

Kapitel 4

Optimalitätskriterien

Bemerkung 4.1 Motivation. 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. \square

4.1 Einleitung

Bemerkung 4.2 Methode der Lagrangeschen Multiplikatoren. 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

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

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

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

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 $\lambda_0, \dots, \lambda_p$ nennt man Lagrangesche Multiplikatoren. \square

Satz 4.3 Optimalitätskriterium für Problem (4.1). Seien $\mathbf{x}_0 \in \Omega$ und $\text{rg}(H_0(\mathbf{x}_0)) = \text{rg}(H(\mathbf{x}_0))$. Dann gilt: besitzt $f(\mathbf{x})$ in \mathbf{x}_0 ein lokales Minimum bezüglich Ω , so existiert ein $\boldsymbol{\lambda}^{(0)} = (\lambda_0^{(0)}, \dots, \lambda_p^{(0)})^T \in \mathbb{R}^{p+1}$, mit $\lambda_0^{(0)} \neq 0$ und

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

Beweis: Siehe Literatur. ■

Bemerkung 4.4 Fortsetzung: Methode der Lagrangeschen Multiplikatoren. 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

Bemerkung 4.5 Ziel. Wir betrachten das Optimierungsproblem

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

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

Zunächst werden spezielle konvergente Folgen betrachtet.

Definition 4.6 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 $\mathbf{x}_k \rightarrow \mathbf{x}_0$ heißt konvergent gegen \mathbf{x}_0 in Richtung $\mathbf{y} \in \mathbb{R}^n$ oder gerichtet konvergent, wenn gilt

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

Die Schreibweise ist

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

Beispiel 4.7 Sei $\{\mathbf{x}_k\}_{k \in \mathbb{N}}$ mit

$$\mathbf{x}_k = \begin{pmatrix} \frac{1}{k} & \frac{1}{k} \end{pmatrix}^T.$$

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

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

Bemerkung 4.8 Gerichtet konvergente Folgen. Der Vektor \mathbf{y} beschreibt die Richtung, aus der man sich mit der Folge $\{\mathbf{x}_k\}_{k \in \mathbb{N}}$ gegen \mathbf{x}_0 annähert.

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 $\{\mathbf{y}_k\}_{k \geq k_0}$ einen Häufungspunkt in dieser Menge und man findet in $\{\mathbf{y}_k\}_{k \geq k_0}$ eine konvergente Teilfolge. \square

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

Lemma 4.9 Die Folge $\{\mathbf{x}_k\}_{k \in \mathbb{N}}$ sei konvergent gegen \mathbf{x}_0 in Richtung \mathbf{y} . Dann gelten:

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

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

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

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

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

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

Außerdem existiert nach Voraussetzung der Grenzwert

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

Damit folgt

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

■

Definition 4.10 Tangentenkegel. Der Tangentenkegel $T(\mathbf{x}_0)$ an die Menge Ω im Punkt $\mathbf{x}_0 \in \Omega$ ist gegeben durch

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

□

Bemerkung 4.11 Tangentenkegel. 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 Tangentialhyperebene. 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 $T(\mathbf{x}_0) = \{\mathbf{0}\}$. \square

Beispiel 4.12 Sei

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

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

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

Analytisches Nachrechnen: Übungsaufgabe □

Lemma 4.13 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.

Satz 4.14 Notwendiges Kriterium für Existenz eines lokalen Minimums.

Sei $f : \mathbb{R}^n \rightarrow \mathbb{R}$ 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 $f(\mathbf{x}_k) \geq f(\mathbf{x}_0)$. Mit dieser Beziehung folgt, unter Beachtung von Lemma 4.9, 1),

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

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

Bemerkung 4.15 Minimum auf Rand des Gebiets. 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.

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 4.16 Zweites notwendiges Kriterium. Sei $f : \mathbb{R}^n \rightarrow \mathbb{R}$ 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 4.14, wobei jedoch jetzt Lemma 4.9, 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

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

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

Satz 4.17 Hinreichende Bedingung für die Existenz eines isolierten lokalen Minimums. Sei $f : \mathbb{R}^n \rightarrow \mathbb{R}$ 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 $\mathbf{x}_k \rightarrow \mathbf{x}_0$, 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 man annehmen, dass die gesamte Folge gegen \mathbf{x}_0 aus Richtung $\tilde{\mathbf{y}}$ konvergiert: $\mathbf{x}_k \xrightarrow{\tilde{\mathbf{y}}} \mathbf{x}_0$. Dann folgt aus Lemma 4.9

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

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

Bemerkung 4.18 Bekannte Kriterien für innere Punkte. Für jeden inneren Punkt $\mathbf{x}_0 \in \Omega$ gilt $T(\mathbf{x}_0) = \mathbb{R}^n$ und aus den Aussagen der Sätze 4.14, 4.16 und 4.17 folgen die bekannten notwendigen und hinreichenden Bedingungen für die Existenz lokaler Minima einer Funktion $f : \mathbb{R}^n \rightarrow \mathbb{R}$. *Übungsaufgabe* □

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

Bemerkung 4.19 Ziel. In diesem Abschnitt wird das Optimierungsproblem

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

mit $f \in C^1(\mathbb{R}^n)$ und $\mathbf{g} : \mathbb{R}^n \rightarrow \mathbb{R}^m, \mathbf{g} \in C^1(\mathbb{R}^n)$ 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. Lokal bedeutet, dass das Gebiet lokal durch affine Bedingungen approximiert wird. □

Definition 4.20 Aktive Nebenbedingung. Eine Nebenbedingung $g_i(\mathbf{x}) \leq 0, i = 1, \dots, m$, wird im Punkt $\mathbf{x}_0 \in \Omega$ aktiv genannt, wenn gilt $g_i(\mathbf{x}_0) = 0$. Bezeichnung:

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

□

Bemerkung 4.21 Sei $I_0 \neq \emptyset$. Die in \mathbf{x}_0 aktiven Nebenbedingungen werden nun durch affine Funktionen ersetzt

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

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:

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

Ausgehend von der Anschauung, könnte man die Menge

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

als eine lineare Approximation der Menge Ω im Punkt \mathbf{x}_0 ansehen. Dies ist jedoch in gewissen ausgearteten Punkten nicht zutreffend, siehe Beispiel 4.25 \square

Definition 4.22 Linearisierter Kegel. Die Menge

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

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

Lemma 4.23 *Es gilt $T(\mathbf{x}_0) \subseteq K(\mathbf{x}_0)$.*

Beweis: Für $I_0 = \emptyset$ gilt die Behauptung trivialerweise. Sei also $I_0 \neq \emptyset$. Der Nullvektor gehört per Definition zu beiden Mengen. Sei $\mathbf{y} \in T(\mathbf{x}_0)$, $\mathbf{y} \neq \mathbf{0}$. 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 $g_i(\mathbf{x}_k) \leq 0$ für alle i und $g_i(\mathbf{x}_0) = 0$ für alle $i \in I_0$ hat man

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

wobei Lemma 4.9, 1) verwendet wurde. Folglich gilt $\mathbf{y} \in K(\mathbf{x}_0)$. \blacksquare

Lemma 4.24 *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]. \blacksquare

Beispiel 4.25 Tangentenkegel und linearisierter Kegel. Sei

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

das heißt 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

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

Für den Tangentialkegel gilt $T(\mathbf{x}_0) = \{ \mathbf{y} \in \mathbb{R}^2 : y_1 \geq 0, y_2 = 0 \}$. 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. Der Grund ist die flache Spitze in \mathbf{x}_0 . Falls die Spitze in \mathbf{x}_0 weniger flach wäre, wäre eine lineare Approximation möglich. \square

Beispiel 4.26 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 4.25 kann auch dargestellt werden durch

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

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

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

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

Für den nächsten Beweis benötigen wir einen Hilfssatz.

Lemma 4.27 Alternativsatz von Gordan. *Sei $A \in \mathbb{R}^{m \times n}$, dann hat von den beiden Systemen*

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

und

$$\mathbf{z}^T A = \mathbf{0}, \quad \mathbf{z} \geq \mathbf{0}, \quad \mathbf{z} \neq \mathbf{0},$$

genau eines eine Lösung.

Beweis: Literatur. ■

Bemerkung 4.28 Seien die Voraussetzungen von Satz 4.14 erfüllt und $I_0 \neq \emptyset$. Dann folgt wegen $K_0(\mathbf{x}_0) \subseteq T(\mathbf{x}_0)$

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

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 4.27 besitzt dieses System genau dann keine Lösung, wenn das System

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

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:

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

Sei $I_0 = \emptyset$ und hat $f(\mathbf{x})$ in $\mathbf{x}_0 \in \Omega$ ein lokales Minimum, das heißt es gelten $\mathbf{x}_0 \in \text{int}(\Omega)$, $\mathbf{z} = \mathbf{0}$ und $\nabla f(\mathbf{x}_0) = \mathbf{0}$, dann besitzt dieses System offensichtlich eine Lösung $\eta > 0$, $\mathbf{z} = \mathbf{0}$. Damit hat man auch für den Fall $I_0 = \emptyset$ ein sinnvolles Problem definiert. □

Bemerkung 4.29 Ziel. Wir betrachten jetzt das Problem: Gesucht ist ein Tripel $(\mathbf{x}_0, \mathbf{z}_0, \eta_0)^T \in \mathbb{R}^n \times \mathbb{R}_+^m \times \mathbb{R}_+(\geq 0)$, $(\eta_0, \mathbf{z}_0)^T \neq \mathbf{0}$, welches das System

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

löst. Mit dem Lagrange-Funktional von (4.4)

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

ergibt sich die äquivalente Formulierung

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

□

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

Satz 4.30 Optimalitätskriterium für den Fall $\eta_0 \geq 0$. Seien $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $\mathbf{g} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ 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) beziehungsweise von (4.7) sind.

Beweis: Literatur, Fritz John (1948). ■

Bemerkung 4.31 Regularitätsbedingung. 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 4.30 für jede Lösung von (4.6) $\eta_0 > 0$ gilt.

$$\text{Regularitätsbedingung:} \quad K(\mathbf{x}_0) = T(\mathbf{x}_0). \quad (4.8)$$

□

Bemerkung 4.32 Zur Regularitätsbedingung.

- 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 4.25 und 4.26.
- Man kann andere Regularitätsbedingungen formulieren, die die obige Bedingungen implizieren, aber leichter zu überprüfen sind.

□

Bemerkung 4.33 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

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

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

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

erhält man eine zu (4.9) äquivalente Formulierung

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

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

Lemma 4.34 Satz von Farkas. Seien $A \in \mathbb{R}^{m \times n}$ und $\mathbf{b} \in \mathbb{R}^m$. 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 4.35 Satz von Kuhn/Tucker. Seien $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $\mathbf{g} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ 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) beziehungsweise von (4.10) ist.

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

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

Da (4.8) erfüllt ist, folgt aus Satz 4.14 für $I_0 \neq \emptyset$

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

Aus Lemma 4.34 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 $(\mathbf{x}_0, \mathbf{0})$ eine Lösung von (4.9). ■

Bemerkung 4.36 Im Falle $\mathbf{x}_0 \in \text{int}(\Omega)$ 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 4.14 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

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

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

Bemerkung 4.37 Ziel. Wir betrachten das Optimierungsproblem

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

Im folgenden werden Optimalitätskriterien in Form von Sattelpunktaussagen formuliert. □

Definition 4.38 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

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

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. \square

Bemerkung 4.39 Formulierung als Sattelpunktproblem. 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

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

gilt

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

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

Lemma 4.40 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 $\bar{\mathbf{x}} \in \text{int}(\Omega)$ welches zusätzlich $\mathbf{g}(\bar{\mathbf{x}}) < \mathbf{0}$ erfüllt, dann gilt $u \neq 0$.

Beweis: Siehe Literatur. \blacksquare

Satz 4.41 Notwendiges Optimalitätskriterium. Seien $\Omega \subseteq \mathbb{R}^n$ eine konvexe Menge, $f : \Omega \rightarrow \mathbb{R}$ und $\mathbf{g} : \Omega \rightarrow \mathbb{R}^m$ 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

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

keine Lösung. Nach Lemma 4.40, 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

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

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

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

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

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

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

$$\eta_0 f(\mathbf{x}_0) + \mathbf{y}_0^T \mathbf{g}(\mathbf{x}_0) \geq \eta_0 f(\mathbf{x}_0) + \mathbf{y}^T \mathbf{g}(\mathbf{x}_0) \quad \forall \mathbf{y} \in \mathbb{R}_+^m.$$

Aus den letzten beiden Ungleichungen folgt die Behauptung. \blacksquare

Bemerkung 4.42 Regularitätsbedingung. Die Aussage dieses Satzes hat in gewissem Sinne Ähnlichkeit mit der von Satz 4.30. 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 $g_i : \Omega \rightarrow \mathbb{R}$, $i \in \{1, \dots, m\} =: I$, konvex. Es existiere ein $\bar{\mathbf{x}} \in \text{int}(\Omega)$ mit

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

Hierbei ist I_L die Menge aller $i \in \{1, \dots, m\}$, 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

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

betrachtet. □

Satz 4.43 Existenz eines Sattelpunktes. *Seien die Menge $\Omega \subseteq \mathbb{R}^n$ konvex, die Funktionen $f : \Omega \rightarrow \mathbb{R}$, $g_i : \Omega \rightarrow \mathbb{R}$, $i \in I$, 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

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

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

$$\mathbf{g}_{I_N}(\mathbf{x}) < \mathbf{0}, \quad \mathbf{g}_{I_L}(\mathbf{x}) \leq \mathbf{0}, \quad \mathbf{x} \in \text{int}(\Omega)$$

eine Lösung. Nach Lemma 4.40 existiert ein $(\eta_0, \mathbf{y}_0)^T \in \mathbb{R}_+^{m+1}$, $\eta_0 \neq 0$, (das $\mathbf{g}_{I_N}(\mathbf{x})$ spielt die Rolle von $\mathbf{g}(\mathbf{x})$ aus Lemma 4.40) mit

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

Bei dieser Darstellung wurden die Vektoren \mathbf{v} und \mathbf{w} aus Lemma 4.40 zum Vektor \mathbf{y}_0 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-Funktional (4.18). ■

Nun wird eine hinreichende Bedingung für eine Lösung von (4.13) formuliert. Dabei gilt eine gewisse Umkehrung von Satz 4.43, die ohne weitere Voraussetzungen an (4.13) gilt.

Satz 4.44 Hinreichendes Optimalitätskriterium. *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

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

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 $\mathbf{x} \in \Omega$

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

■

Bemerkung 4.45 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

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

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

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

zugeordnet, wobei

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

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. □

Kapitel 5

Lösungsverfahren

Dieses Kapitel gibt einen Überblick über Lösungsverfahren für nichtlineare Optimierungsprobleme. Es wird vor allem auf die wesentlichen Ideen der Verfahren eingegangen und weniger auf Details.

5.1 Projektionsverfahren

Projektionsverfahren sind recht einfache Verfahren, die das Konzept von Abstiegsverfahren zur Minimierung von Funktionen ohne Nebenbedingungen auf Minimierungsprobleme mit konvexen Nebenbedingungen übertragen. Bei einem Abstiegsverfahren wird, ausgehend von einer Iterierten $\mathbf{x}^{(k)}$, die nächste Iterierte $\mathbf{x}^{(k+1)}$ so gewählt, dass sich der Wert der zu minimierenden Funktion verkleinert. Hat man ein Problem mit Nebenbedingungen, so muss man natürlich zusätzlich darauf achten, dass $\mathbf{x}^{(k+1)}$ zum zulässigen Bereich gehört. Anderenfalls kann es zum Beispiel vorkommen, dass die Zielfunktion gar nicht definiert ist. Projektionsverfahren projizieren die Abstiegsrichtung für die Zielfunktion in geeigneter Weise in die zulässige Menge.

Wir betrachten das Optimierungsproblem

$$z = \min\{f(\mathbf{x}) : \mathbf{x} \in \Omega\} \quad (5.1)$$

mit $f \in C^1(\mathbb{R}^n)$ und die zulässige Menge Ω sei nichtleer, konvex und abgeschlossen.

Von besonderer Bedeutung sind die Probleme, bei denen $f(\mathbf{x})$ quadratisch und

$$\Omega = \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} \leq \mathbf{b}\} \quad (5.2)$$

ein Polyeder ist, $A \in \mathbb{R}^{m \times n}$, $\mathbf{b} \in \mathbb{R}^m$. Solche Probleme treten als Teilprobleme bei den sogenannten SQP-Verfahren (sequential quadratic programming) auf, wo sie wiederholt mit unterschiedlichen Daten A, \mathbf{b}, f gelöst werden müssen, siehe Abschnitt 5.4.

In anderen Anwendungen ist der zulässige Bereich sogar nur ein n -dimensionaler Quader („box constraints“):

$$\Omega = \times_{i=1}^n [l_i, u_i]. \quad (5.3)$$

Für $\mathbf{x} \in \mathbb{R}^n$ sei $P_\Omega(\mathbf{x})$ die Lösung von

$$\inf_{\mathbf{y} \in \Omega} \|\mathbf{y} - \mathbf{x}\|_2,$$

die Projektion von \mathbf{x} auf Ω bezüglich der Euklidischen Norm. Da die Menge Ω nach Voraussetzung abgeschlossen ist, kann man das Infimum durch das Minimum ersetzen, also

$$P_\Omega(\mathbf{x}) = \arg \min_{\mathbf{y} \in \Omega} \|\mathbf{y} - \mathbf{x}\|_2.$$

Beispiel 5.1 Falls Ω durch (5.2) gegeben ist, muss man zur Berechnung der Projektion $\bar{\mathbf{y}} = P_{\Omega}(\mathbf{x})$ das konvexe quadratische Programm

$$\bar{\mathbf{y}} = \arg \min_{\mathbf{y} \in \Omega} \{\|\mathbf{y} - \mathbf{x}\|_2^2 : \mathbf{A}\mathbf{y} \leq \mathbf{b}\} = \arg \min_{\mathbf{y} \in \Omega} \{\mathbf{y}^T \mathbf{y} - 2\mathbf{y}^T \mathbf{x} : \mathbf{A}\mathbf{y} \leq \mathbf{b}\}$$

lösen. Die Menge Ω ist ein konvexes Polyeder. Die Zielfunktion $\|\mathbf{y} - \mathbf{x}\|_2^2$ ist eine konvexe Funktion.

Ist der zulässige Bereich ein n -dimensionaler Quader (5.3), kann man die Projektion komponentenweise berechnen:

$$(\bar{\mathbf{y}})_i = \begin{cases} x_i & \text{falls } x_i \in [l_i, u_i], \\ u_i & \text{falls } x_i > u_i, \\ l_i & \text{falls } x_i < l_i. \end{cases}$$

□

Definition 5.2 Stationärer Punkt. Der Punkt \mathbf{x}_0 wird stationärer Punkt des Problems (5.1) genannt, falls für alle Punkte des Tangentenkegels $\mathbf{y} \in T(\mathbf{x}_0)$ die Ungleichung

$$\mathbf{y}^T \nabla f(\mathbf{x}_0) \geq 0$$

gilt.

□

Ein stationärer Punkt erfüllt also die im Satz 4.14 bewiesene notwendige Bedingung für ein lokales Minimum bezüglich Ω .

Algorithmus 5.3 Projektionsverfahren.

1. *Initialisierung.*

Bestimme $\mathbf{x}^{(0)} \in \Omega$ und wähle drei reelle Parameter $0 < \beta$, $\mu < 1$ und $\gamma > 0$.

2. *Iteration.* $k = 0, 1, 2, \dots$

Falls $\mathbf{x}^{(k)}$ ein stationärer Punkt ist, dann

Stopp

sonst

Betrachte für $\alpha > 0$ den Pfad

$$\mathbf{x}^{(k)}(\alpha) := P_{\Omega}(\mathbf{x}^{(k)} + \alpha \nabla f(\mathbf{x}^{(k)}))$$

Setze

$$\mathbf{x}^{(k+1)}(\alpha) := \mathbf{x}^{(k)}(\alpha^{(k)}),$$

wobei $\alpha^{(k)} = \gamma \beta^{m^{(k)}}$ und $m^{(k)}$ die kleinste natürliche Zahl größer oder gleich Null ist mit

$$\mathbf{f}(\mathbf{x}^{(k+1)}) \leq \mathbf{f}(\mathbf{x}^{(k)}) + \mu \nabla \mathbf{f}(\mathbf{x}^{(k)})^T (\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)})$$

(Armijo-Liniensuche)

□

Im Algorithmus 5.3 hat man einen gekrümmten Pfad $\mathbf{x}^{(k)}(\alpha)$. Für jedes α hat man im allgemeinen ein konvexes, quadratisches Minimierungsproblem über Ω zu lösen. Das ist recht teuer.

Falls Ω ein Polyeder ist, kann man eine Startiterierte $\mathbf{x}^{(0)}$ mittels eines linearen Programms bestimmen.

Da der Algorithmus 5.3 im wesentlichen ein Gradientenverfahren ist, kann man im allgemeinen nur langsame Konvergenz erwarten. Außerdem ist die Berechnung von $P_\Omega(\mathbf{x}^{(k)} + \alpha \nabla f(\mathbf{x}^{(k)}))$ für allgemeine Polyeder aufwendig. Wir betrachten jetzt noch eine Variante von Algorithmus 5.3, die zusätzlich leichter berechenbare Zwischenwerte $\mathbf{x}^{(k)} \in \Omega$ berechnet, bei denen der Funktionswert zumindest nicht ansteigt.

Algorithmus 5.4 Modifiziertes Projektionsverfahren.

1. *Initialisierung.*

Bestimme $\mathbf{x}^{(0)} \in \Omega$ und wähle Parameter falls nötig.

2. *Iteration.* $k = 0, 1, 2, \dots$

Bestimme $\mathbf{x}^{(k+1)} = P_\Omega(\mathbf{x}^{(k)} + \alpha \nabla f(\mathbf{x}^{(k)}))$ entweder wie im Algorithmus 5.3 oder

bestimme $\mathbf{x}^{(k+1)} \in \Omega$, so dass $f(\mathbf{x}^{(k+1)}) \leq f(\mathbf{x}^{(k)})$.

□

Die Umsetzung der zweiten Strategie hängt von der Art der Nebenbedingungen ab. Wir betrachten affine Nebenbedingungen (5.2). Bezeichne $\hat{\mathbf{a}}_i$ die Zeilen von A . Für $\mathbf{x} \in \Omega$ sei die Menge der aktiven Nebenbedingungen

$$I(\mathbf{x}) = \{i \in \{1, \dots, m\} : \hat{\mathbf{a}}_i^T \mathbf{x} = b_i\}.$$

Dann wird die zweite Strategie häufig so realisiert, dass $I(\mathbf{x}^{(k)}) \subseteq I(\mathbf{x}^{(k+1)})$ gilt. Man wählt dazu in $\mathbf{x}^{(k)}$ eine Abstiegsrichtung

$$\mathbf{v}^{(k)} \in L(\mathbf{x}^{(k)}) := \{\mathbf{v} : A_{I(\mathbf{x}^{(k)})} \mathbf{v} = \mathbf{0}\},$$

wobei $A_{I(\mathbf{x}^{(k)})}$ die Teilmatrix von A mit den Zeilen ist, deren Indizes in $I(\mathbf{x}^{(k)})$ enthalten sind. Die Abstiegsrichtung wird dann nur in schwacher Form

$$\nabla f(\mathbf{x}^{(k)})^T \mathbf{v}^{(k)} \leq 0$$

verlangt. Die neue Iterierte besitzt die Gestalt $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha \mathbf{v}^{(k)}$ für ein geeignetes α . Aus der Wahl von $\mathbf{v}^{(k)}$ folgt

$$A(\mathbf{x}^{(k+1)}) = A(\mathbf{x}^{(k)} + \alpha \mathbf{v}^{(k)}) = A\mathbf{x}^{(k)} + \alpha A\mathbf{v}^{(k)}.$$

Für die aktiven Nebenbedingungen verschwindet der zweite Summand. Für die anderen Nebenbedingungen, $i \notin I(\mathbf{x}^{(k)})$, ist der erste Summand kleiner als b_i und man findet ein hinreichend kleines $\alpha > 0$, so dass die Summe der beiden Summanden kleiner oder gleich b_i bleibt. Mit

$$\bar{\alpha}^{(k)} := \sup\{\alpha : \mathbf{x}^{(k)} + \alpha \mathbf{v}^{(k)} \in \Omega\}$$

wird nun eine Liniensuche gestartet um einen Faktor $\alpha^{(k)}$ und damit ein Argument $\mathbf{x}^{(k+1)}$ zu finden, so dass $f(\mathbf{x}^{(k+1)}) \leq f(\mathbf{x}^{(k)})$ gilt. Ist $\alpha^{(k)} < \bar{\alpha}^{(k)}$, dann ist $I(\mathbf{x}^{(k)}) = I(\mathbf{x}^{(k+1)})$, sonst $I(\mathbf{x}^{(k)}) \subset I(\mathbf{x}^{(k+1)})$.

5.2 Penalty–Verfahren (Strafverfahren)

Wir betrachten wieder das Optimierungsproblem

$$z = \min\{f(\mathbf{x}) : \mathbf{x} \in \Omega\}, \tag{5.4}$$

diesmal aber zunächst mit $f \in C^0(\mathbb{R}^n)$ und die Menge $\Omega \subset \mathbb{R}^n$ sei abgeschlossen. Um die Lösung von (5.4) mit einer Folge einfacherer Optimierungsprobleme ohne Nebenbedingungen zu approximieren, wird die Straffunktion

$$l : \mathbb{R}^n \rightarrow \mathbb{R}_+, \quad l(\mathbf{x}) \begin{cases} > 0 & \text{für } \mathbf{x} \notin \Omega, \\ = 0 & \text{für } \mathbf{x} \in \Omega \end{cases}$$

eingeführt. Diese Funktion bestraft Punkte, die nicht zum zulässigen Bereich gehören, mit positiven Funktionswerten.

Beispiel 5.5 Für

$$\Omega = \{\mathbf{x} \in \mathbb{R}^n : g_i(\mathbf{x}) \leq 0, i = 1, \dots, p; g_i(\mathbf{x}) = 0, i = p + 1, \dots, m\}$$

ist eine mögliche Straffunktion

$$l(\mathbf{x}) = \sum_{i=1}^p (g_i^+(\mathbf{x}))^\alpha + \sum_{i=p+1}^m |g_i(\mathbf{x})|^\alpha$$

mit $\alpha > 0$, $g_i^+(\mathbf{x}) = \max\{0, g_i(\mathbf{x})\}$. □

Definition 5.6 Penalty-Funktion Die gewichtete Summe aus Zielfunktion und Straffunktion

$$p(\mathbf{x}, r) := f(\mathbf{x}) + rl(\mathbf{x}), \quad r \in \mathbb{R}_+, r > 0,$$

wird Penalty-Funktion genannt. Der Parameter r heißt Strafparameter. □

Für fest gewählte Parameter r werden jetzt die Minimierungsprobleme ohne Nebenbedingungen

$$\min_{\mathbf{x} \in \mathbb{R}^n} p(\mathbf{x}, r) \tag{5.5}$$

betrachtet. Der Strafterm belegt die Punkte, die nicht zum zulässigen Bereich gehören, mit positiven Werten, die für große Strafparameter r groß sind. Deswegen hofft man, dass die Minima von (5.5) für große Strafparameter im zulässigen Bereich liegen.

Algorithmus 5.7 Allgemeines Penalty-Verfahren.

1. *Initialisierung.*

Wähle $\mathbf{r}^{(1)} > 0$.

2. *Iteration.* $k = 0, 1, 2, \dots$

Bestimme ein lokales Minimum $\mathbf{x}^{(k)}$ für $p(\mathbf{x}, \mathbf{r}^{(k)})$

Falls $\mathbf{x}^{(k)} \in \Omega$, dann Stopp

sonst wähle $\mathbf{r}^{(k+1)} \geq 2\mathbf{r}^{(k)}$

□

Man kann zeigen, dass die $\mathbf{x}^{(k)}$ unter gewissen Voraussetzungen tatsächlich Näherungen eines lokalen Minimums von (5.4) sind.

Satz 5.8 Seien $f : \mathbb{R}^n \rightarrow \mathbb{R}$ eine stetige Funktion, \mathbf{x}_0 ein striktes lokales Minimum und $l : \mathbb{R}^n \rightarrow \mathbb{R}_+$ eine stetige Straffunktion. Dann gibt es ein $r_0 > 0$ so, dass für alle $r > r_0$ die Penalty-Funktion $p(\mathbf{x}, r)$ ein lokales Minimum $\mathbf{x}(r)$ besitzt, dass für $r \rightarrow \infty$ gegen \mathbf{x}_0 konvergiert.

Beweis: Literatur, [JS04, S. 294]. ■

Bemerkung 5.9 Zwei Eigenschaften sind für Penalty-Verfahren von Bedeutung:

1. In vielen Fällen ist die Zielfunktion $f(\mathbf{x})$ differenzierbar. Damit die Anwendung eines Verfahrens vom Newton-Typ zur Bestimmung des Minimums von (5.5) möglich ist, muss die Straffunktion $l(\mathbf{x})$ auch differenzierbar sein.
2. Damit das Verfahren nach endlich vielen Schritten abbricht, ist es wünschenswert, wenn es bereits einen endlichen Wert $\bar{r} > 0$ gibt, so dass ein lokales Minimum \mathbf{x}_0 von (5.4) auch lokales Minimum für jedes Problem ohne Nebenbedingungen (5.5) mit $r \geq \bar{r}$ ist. In diesem Fall nennt man die Penalty-Funktion exakt in \mathbf{x}_0 .

Es stellt sich leider heraus, dass diese beiden wünschenswerten Eigenschaften in der Regel unvereinbar sind. Aus diesem Grunde werden Penalty-Verfahren in der Form von Algorithmus 5.7 praktisch nicht genutzt. Stattdessen betrachtet man modifizierte Penalty-Funktionen, die auf dem Konzept einer erweiterten Lagrange-Funktion (augmented Lagrange-Funktion) beruhen, siehe Literatur. \square

5.3 Barrieremethoden

Wir betrachten das Problem

$$z = \min_{\mathbf{x} \in \Omega} f(\mathbf{x}) \quad (5.6)$$

mit den Nebenbedingungen

$$g_i(\mathbf{x}) \leq 0 \text{ für } 1 \leq i \leq p, \quad g_i(\mathbf{x}) = 0 \text{ für } p+1 \leq i \leq m. \quad (5.7)$$

Dabei seien $f, g_i \in C^2(\mathbb{R}^n)$, $i = 1, \dots, m$, und wir nehmen an, dass (5.6) eine Optimallösung besitzt, die mit \mathbf{x}_0 bezeichnet wird.

Barrieremethoden sind eng verwandt mit den Penalty-Verfahren. Auch bei diesen Methoden betrachtet man eine Folge von Hilfsproblemen, bei denen die Zielfunktion $f(\mathbf{x})$ durch gewichtete Strafterme erweitert wird. Die Barrieremethoden erzeugen eine Folge von inneren Punkten, das heißt von Punkten welche die Ungleichungsrestriktionen sogar strikt erfüllen, $g_i(\mathbf{x}) < 0, i = 1, \dots, p$, während die Gleichungsrestriktionen verletzt sein können.

Bezeichne

$$\begin{aligned} \hat{\Omega} &:= \{\mathbf{x} \in \mathbb{R}^n : g_i(\mathbf{x}) \leq 0, i = 1, \dots, p\}, \\ \hat{\Omega}_0 &:= \{\mathbf{x} \in \mathbb{R}^n : g_i(\mathbf{x}) < 0, i = 1, \dots, p\}. \end{aligned}$$

Man beachte, die Menge $\hat{\Omega}_0$ muss nicht notwendig die topologischen inneren Punkte von $\hat{\Omega}$ enthalten, wähle zum Beispiel $n = p = 1, g_1(x) = 0$. Dann sind $\hat{\Omega} = \mathbb{R}$ und $\hat{\Omega}_0 = \emptyset$.

Barrieremethoden bestrafen solche Punkte aus $\hat{\Omega}_0$, die sich dem Rand von $\hat{\Omega}_0$ nähern. Die Gleichheitsnebenbedingungen werden direkt mit Hilfe von Linearisierungen behandelt. Diese Gleichheitsnebenbedingungen sind grundsätzlich einfacher zu behandeln als Ungleichungsnebenbedingungen. Die Strafterme in den Barrieremethoden, die sogenannten Barriereterme, sind in $\hat{\Omega}_0$ endlich und wachsen zum Rand dieser Menge nach unendlich an. Außerhalb von $\hat{\Omega}$ besitzen sie den Wert ∞ . Im Gegensatz zu den Penalty-Verfahren, bei denen die Strafterme sukzessive immer stärker gewichtet werden, siehe Algorithmus 5.7, muss bei den Barrieremethoden der Einfluss der Strafterme immer weiter abgeschwächt werden. Damit wird das Gewicht der Zielfunktion im Barriereproblem erhöht und man kann hoffen, dass die Minima der Barriereprobleme unter geeigneten Voraussetzungen gegen ein Minimum von $f(\mathbf{x})$ konvergieren. An der Eigenschaft, dass die Barriereterme außerhalb von $\hat{\Omega}$ unendlich sind, ändert man nichts. Somit ist garantiert, dass die Minima der Barriereprobleme immer in $\hat{\Omega}_0$ liegen.

Definition 5.10 Skalare Barrierefunktion. Eine skalare Barrierefunktion ist eine streng monoton fallende, glatte, konvexe Funktion $b : (0, \infty) \rightarrow \mathbb{R}$ mit

$$\lim_{t \rightarrow 0+0} b(t) = \infty \quad \text{und} \quad \lim_{t \rightarrow 0+0} b'(t) = -\infty.$$

□

Außerdem wird stets $b(t) = \infty$ für $t \leq 0$ gesetzt, so dass $b(t)$ formal eine auf \mathbb{R} definierte konvexe Funktion ist, $b : \mathbb{R} \rightarrow \mathbb{R} \cup \{\infty\}$.

Beispiel 5.11 Beispiele für Barrierefunktionen sind

$$b(t) = -\log t, \quad b(t) = \frac{1}{t^\alpha}, \quad \alpha > 0.$$

Die logarithmische Barrierefunktion ist in gewisser Hinsicht optimal. □

Zur Konstruktion von Barriereverfahren zur Lösung von Problem (5.6) mit den Nebenbedingungen (5.7) werden nun Hilfsprobleme der Form

$$\inf_{\mathbf{x} \in \mathbb{R}^n} \left\{ f(\mathbf{x}) + \mu \sum_{i=1}^p b(d_i - g_i(\mathbf{x})) : g_i(\mathbf{x}) = 0, i = p+1, \dots, m \right\} \quad (5.8)$$

betrachtet. In (5.8) ist $\mu > 0$ ein Gewicht für die Barriereterme und die Zahlen $d_i \geq 0, i = 1, \dots, p$, sind Verschiebungen der Ungleichungsnebenbedingungen in (5.7), das heisst anstatt $g_i(\mathbf{x}) \leq 0$ ist nun $g_i(\mathbf{x}) \leq d_i$ erlaubt. Diese Verschiebungen gestatten es, dass man das Verfahren auch dann anwenden kann, wenn kein innerer Punkt für (5.6), (5.7) bekannt ist. Die Zielfunktion von (5.8) wird abkürzend mit

$$\Phi(\mathbf{x}; \mu, \mathbf{d}) = f(\mathbf{x}) + \mu \sum_{i=1}^p b(d_i - g_i(\mathbf{x}))$$

bezeichnet.

Wir nehmen an, dass (5.8) ein endliches lokales Minimum besitzt. Die gewichtete Summe der Barriereterme in der Zielfunktion garantiert, dass jedes \mathbf{x} mit $\Phi(\mathbf{x}; \mu, \mathbf{d}) \in \mathbb{R}$ die abgeschwächten Nebenbedingungen $g_i(\mathbf{x}) < d_i, i = 1, \dots, p$, erfüllt.

Lemma 5.12 Falls $f(\mathbf{x})$ und $g_i(\mathbf{x}), i = 1, \dots, p$ konvex sind, so ist auch $\Phi(\mathbf{x}; \mu, \mathbf{d})$ konvex.

Beweis: Es ist bekannt, dass die Linearkombination konvexer Funktionen mit nicht-negativen Koeffizienten in der Linearkombination eine konvexe Funktion ist. Damit bleibt zu zeigen, dass die Funktionen $b(d_i - g_i(\mathbf{x})), i = 1, \dots, p$, konvex sind.

Da die Funktionen, $g_i(\mathbf{x})$ konvex sind, gilt für $\lambda \in [0, 1]$

$$\begin{aligned} d_i - g_i(\lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) &\geq d_i - (\lambda g_i(\mathbf{x}_1) + (1-\lambda)g_i(\mathbf{x}_2)) \\ &= \lambda(d_i - g_i(\mathbf{x}_1)) + (1-\lambda)(d_i - g_i(\mathbf{x}_2)). \end{aligned}$$

Mit dieser Aussage, mit der Monotonie von $b(t)$ und der Konvexität von $b(t)$ folgt

$$\begin{aligned} b(d_i - g_i(\lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2)) &\leq b(\lambda(d_i - g_i(\mathbf{x}_1)) + (1-\lambda)(d_i - g_i(\mathbf{x}_2))) \\ &\leq \lambda b(d_i - g_i(\mathbf{x}_1)) + (1-\lambda)b(d_i - g_i(\mathbf{x}_2)). \end{aligned}$$

■

Weiterhin gilt folgende stärkere Aussage.

Satz 5.13 *Gelten die Voraussetzungen von Lemma 5.12 sowie $\lim_{t \rightarrow \infty} b'(t) = 0$. Die Gleichheitsnebenbedingungen $g_i(\mathbf{x})$, $i = p + 1, \dots, m$ seien affin und die Menge der Optimallösungen von (5.6) sei nicht leer und beschränkt. Dann besitzt das Hilfsproblem (5.8) für jedes $\mu > 0$ eine Optimallösung und die Minima der Barriereprobleme nähern sich der Optimalmenge von (5.6).*

Beweis: Siehe Literatur. ■

Für Probleme (5.6), die die Bedingungen dieses Satzes erfüllen, kann man nun das folgende Verfahren konstruieren.

Algorithmus 5.14 **Barriermethode für konvexe Probleme.**

1. *Initialisierung.*

Bestimme $\mathbf{x}^{(0)} \in \mathbb{R}^n$ mit $g_i(\mathbf{x}^{(0)}) = 0$ für $i = p + 1, \dots, m$. Wähle $\mu^{(0)} > 0$ und $\mathbf{d}^{(0)} \geq \mathbf{0}$ so dass $d_i^{(0)} > g_i(\mathbf{x}^{(0)})$ für $i = 1, \dots, p$.

2. *Iteration.* $k = 1, 2, \dots$

Wähle $\lambda^{(k)} \in (0, 1)$ so, dass mit $(\mu^{(k)}, \mathbf{d}^{(k)}) := \lambda^{(k)} (\mu^{(k-1)}, \mathbf{d}^{(k-1)})$ gilt

$$g_i(\mathbf{x}^{(k-1)}) < d_i^{(k)}, \quad \text{für } i = 1, \dots, p.$$

Ausgehend von $\mathbf{x}^{(k-1)}$ führt man nun einige Schritte des Newton-Verfahrens (mit Liniensuche) zum lösen des Barriereproblems aus. Das Ergebnis ist $\mathbf{x}^{(k)}$.

□

Im ersten Schritt der Iteration werden sowohl das Gewicht als auch der Verschiebevektor verkleinert. Der Verkleinerungsfaktor wird so gewählt, dass mit dem neuen Verschiebevektor noch alle Ungleichungsnebenbedingungen erfüllt sind. Das Minimum $\mathbf{x}_0(\mu^{(k)}, \mathbf{d}^{(k)})$ des Barriereproblems (5.8) zu den Parametern $(\mu^{(k)}, \mathbf{d}^{(k)})$ wird im zweiten Schritt approximiert. Da die Barriereterme das Minimum vom Rand der Menge $\{\mathbf{x} : g_i(\mathbf{x}) \leq d_i^{(k)}\}$ abstoßen, kann man nach der Berechnung der Näherung $\mathbf{x}^{(k)}$ von $\mathbf{x}_0(\mu^{(k)}, \mathbf{d}^{(k)})$ die Verschiebeparameter $d_i^{(k)}$ in der folgenden Iteration wieder etwas verkleinern.

Die Schwierigkeiten von Algorithmus 5.14 bestehen darin, dass das Newton-Verfahren für $\mu^{(k)} \rightarrow 0$ oft schlecht konvergiert. Deshalb wird diese Basiserangeheungsweise nicht genutzt. Man kann diese Herangeheungsweise durch Verfeinerung der Barriermethode verbessern.

5.4 SQP-Verfahren

In diesem Abschnitt wird ein Zugang vorgestellt, der Punkte berechnet, die die notwendige Optimalitätsbedingung, die im Satz 4.35 (Kuhn/Tucker) formuliert ist, erfüllen, die sogenannten SQP-Verfahren (sequential quadratic programming). Es wird also eine Iteration durchgeführt, bei welcher in jedem Schritt ein quadratisches Optimierungsproblem gelöst wird.

Wir betrachten wieder das Optimierungsproblem (5.6) mit den Nebenbedingungen (5.7) und den gleichen Regularitätsvoraussetzungen wie im Abschnitt 5.3. Seien \mathbf{x}_0 ein lokales Minimum von (5.6), (5.7) und \mathbf{z}_0 der zugehörige Lagrange-Multiplikator zum Lagrange-Problem (4.10). Insgesamt erfüllen $(\mathbf{x}_0, \mathbf{z}_0)$ das Problem, siehe (4.10),

$$\Phi(\mathbf{x}_0, \mathbf{z}_0) = \begin{pmatrix} \nabla_{\mathbf{x}} L(\mathbf{x}_0, \mathbf{z}_0) \\ \mathbf{z}_0^T \mathbf{g}(\mathbf{x}_0) \end{pmatrix} = \begin{pmatrix} \nabla f(\mathbf{x}_0) + (\nabla \mathbf{g}(\mathbf{x}_0))^T \mathbf{z}_0 \\ \mathbf{z}_0^T \mathbf{g}(\mathbf{x}_0) \end{pmatrix} = \mathbf{0} \quad (5.9)$$

mit $\mathbf{z}_0 \geq \mathbf{0}$. Da die Gleichheitsbedingungen ohnehin verschwinden, kann man (5.9) sogar wie folgt schreiben

$$\Phi(\mathbf{x}_0, \mathbf{z}_0) = \begin{pmatrix} \nabla f(\mathbf{x}_0) + (\nabla \mathbf{g}(\mathbf{x}_0))^T \mathbf{z}_0 \\ z_{0,1} g_1(\mathbf{x}_0) \\ \vdots \\ z_{0,p} g_p(\mathbf{x}_0) \\ g_{p+1}(\mathbf{x}_0) \\ \vdots \\ g_m(\mathbf{x}_0) \end{pmatrix} = \mathbf{0}. \quad (5.10)$$

SQP-Verfahren wollen das nichtlineare Problem (5.10) mit einem Verfahren vom Newton-Typ lösen. Dazu benötigt man die Jacobi-Matrix von $\Phi(\mathbf{x}, \mathbf{z})$, die durch

$$\begin{aligned} & \Psi(\mathbf{x}, \mathbf{z}, H_{L,\mathbf{x}}(\mathbf{x}, \mathbf{z})) \\ &= \begin{pmatrix} H_{L,\mathbf{x}}(\mathbf{x}, \mathbf{z}) & \nabla g_1(\mathbf{x}) & \cdots & \nabla g_p(\mathbf{x}) & \nabla g_{p+1}(\mathbf{x}) & \cdots & \nabla g_m(\mathbf{x}) \\ z_1 (\nabla g_1(\mathbf{x}))^T & g_1(\mathbf{x}) & & & & & \\ \vdots & & \ddots & & & & 0 \\ z_p (\nabla g_p(\mathbf{x}))^T & & & g_p(\mathbf{x}) & & & \\ (\nabla g_{p+1}(\mathbf{x}))^T & & & & & & \\ \vdots & & & & & & \\ (\nabla g_m(\mathbf{x}))^T & & & & 0 & & 0 \end{pmatrix} \\ & \in \mathbb{R}^{(n+m) \times (n+m)} \end{aligned}$$

gegeben ist. Ein wesentliches Merkmal eines SQP-Verfahrens besteht darin, dass die teure Hesse-Matrix $H_{L,\mathbf{x}}(\mathbf{x}, \mathbf{z})$ in der Regel durch eine einfacher zu berechnende Matrix ersetzt wird.

Unter geeigneten Voraussetzungen kann man zeigen, dass die Jacobi-Matrix $\Psi(\mathbf{x}_0, \mathbf{z}_0, H_{L,\mathbf{x}}(\mathbf{x}_0, \mathbf{z}_0))$ nichtsingulär ist und dass das Newton-Verfahren zur Nullstellenbestimmung von $\Phi(\mathbf{x}, \mathbf{z})$ quadratisch konvergiert. Sei eine aktuelle Iterierte $(\mathbf{x}^{(k)}, \mathbf{z}^{(k)})^T$ gegeben. Die Newton-Korrektur $(\Delta \mathbf{x}^{(k)}, \Delta \mathbf{z}^{(k)})^T$ berechnet sich als Lösung des Gleichungssystems

$$\Psi(\mathbf{x}^{(k)}, \mathbf{z}^{(k)}, H_{L,\mathbf{x}}(\mathbf{x}^{(k)}, \mathbf{z}^{(k)})) (\Delta \mathbf{x}^{(k)}, \Delta \mathbf{z}^{(k)})^T = -\Phi(\mathbf{x}^{(k)}, \mathbf{z}^{(k)}).$$

Beim Newton-Verfahren kann man im allgemeinen jedoch nur lokale Konvergenz erwarten, das heißt, der Startwert muss nahe genug an der (unbekannten) Lösung sein. Speziell für die Funktion $\Phi(\mathbf{x}, \mathbf{z})$ kann man auch nicht garantieren, dass alle Iterierten die Ungleichungen $g_i(\mathbf{x}^{(k)}) \leq 0$ und die Nichtnegativitätsbedingung $z_i^{(k)} \geq 0$ erfüllen. Damit ist es möglich, dass das Newton-Verfahren gegen eine nicht zulässige Lösung von $\Phi(\mathbf{x}, \mathbf{z}) = \mathbf{0}$ konvergiert, bei der Nebenbedingungen nicht erfüllt sind oder die Lagrange-Multiplikatoren negativ sind. Die Konvergenz gegen eine solche Lösung muss verhindert werden. Dazu wird anstelle des Newton-Verfahrens das System

$$\Psi(\mathbf{x}^{(k)}, \mathbf{z}^{(k+1)}, B^{(k)}) (\Delta \mathbf{x}^{(k)}, \Delta \mathbf{z}^{(k)})^T = -\Phi(\mathbf{x}^{(k)}, \mathbf{z}^{(k)}) \quad (5.11)$$

betrachtet, wobei $\Delta \mathbf{x}^{(k)}$, $\Delta \mathbf{z}^{(k)}$ und $\mathbf{z}^{(k+1)} = \mathbf{z}^{(k)} + \Delta \mathbf{z}^{(k)}$ die zusätzlichen Forderungen

$$z_i^{(k+1)} \geq 0 \quad \text{für } i = 1, \dots, p, \quad (5.12)$$

$$g_i(\mathbf{x}^{(k)}) + \nabla(g_i(\mathbf{x}^{(k)}))^T \Delta \mathbf{x}^{(k)} \leq 0 \quad \text{für } i = 1, \dots, p \quad (5.13)$$

erfüllen. Im Vergleich zum Newton-Verfahren ersetzt man $H_{L,\mathbf{x}}(\mathbf{x}^{(k)}, \mathbf{z}^{(k)})$ durch eine Matrix $B^{(k)}$, die in der Regel durch gewisse Quasi-Newton-Korrekturen (so genannte Broyden-Verfahren) erzeugt wird. Des Weiteren wird der Vektor $\mathbf{z}^{(k)}$ auf der linken Seite durch $\mathbf{z}^{(k+1)}$ ersetzt. Man erhält ein implizites Gleichungssystem, welches nicht mehr linear bezüglich der Lagrange-Multiplikatoren ist. Außerdem werden noch die linearen Ungleichungsbedingungen (5.12), (5.13) an $\Delta\mathbf{x}^{(k)}$ und $\Delta\mathbf{z}^{(k)}$ gestellt.

Ausgeschrieben besagt (5.11) – (5.13)

$$\begin{aligned} \nabla f(\mathbf{x}^{(k)}) + B^{(k)} \Delta\mathbf{x}^{(k)} + \sum_{i=1}^m z_i^{(k+1)} \nabla g_i(\mathbf{x}^{(k)}) &= 0, \\ z_i^{(k+1)} \left(g_i(\mathbf{x}^{(k)}) + \left(\nabla g_i(\mathbf{x}^{(k)}) \right)^T \Delta\mathbf{x}^{(k)} \right) &= 0, \quad i = 1, \dots, p, \quad (5.14) \\ g_i(\mathbf{x}^{(k)}) + \left(\nabla g_i(\mathbf{x}^{(k)}) \right)^T \Delta\mathbf{x}^{(k)} &= 0, \quad i = p+1, \dots, m. \end{aligned}$$

Wir betrachten das folgende quadratische Programm

$$z = \min_{\mathbf{s} \in \mathbb{R}^n} \left(\nabla f(\mathbf{x}^{(k)}) \right)^T \mathbf{s} + \frac{1}{2} \mathbf{s}^T B^{(k)} \mathbf{s} \quad (5.15)$$

unter den Nebenbedingungen

$$g_i(\mathbf{x}^{(k)}) + \left(\nabla g_i(\mathbf{x}^{(k)}) \right)^T \mathbf{s} \leq 0, \quad i = 1, \dots, p, \quad (5.16)$$

$$g_i(\mathbf{x}^{(k)}) + \left(\nabla g_i(\mathbf{x}^{(k)}) \right)^T \mathbf{s} = 0, \quad i = p+1, \dots, m. \quad (5.17)$$

Erfülle dieses Problem die Voraussetzungen des Satzes 4.35 (Kuhn/Tucker). Dann sind die notwendigen Bedingungen für ein Minimum gerade die Gleichungen (5.14).

Übungsaufgabe Damit ergibt sich folgendes Verfahren:

Algorithmus 5.15 SQP-Algorithmus (Grundform).

1. *Initialisierung.*

Wähle $\mathbf{x}^{(0)} \in \mathbb{R}^n, \mathbf{B}^{(0)} = (\mathbf{B}^{(0)})^T (\approx H_{L,\mathbf{x}}(\mathbf{x}^{(0)}, \mathbf{z}^{(0)}))$ für ein $\mathbf{z}^{(0)} \in \mathbb{R}^m$ mit $z_i^{(0)} > 0, i = 1, \dots, p$

2. *Iteration.* $k = 0, 1, 2, \dots$

Bestimme die Lösung \mathbf{s} von (5.15) – (5.17) und einen zugehörigen Lagrange-Multiplikator \mathbf{z} . Setze

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \mathbf{s}, \quad \mathbf{z}^{(k+1)} = \mathbf{z}.$$

Bestimme eine symmetrische Matrix

$$\mathbf{B}^{(k+1)} \approx H_{L,\mathbf{x}}(\mathbf{x}^{(k+1)}, \mathbf{z}^{(k+1)}).$$

□

Falls $B^{(k)}$ positiv semidefinit ist, dann ist (5.15) – (5.17) ein konvexes quadratisches Programm. In diesem Fall sind die Bedingungen des Satzes 4.35 notwendig und hinreichend für ein globales Minimum, siehe auch Satz 3.27. Zur Lösung kann man beispielsweise ein Projektionsverfahren nehmen, siehe Abschnitt 5.1.