Sei also $I_0 \ne \emptyset$. Der Nullvektor gehört per Definition zu beiden Mengen. Sei $\mathbf{y} \in T(\mathbf{x}_0)$, $\mathbf{y} \ne \mathbf{0}$. Dann existieren ein $\mathbf{y}^* \in T(\mathbf{x}_0)$, $\|\mathbf{y}^*\|_2 = 1$ und ein $\lambda > 0$ mit $\mathbf{y} = \lambda \mathbf{y}^*$ und eine Folge $\{\mathbf{x}_k\}_{k \in \mathbb{N}} \subset \Omega$ mit $\mathbf{x}_k \xrightarrow{\mathbf{y}^*} \mathbf{x}_0$. Wegen $g_i(\mathbf{x}_k) \le 0$ für alle i und $g_i(\mathbf{x}_0) = 0$ für alle $i \in I_0$ hat man

$$0 \ge \lim_{k \to \infty} \frac{g_i(\mathbf{x}_k) - g_i(\mathbf{x}_0)}{\|\mathbf{x}_k - \mathbf{x}_0\|_2} = \nabla g_i(\mathbf{x}_0)^T \mathbf{y}^* \quad \forall i \in I_0,$$

wobei Lemma 4.4, 1) verwendet wurde. Folglich gilt $\mathbf{y} \in K(\mathbf{x}_0)$. ■

Lemma 4.14 *Es gilt* $K_0(\mathbf{x}_0) := \{\mathbf{y} \in \mathbb{R}^n \ : \ (\nabla \mathbf{g}_{I_0}(\mathbf{x}_0))^T \mathbf{y} < \mathbf{0}\} \subseteq T(\mathbf{x}_0)$.

Beweis: Siehe Literatur, zum Beispiel [ERSD77, S. 145]. ■

Beispiel 4.15 Sei

$$\Omega = \{\mathbf{x} \in \mathbb{R}^2 \ : \ -x_1^3 + x_2 \le 0, -x_2 \le 0\},$$

d.h. die Menge Ω ist begrenzt von der positiven x_1–Achse und der Funktion x_1^3. Damit sind

$$\mathbf{g}(\mathbf{x}) = \begin{pmatrix} -x_1^3 + x_2 \\ -x_2 \end{pmatrix}, \quad (\nabla \mathbf{g}(\mathbf{x}))^T = \begin{pmatrix} -3x_1^2 & 1 \\ 0 & -1 \end{pmatrix}.$$

Wir betrachten den Punkt $\mathbf{x}_0 = (0,0)^T$. In diesem Punkt sind beide Nebenbedingungen mit Gleichheit erfüllt, also $I_0 = \{1,2\}$. Es folgt

$$(\nabla \mathbf{g}_{I_0}(\mathbf{x}_0))^T = \begin{pmatrix} 0 & 1 \\ 0 & -1 \end{pmatrix}, \quad \Longrightarrow \quad K(\mathbf{x}_0) = \begin{pmatrix} y_1 \\ 0 \end{pmatrix}, \ y_1 \in \mathbb{R}.$$

Für den Tangentialkegel gilt $T(\mathbf{x}_0) = \{\mathbf{y} \in \mathbb{R}^2 \ : \ y_1 \ge 0, y_2 = 0\}$. Es ist also $T(\mathbf{x}_0) \subsetneq K(\mathbf{x}_0)$. Dieses Beispiel zeigt insbesondere, dass man für den Punkt $\mathbf{x}_0 = (0,0)^T$ mit $K(\mathbf{x}_0)$ keine lineare Approximation von Ω erhält. □

Beispiel 4.16 Der linearisierte Kegel $K(\mathbf{x}_0)$ ist im Gegensatz zum Tangentenkegel $T(\mathbf{x}_0)$ von der analytischen Darstellung von Ω abhängig. Die Menge aus Beispiel 4.15 kann auch dargestellt werden durch

$$\Omega = \{\mathbf{x} \in \mathbb{R}^2 \ : \ -x_1^3 + x_2 \le 0, -x_2^3 \le 0\}$$

In diesem Fall hat man

$$\mathbf{g}(\mathbf{x}) = \begin{pmatrix} -x_1^3 + x_2 \\ -x_2^3 \end{pmatrix}, \quad (\nabla \mathbf{g}(\mathbf{x}))^T = \begin{pmatrix} -3x_1^2 & 1 \\ 0 & -3x_2^2 \end{pmatrix}$$

und man erhält für $\mathbf{x}_0 = (0,0)^T$, dass $I_0 = \{1,2\}$ und

$$(\nabla \mathbf{g}_{I_0}(\mathbf{x}_0))^T = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad \Longrightarrow \quad K(\mathbf{x}_0) = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

mit $y_1 \in \mathbb{R}$ und $y_2 \le 0$. □

Für den nächsten Beweis benötigen wir einen Hilfssatz.

Lemma 4.17 Alternativsatz von Gordan. *Sei $A \in \mathbb{R}^{m \times n}$, dann hat von den beiden Systemen*

$$A\mathbf{y} < \mathbf{0}$$

und

$$\mathbf{z}^T A = \mathbf{0}, \ \mathbf{z} \ge \mathbf{0}, \ \mathbf{z} \ne \mathbf{0}$$

genau eines eine Lösung.

Beweis: Literatur. ■

Seien die Voraussetzungen von Satz 4.8 erfüllt und $I_0 \ne \emptyset$. Dann folgt wegen $K_0(\mathbf{x}_0) \subseteq T(\mathbf{x}_0)$

$$(\nabla \mathbf{g}_{I_0}(\mathbf{x}_0))^T \mathbf{y} < \mathbf{0} \quad \Longrightarrow \quad \mathbf{y}^T \nabla f(\mathbf{x}_0) \ge 0 \quad \forall\, \mathbf{y} \in K_0(\mathbf{x}_0).$$

Diese Beziehung ist äquivalent damit, dass das Ungleichungssystem

$$(\nabla \mathbf{g}_{I_0}(\mathbf{x}_0))^T \mathbf{y} < \mathbf{0}, \quad \mathbf{y}^T \nabla f(\mathbf{x}_0) < 0$$

keine Lösung $\mathbf{y} \in K_0(\mathbf{x}_0)$ besitzt. Nach Lemma 4.17 besitzt dieses System genau dann keine Lösung, wenn das System

$$\eta \nabla f(\mathbf{x}_0) + \nabla \mathbf{g}_{I_0}(\mathbf{x}_0)\mathbf{z}_{I_0} = \mathbf{0}, \quad \begin{pmatrix} \eta \\ \mathbf{z}_{I_0} \end{pmatrix} \geq \mathbf{0}, \quad \eta \in \mathbb{R}_+, \tag{4.5}$$

eine nichttriviale Lösung besitzt. Hierbei ist $\mathbf{z}_{I_0}$ ein $|I_0|$–dimensionaler Vektor, dessen Einträge durch die Indizes der aktiven Nebenbedingungen (unter Beibehaltung der natürlichen Reihenfolge) bestimmt sind. Setzt man $z_i = 0$ für $i \notin I_0$, so kann man (4.5) mit $\mathbf{z} \in \mathbb{R}^m$ wie folgt formulieren:

$$\eta \nabla f(\mathbf{x}_0) + \nabla \mathbf{g}(\mathbf{x}_0)\mathbf{z} = \mathbf{0}, \ \mathbf{g}(\mathbf{x}_0) \leq \mathbf{0}, \ \underbrace{\mathbf{z}^T}_{=0, i \notin I_0} \ \underbrace{\mathbf{g}(\mathbf{x}_0)}_{=0, i \in I_0} = 0, \begin{pmatrix} \eta \\ \mathbf{z} \end{pmatrix} \geq \mathbf{0}, \ \begin{pmatrix} \eta \\ \mathbf{z} \end{pmatrix} \neq \mathbf{0}.$$

Sei $I_0 = \emptyset$ und hat $f(\mathbf{x})$ in $\mathbf{x}_0 \in \Omega$ ein lokales Minimum, das heißt es gelten $\mathbf{x}_0 \in \operatorname{int}(\Omega)$, $\mathbf{z} = \mathbf{0}$ und $\nabla f(\mathbf{x}_0) = \mathbf{0}$, dann besitzt dieses System eine Lösung $\eta > 0, \mathbf{z} = \mathbf{0}$.

Wir betrachten jetzt das Problem: Gesucht ist ein Paar $(\mathbf{x}_0, \mathbf{z}_0)^T \in \mathbb{R}^n \times \mathbb{R}^m_+(\geq 0)$, $\eta_0 \geq 0$, $(\eta_0, \mathbf{z}_0)^T \neq \mathbf{0}$, welches das System

$$\eta \nabla f(\mathbf{x}) + \nabla \mathbf{g}(\mathbf{x})\mathbf{z} = \mathbf{0}, \ \mathbf{g}(\mathbf{x}) \leq \mathbf{0}, \ \mathbf{z}^T \mathbf{g}(\mathbf{x}) = 0, \begin{pmatrix} \eta \\ \mathbf{z} \end{pmatrix} \geq \mathbf{0} \tag{4.6}$$

löst. Mit dem Lagrange–Funktional

$$L_\eta(\mathbf{x}, \mathbf{z}) = \eta f(\mathbf{x}) + \mathbf{z}^T \mathbf{g}(\mathbf{x})$$

ergibt sich die äquivalente Formulierung

$$\nabla_{\mathbf{x}} L_\eta(\mathbf{x}, \mathbf{z}) = \mathbf{0}, \ \nabla_{\mathbf{z}} L_\eta(\mathbf{x}, \mathbf{z}) \leq \mathbf{0}, \ \mathbf{z}^T \nabla_{\mathbf{z}} L_\eta(\mathbf{x}, \mathbf{z}) = 0, \ \begin{pmatrix} \eta \\ \mathbf{z} \end{pmatrix} \geq \mathbf{0}. \tag{4.7}$$

Der Zusammenhang zwischen den Problemen (4.4) und (4.6) ist wie folgt.

Satz 4.18 *Seien $f : \mathbb{R}^n \to \mathbb{R}$, $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^m$ differenzierbar und $\mathbf{x}_0 \in \Omega$ Stelle eines lokalen Minimums von $f(\mathbf{x})$ bezüglich Ω. Dann existieren ein $\mathbf{z}_0 \in \mathbb{R}^m_+$ und ein $\eta_0 \geq 0$, so dass $(\mathbf{x}_0, \mathbf{z}_0)^T$ und η_0 eine Lösung von (4.6) bzw. von (4.7) sind.*

Beweis: Literatur, Fritz John (1948). ■

Damit haben wir für den Fall $\eta_0 > 0$ ein Optimalitätskriterium gefunden. Ist jedoch $\eta_0 = 0$, so kann der Funktionswert von $f(\mathbf{x})$ in (4.6) beliebig sein. Es wird jetzt eine Regularitätsbedingung eingeführt, die sichert, dass unter den Voraussetzungen von Satz 4.18 für jede Lösung von (4.6) $\eta_0 > 0$ gilt.

Regularitätsbedingung. Seien $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^m$ differenzierbar und

$$K(\mathbf{x}_0) = T(\mathbf{x}_0). \tag{4.8}$$

Bemerkung 4.19

- Für Optimierungsprobleme mit ausschließlich affinen Nebenbedingungen ist die Regularitätsbedingung stets erfüllt.
- Die Regularitätsbedingung stellt eine Bedingung an die analytische Darstellung der Menge Ω dar, vergleiche Beispiele 4.15 und 4.16.

- Man kann andere Regularitätsbedingungen formulieren, die die obige Bedingung implizieren, aber leichter zu überprüfen sind.

□

Anstelle von (4.6) betrachten wir jetzt das Optimierungsproblem: Finde ein Paar $(\mathbf{x}_0, \mathbf{z}_0)^T \in \mathbb{R}^n \times \mathbb{R}^m_+$, welches das Problem

$$\nabla f(\mathbf{x}) + \nabla \mathbf{g}(\mathbf{x})\mathbf{z} = \mathbf{0},\ \mathbf{g}(\mathbf{x}) \leq \mathbf{0},\ \mathbf{z}^T\mathbf{g}(\mathbf{x}) = 0, \mathbf{z} \geq \mathbf{0} \tag{4.9}$$

löst. Man bezeichnet (4.9) als lokale Kuhn–Tucker–Bedingung (Kuhn, Tucker (1951)). Unter Verwendung des Lagrange–Funktionals

$$L(\mathbf{x}, \mathbf{z}) := f(\mathbf{x}) + \mathbf{z}^T\mathbf{g}(\mathbf{x}), \quad (\mathbf{x}, \mathbf{z})^T \in \mathbb{R}^n \times \mathbb{R}^m_+$$

erhält man eine zu (4.9) äquivalente Formulierung

$$\nabla_{\mathbf{x}} L(\mathbf{x}, \mathbf{z}) = \mathbf{0},\ \nabla_{\mathbf{z}} L(\mathbf{x}, \mathbf{z}) \leq \mathbf{0},\ \mathbf{z}^T \nabla_{\mathbf{z}} L(\mathbf{x}, \mathbf{z}) = 0,\ \mathbf{z} \geq \mathbf{0}. \tag{4.10}$$

Die Optimierungsprobleme (4.9), (4.10) hat man aus (4.6), (4.7) durch die Festsetzung von $\eta = 1$ erhalten.

Lemma 4.20 Satz von Farkas. *Seien $A \in \mathbb{R}^{m\times n}$ und $\mathbf{b} \in \mathbb{R}^n$. Dann hat von den beiden Systemen*

$$A\mathbf{x} \leq \mathbf{0}, \quad \mathbf{b}^T\mathbf{x} > 0,$$

und

$$\mathbf{y}^T A = \mathbf{b}^T, \quad \mathbf{y} \geq \mathbf{0}$$

genau eines eine Lösung.

Beweis: Siehe Literatur. ■

Satz 4.21 Satz von Kuhn/Tucker. *Seien $f : \mathbb{R}^n \to \mathbb{R}$, $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^m$ stetig differenzierbar und $\mathbf{x}_0 \in \Omega$ Stelle eines lokalen Minimums von $f(\mathbf{x})$ bezüglich Ω. Ist in $\mathbf{x}_0$ die Regularitätsbedingung (4.8) erfüllt, dann existiert ein $\mathbf{z}_0 \in \mathbb{R}^m_+$, so dass $(\mathbf{x}_0, \mathbf{z}_0)^T$ eine Lösung von (4.9) bzw. von (4.10) ist.*

Beweis: Sei $I_0 \neq \emptyset$. Per Definition des linearisierten Kegels gilt

$$\left(\nabla \mathbf{g}_{I_0}(\mathbf{x}_0)\right)^T \mathbf{y} \leq \mathbf{0} \quad \forall\ \mathbf{y} \in K(\mathbf{x}_0). \tag{4.11}$$

Da (4.8) erfüllt ist, folgt aus Satz 4.8 für $I_0 \neq \emptyset$

$$\left(\nabla f(\mathbf{x}_0)\right)^T \mathbf{y} \geq 0 \implies \left(-\nabla f(\mathbf{x}_0)\right)^T \mathbf{y} \leq 0 \quad \forall\ \mathbf{y} \in T(\mathbf{x}_0) = K(\mathbf{x}_0). \tag{4.12}$$

Aus Lemma 4.20 folgt, dass (4.11) und (4.12) genau dann gleichzeitig gelten, wenn das System

$$\nabla \mathbf{g}_{I_0}(\mathbf{x}_0)\mathbf{z}_{I_0} = -\nabla f(\mathbf{x}_0), \quad \mathbf{z}_{I_0} \geq \mathbf{0}$$

eine Lösung besitzt. Setzt man $z_i = 0$ für $i \notin I_0$, so folgt die erste Gleichung von (4.9) für $I_0 \neq \emptyset$. Die Beziehung $\mathbf{g}(\mathbf{x}_0) \leq \mathbf{0}$ gilt per Definition des Optimierungsproblems. Aus der Konstruktion von $\mathbf{z}$ folgt schließlich $\mathbf{z}^T\mathbf{g}(\mathbf{x}_0) = 0$.

Für $I_0 = \emptyset$ ist $\mathbf{z} = \mathbf{0}$ eine Lösung von (4.9). ■

Im Falle $\mathbf{x}_0 \in \mathrm{int}(\Omega)$ gilt wegen der Regularitätsbedingung (4.8) $T(\mathbf{x}_0) = K(\mathbf{x}_0) = \mathbb{R}^n$. Aus der Definition von $K(\mathbf{x}_0)$ folgt dann, dass $\nabla \mathbf{g}(\mathbf{x}_0) = \mathbf{0}$ sein muss, da die Bedingung aus Satz 4.8 für jedes $\mathbf{y} \in \mathbb{R}^n$, insbesondere für $-\mathbf{y}$ gelten muss. Aus der ersten Gleichung der Kuhn–Tucker–Bedingung (4.9) folgt dann die bekannte notwendige Bedingung für ein lokales Minimum: $\nabla f(\mathbf{x}_0) = \mathbf{0}$.

Die Gültigkeit der Implikation von (4.11) nach (4.12) kann mit Hilfe eines linearen Programms überprüft werden. Diese Implikation gilt genau dann, wenn 0 der optimale Wert des Problem

$$\begin{aligned} z = \mathbf{z}^T \nabla f(\mathbf{x}_0) \quad &\rightarrow \quad \min\ ! \\ \mathbf{z}^T \nabla \mathbf{g}_{I_0}(\mathbf{x}_0) \quad &\leq \mathbf{0} \end{aligned}$$

ist.

Sind die Zielfunktion und die Nebenbedingungen zweimal differenzierbar, kann man weitere Kriterien mit Hilfe der Hesse–Matrix formulieren.

4.4 Globale Theorie der Lagrange–Multiplikatoren

Wir betrachten das Optimierungsproblem

$$z = \min\{f(\mathbf{x}) \ : \ \mathbf{x} \in G_\Omega\} \qquad \text{mit } G_\Omega = \{\mathbf{x} \in \mathbb{R}^n \ : \ \mathbf{g}(\mathbf{x}) \leq \mathbf{0}, \ \mathbf{x} \in \Omega\}, \tag{4.13}$$

wobei $\Omega \subset \mathbb{R}^n$ eine nichtleere Menge ist, $f : \Omega \rightarrow \mathbb{R}$, $\mathbf{g} : \Omega \rightarrow \mathbb{R}^m$. Bei dieser sogenannten konischen Formulierung der Menge G_Ω ist $\mathbf{x}$ genau dann zulässig, wenn $\mathbf{x} \in \Omega$ und $\mathbf{g}(\mathbf{x}) \leq \mathbf{0}$. Für $\Omega = \mathbb{R}^n$ haben wir als Spezialfall das Problem (4.4). Im folgenden werden Optimalitätskriterien in Form von Sattelpunktaussagen formuliert.

Definition 4.22 Sattelpunkt. Die Funktion $L(\mathbf{x}, \boldsymbol{\lambda})$ mit $\mathbf{x} \in \Omega$, $\boldsymbol{\lambda} \in D_{\boldsymbol{\lambda}}$ besitzt in $(\mathbf{x}_0, \boldsymbol{\lambda}_0)$ einen lokalen Sattelpunkt, wenn es ein $\varepsilon > 0$ gibt, so dass

$$L(\mathbf{x}_0, \boldsymbol{\lambda}) \leq L(\mathbf{x}_0, \boldsymbol{\lambda}_0) \leq L(\mathbf{x}, \boldsymbol{\lambda}_0) \quad \forall\ (\mathbf{x}, \boldsymbol{\lambda}) \in (\Omega \times D_{\boldsymbol{\lambda}}) \cap U_\varepsilon(\mathbf{x}_0, \boldsymbol{\lambda}_0), \tag{4.14}$$

wobei $U_\varepsilon(\mathbf{x}_0, \boldsymbol{\lambda}_0)$ eine ε–Umgebung von $(\mathbf{x}_0, \boldsymbol{\lambda}_0)$ ist. Gilt (4.14) für alle $(\mathbf{x}, \boldsymbol{\lambda}) \in \Omega \times D_{\boldsymbol{\lambda}}$, so hat $L(\mathbf{x}, \boldsymbol{\lambda})$ in $(\mathbf{x}_0, \boldsymbol{\lambda}_0)$ einen globalen Sattelpunkt. □

Zur Formulierung von globalen Optimalitätskriterien wird das folgende Sattelpunktproblem betrachtet: Gesucht sind $\mathbf{x}_0 \in \Omega$ und $(\eta_0, \mathbf{y}_0)^T \in \mathbb{R}_+^{m+1}$ mit $(\eta_0, \mathbf{y}_0)^T \neq \mathbf{0}$, so dass für das Lagrange–Funktional

$$L_\eta(\mathbf{x}, \mathbf{y}) := \eta f(\mathbf{x}) + \mathbf{y}^T \mathbf{g}(\mathbf{x}), \quad (\mathbf{x}, \mathbf{y})^T \in \Omega \times \mathbb{R}_+^m, \tag{4.15}$$

gilt

$$L_{\eta_0}(\mathbf{x}_0, \mathbf{y}) \leq L_{\eta_0}(\mathbf{x}_0, \mathbf{y}_0) \leq L_{\eta_0}(\mathbf{x}, \mathbf{y}_0), \quad \forall\ (\mathbf{x}, \mathbf{y})^T \in \Omega \times \mathbb{R}_+^m.$$

Als nächstes soll ein notwendiges Optimalitätskriterium für (4.13) bewiesen werden. Dazu benötigen wir folgendes Lemma.

Lemma 4.23 *Seien $\Omega \subseteq \mathbb{R}^n$ eine konvexe Menge, $f : \Omega \rightarrow \mathbb{R}$, $\mathbf{g} : \Omega \rightarrow \mathbb{R}^q$ konvexe Funktionen und $\mathbf{h} : \Omega \rightarrow \mathbb{R}^r$ affine Funktionen. Das System*

$$f(\mathbf{x}) < 0, \quad \mathbf{g}(\mathbf{x}) \leq \mathbf{0}, \quad \mathbf{h}(\mathbf{x}) \leq \mathbf{0}, \quad \mathbf{x} \in \Omega$$

besitzt genau dann keine Lösung, wenn ein Vektor $(u, \mathbf{v}, \mathbf{w})^T \in \mathbb{R}_+ \times \mathbb{R}_+^q \times \mathbb{R}_+^r$, $(u, \mathbf{v}, \mathbf{w})^T \neq \mathbf{0}$, existiert mit

$$u f(\mathbf{x}) + \mathbf{v}^T \mathbf{g}(\mathbf{x}) + \mathbf{w}^T \mathbf{h}(\mathbf{x}) \geq 0, \quad \forall\ \mathbf{x} \in \Omega.$$

Existiert ein $\bar{\mathbf{x}} \in \operatorname{int}(\Omega)$ welches zusätzlich $\mathbf{g}(\bar{\mathbf{x}}) < \mathbf{0}$ erfüllt, dann gilt $u \neq 0$.

Beweis: Siehe Literatur. ■

Satz 4.24 *Seien $\Omega \subseteq \mathbb{R}^n$ eine konvexe Menge, $f : \Omega \to \mathbb{R}$ und $\mathbf{g} : \Omega \to \mathbb{R}^m$ konvexe Funktionen. Ist $\mathbf{x}_0$ eine Lösung von (4.13), so existieren ein $\eta_0 \in \mathbb{R}_+$ und ein $\mathbf{y}_0 \in \mathbb{R}^m_+$, so dass $(\mathbf{x}_0, \eta_0, \mathbf{y}_0)^T \in \Omega \times \mathbb{R}^{m+1}_+$ eine Lösung des Sattelpunktproblems (4.15) ist.*

Beweis: Sei $\mathbf{x}_0$ eine Lösung von (4.13). Dann besitzt das System

$$f(\mathbf{x}) - f(\mathbf{x}_0) < 0, \quad \mathbf{g}(\mathbf{x}) \le \mathbf{0}, \quad \mathbf{x} \in \Omega$$

keine Lösung. Nach Lemma 4.23, mit $\mathbf{h}(\mathbf{x}) \equiv \mathbf{0}$, existiert dann ein Vektor $(\eta_0, \mathbf{y}_0)^T \in \mathbb{R}^{m+1}_+$, $(\eta_0, \mathbf{y}_0)^T \neq \mathbf{0}$, mit

$$L_{\eta_0}(\mathbf{x}, \mathbf{y}_0) = \eta_0 f(\mathbf{x}) + \mathbf{y}_0^T \mathbf{g}(\mathbf{x}) \ge \eta_0 f(\mathbf{x}_0) \quad \forall\, \mathbf{x} \in \Omega.$$

Wegen $\mathbf{g}(\mathbf{x}_0) \le \mathbf{0}$ folgt für alle $\mathbf{y} \ge \mathbf{0}$ damit

$$L_{\eta_0}(\mathbf{x}, \mathbf{y}_0) = \eta_0 f(\mathbf{x}) + \mathbf{y}_0^T \mathbf{g}(\mathbf{x}) \ge \eta_0 f(\mathbf{x}_0) + \mathbf{y}^T \mathbf{g}(\mathbf{x}_0) = L_{\eta_0}(\mathbf{x}_0, \mathbf{y}) \quad \forall\, (\mathbf{x}, \mathbf{y}) \in \Omega \times \mathbb{R}^m_+. \tag{4.16}$$

Setzt man in (4.16) rechts $\mathbf{y} = \mathbf{y}_0$, ergibt sich

$$\eta_0 f(\mathbf{x}) + \mathbf{y}_0^T \mathbf{g}(\mathbf{x}) \ge \eta_0 f(\mathbf{x}_0) + \mathbf{y}_0^T \mathbf{g}(\mathbf{x}_0) = L_{\eta_0}(\mathbf{x}_0, \mathbf{y}_0) \quad \forall\, \mathbf{x} \in \Omega.$$

Setzt man in (4.16) links $\mathbf{x} = \mathbf{x}_0$, erhält man

$$\eta_0 f(\mathbf{x}_0) + \mathbf{y}_0^T \mathbf{g}(\mathbf{x}_0) \ge \eta_0 f(\mathbf{x}_0) + \mathbf{y}^T \mathbf{g}(\mathbf{x}_0) \quad \forall\, \mathbf{y} \in \mathbb{R}^m_+.$$

Aus den letzten beiden Ungleichungen folgt die Behauptung. ■

Die Aussage dieses Satzes hat in gewissem Sinne Ähnlichkeit mit der von Satz 4.18. Für einen gegebenen zulässigen Bereich ist die notwendige Bedingung (nichtnegative Lösung von (4.15)) für beliebige Werte der Zielfunktion erfüllt, falls $\eta_0 = 0$ ist. Analog wie bei den lokalen Lagrange–Multiplikatoren wird deshalb eine Regularitätsbedingung eingeführt, die $\eta_0 > 0$ sichert.

Regularitätsbedingung. Seien die Menge $\Omega \subseteq \mathbb{R}^n$ sowie die Funktionen $g_i : \Omega \to \mathbb{R}$, $i \in \{1, \ldots, m\} =: I$, konvex. Es existiere ein $\bar{\mathbf{x}} \in \operatorname{int}(\Omega)$ mit

$$g_i(\bar{\mathbf{x}}) < 0 \;\text{ für } i \in I_N, \quad g_i(\bar{\mathbf{x}}) \le 0 \;\text{ für } i \in I_L. \tag{4.17}$$

Hierbei ist I_L die Menge aller $i \in \{1, \ldots, m\}$, für die $g_i(\mathbf{x})$ eine affine Funktion ist und I_N die Menge aller Indizes, für die $g_i(\mathbf{x})$ keine affine Funktion ist.

Unter dieser Regularitätsbedingung wird das Sattelpunktproblem: Finde einen Sattelpunkt $(\mathbf{x}_0, \mathbf{y}_0) \in \Omega \times \mathbb{R}^m_+$ des Lagrange–Funktional

$$L(\mathbf{x}, \mathbf{y}) := f(\mathbf{x}) + \mathbf{y}^T \mathbf{g}(\mathbf{x}), \quad (\mathbf{x}, \mathbf{y})^T \in \Omega \times \mathbb{R}^m_+, \tag{4.18}$$

betrachtet.

Satz 4.25 *Seien die Menge $\Omega \subseteq \mathbb{R}^n$ konvex, die Funktionen $f : \Omega \to \mathbb{R}$, $g_i : \Omega \to \mathbb{R}$, $i \in I$, konvex und die Regularitätsbedingung (4.17) erfüllt. Ist $\mathbf{x}_0 \in \Omega$ eine Lösung von (4.13), dann existiert ein $\mathbf{y}_0 \in \mathbb{R}^m_+$, so dass $(\mathbf{x}_0, \mathbf{y}_0)$ ein Sattelpunkt von (4.18) ist.*

Beweis: Da $\mathbf{x}_0$ eine Lösung von (4.13) ist, ist das System

$$f(\mathbf{x}) - f(\mathbf{x}_0) < 0, \quad \mathbf{g}(\mathbf{x}) \le \mathbf{0}, \quad \mathbf{x} \in \Omega$$

nicht lösbar. Wegen der Regularitätsbedingung (4.17) besitzt jedoch das System

$$\mathbf{g}_{I_N}(\mathbf{x}) < \mathbf{0}, \; \mathbf{g}_{I_L}(\mathbf{x}) \le \mathbf{0}, \quad \mathbf{x} \in \operatorname{int}(\Omega)$$

eine Lösung. Nach Lemma 4.23 existiert ein $(\eta_0, \mathbf{y}_0)^T \in \mathbb{R}_+^{m+1}$, $\eta_0 \neq 0$, (das $\mathbf{g}_{I_N}$ spielt die Rolle von $\mathbf{g}$ aus Lemma 4.23) mit

$$\eta_0 \left(f(\mathbf{x}) - f(\mathbf{x}_0)\right) + \mathbf{y}_0^T \mathbf{g}(\mathbf{x}) \geq 0 \quad \forall\, \mathbf{x} \in \Omega.$$

Bei dieser Darstellung wurden die Vektoren $\mathbf{v}$ und $\mathbf{w}$ aus Lemma 4.23 zum Vektor $\mathbf{y}$ zusammengefasst. Man kann so skalieren, dass $\eta_0 = 1$ gilt. Dann folgt

$$L(\mathbf{x}, \mathbf{y}_0) = f(\mathbf{x}) + \mathbf{y}_0^T \mathbf{g}(\mathbf{x}) \geq f(\mathbf{x}_0) \quad \forall\, \mathbf{x} \in \Omega.$$

Wegen $\mathbf{g}(\mathbf{x}_0) \leq \mathbf{0}$ folgt für alle $\mathbf{y} \geq \mathbf{0}$ gilt

$$L(\mathbf{x}_0, \mathbf{y}) = f(\mathbf{x}_0) + \mathbf{y}^T \mathbf{g}(\mathbf{x}_0) \leq f(\mathbf{x}_0) \quad \forall\, \mathbf{y} \in \mathbb{R}_+^m.$$

Aus den letzten beiden Beziehungen folgt, indem man $\mathbf{x} = \mathbf{x}_0$ beziehungsweise $\mathbf{y} = \mathbf{y}_0$ setzt,

$$L(\mathbf{x}_0, \mathbf{y}_0) = f(\mathbf{x}_0).$$

Nach Definition ist $(\mathbf{x}_0, \mathbf{y}_0)^T$ somit ein Sattelpunkt des Lagrange–Funktionals (4.18). ■

Nun wird eine hinreichende Bedingung für eine Lösung von (4.13) formuliert. Dabei gilt eine gewisse Umkehrung von Satz 4.25, die ohne weitere Voraussetzungen an (4.13) gilt.

Satz 4.26 *Sei* $(\mathbf{x}_0, \mathbf{y}_0)^T \in \Omega \times \mathbb{R}_+^m$ *ein Sattelpunkt von (4.18). Dann ist* $\mathbf{x}_0$ *eine Lösung von (4.13).*

Beweis: Zunächst wird gezeigt, dass $\mathbf{x}_0$ die Nebenbedingungen erfüllt. Aus $L(\mathbf{x}_0, \mathbf{y}) \leq L(\mathbf{x}_0, \mathbf{y}_0)$ für alle $\mathbf{y} \in \mathbb{R}_+^m$ folgt

$$(\mathbf{y} - \mathbf{y}_0)^T \mathbf{g}(\mathbf{x}_0) \leq 0 \quad \forall\, \mathbf{y} \in \mathbb{R}_+^m. \tag{4.19}$$

Es gilt $\mathbb{R}_+^m \subset \{\mathbf{y} - \mathbf{y}_0 \ : \ \mathbf{y}, \mathbf{y}_0 \in \mathbb{R}_+^m\}$. Somit gilt

$$\mathbf{y}^T \mathbf{g}(\mathbf{x}_0) \leq 0 \quad \forall\, \mathbf{y} \in \mathbb{R}_+^m.$$

Sei $g_i(\mathbf{x}_0) > 0$. Dann würde diese Beziehung nicht für den i–ten Einheitsvektor gelten, der aber zu $\mathbb{R}_+^m$ gehört. Damit folgt $\mathbf{g}(\mathbf{x}_0) \leq \mathbf{0}$.

Nun wird gezeigt, dass $f(\mathbf{x})$ in $\mathbf{x}_0$ ein globales Minimum annimmt. Aus (4.19) folgt für $\mathbf{y} = \mathbf{0}$ die Ungleichung $\mathbf{y}_0^T \mathbf{g}(\mathbf{x}_0) \geq 0$. Da $\mathbf{y}_0 \geq \mathbf{0}$ und $\mathbf{g}(\mathbf{x}_0) \leq \mathbf{0}$, kann das Skalarprodukt nur nichtpositiv sein, also gilt $\mathbf{y}_0^T \mathbf{g}(\mathbf{x}_0) = 0$. Mit dieser Beziehung und $L(\mathbf{x}_0, \mathbf{y}_0) \leq L(\mathbf{x}, \mathbf{y}_0)$ für alle $\mathbf{x} \in \Omega$ folgt

$$f(\mathbf{x}_0) \leq f(\mathbf{x}) + \mathbf{y}_0^T \mathbf{g}(\mathbf{x}) \quad \forall\, \mathbf{x} \in \Omega$$

und damit für alle $\mathbf{x} \in G_\Omega$

$$f(\mathbf{x}_0) \leq f(\mathbf{x}) + \underbrace{\mathbf{y}_0^T}_{\geq \mathbf{0}} \underbrace{\mathbf{g}(\mathbf{x})}_{\leq \mathbf{0}} \leq f(\mathbf{x}).$$

■

Bemerkung 4.27 Dualitätstheorie. Auch für die nichtlineare Optimierung gibt es eine Dualitätstheorie. Die Dualitätstheorie für lineare Programme ist darin als Spezialfall enthalten.

Gegeben sei das primale Problem

$$z = \min\{f(\mathbf{x}) \ : \ \mathbf{x} \in G_\Omega\} \qquad \text{mit } G_\Omega = \{\mathbf{x} \in \mathbb{R}^n \ : \ \mathbf{g}(\mathbf{x}) \leq \mathbf{0},\ \mathbf{x} \in \Omega\}, \tag{4.20}$$

wobei $\Omega \subseteq \mathbb{R}^n$ eine nichtleere Menge ist und $f : \Omega \to \mathbb{R}$, $\mathbf{g} : \Omega \to \mathbb{R}^m$ Abbildungen sind. Dem Problem (4.20) wird unter Verwendung des Lagrange–Funktionals das duale Problem

$$\tilde{z} = \max\{\phi(\mathbf{y}) \ : \ \mathbf{y} \in \mathbb{R}^m_+\}$$

zugeordnet, wobei

$$\phi(\mathbf{y}) \ : \ \mathbb{R}^m_+ \to \mathbb{R} \cup \{-\infty\}, \quad \phi(\mathbf{y}) = \inf_{\mathbf{x}\in\Omega} L(\mathbf{x},\mathbf{y}) = \inf_{\mathbf{x}\in\Omega} \left(f(\mathbf{x}) + \mathbf{y}^T \mathbf{g}(\mathbf{x})\right).$$

Man sieht, dass hier die Nichtnegativitätsbedingung im dualen Problem steckt, anders als wir es bei linearen Programmen hatten. Da wir dort jedoch gezeigt hatten, dass das duale Problem des dualen Problems wieder das primale Problem ist, ist das kein Widerspruch dazu, dass die Theorie für lineare Programme als Spezialfall enthalten ist.

Man kann wieder schnell zeigen, dass der Maximalwert des dualen Problems, falls er existiert, kleiner oder gleich dem Minimalwert des primalen Problems ist. Gleichheit muss im allgemeinen jedoch nicht gelten. Man spricht dann von einer Dualitätslücke. Des weiteren sind die Fragen bezüglich der Lösbarkeit der beiden Probleme wesentlich komplizierter zu beantworten als bei linearen Programmen. $\square$