GRUNDLAGEN DER OPTIMIERUNG

S. VOLKWEIN

ABSTRACT. In dem vorliegenden Skript sind Grundlagen der finiten und in-finiten Optimierung zusammengestellt. Dabei werden auch für die Optimierung wichtige Resultate aus der konvexen Analysis wiederholt. Im wesentlichen sind die Resultate aus den beiden Büchern [1, 2] von D. G. Luenberger übersetzt worden.

1. Einige Grundlagen aus der finiten Optimierung

In diesem ersten Abschnitt wollen wir uns mit Grundlagen aus der endlich-dimensionalen Optimierung beschäftigen. Als Referenz für die angegebenen Resultate und für ein detaillierteres Studium verweisen wir auf das Buch [1] von D. G. Luenberger.

1.1. Konvexe Mengen. Wir beginnen mit Resultaten für konvexe Mengen.

Lemma 1.1.1 (Eigenschaften konvexer Mengen). a) Wenn $C \subseteq \mathbb{R}^n$ eine konvexe Menge ist und $\beta \in \mathbb{R}$ gilt, dann ist auch die Menge

$$\beta C = \{x : x = \beta c, c \in C\}$$

konvex.

b) Seien C und D konvexe Mengen, dann ist auch

$$C + D = \{x : x = c + d, c \in C \wedge d \in D\}$$

konvex.

c) Der Schnitt von beliebig vielen konvexen Mengen ist wieder konvex.

Definition 1.1.2 (Kegel). Eine Menge C wird ein Kegel genannt, wenn aus $x \in C$ folgt $\alpha x \in C$ für alle $\alpha > 0$.

Definition 1.1.3 (Lineare Mannigfaltigkeit). $V \subseteq \mathbb{R}^n$ heißt eine lineare Mannigfaltigkeit, wenn

$$\lambda x_1 + (1 - \lambda)x_2 \in V$$

für alle $\lambda \in \mathbb{R}$ und $x_1, x_2 \in V$ gilt.

Bemerkung 1.1.4. Bei konvexen Mengen gilt $\lambda \in [0, 1]$. Im $\mathbb{R}^3$ sind lineare Mannigfaltigkeiten Punkte, Geraden, Ebenen oder der gesamte Raum.

Definition 1.1.5 (Hyperebene). Eine Hyperebene im $\mathbb{R}^n$ ist eine $(n - 1)$ -dimensionale Mannigfaltigkeit.

Date: December 9, 2002.

1991 Mathematics Subject Classification. 49-02, 49Kxx, 49Mxx, 90-02, 90Cxx.

Key words and phrases. Konvexe Analysis, Optimalitätsbedingungen, Penalty- und Barriere-Verfahren, Multiplikator-Methoden.

Lemma 1.1.6 (Identifikation einer Hyperebene). *Seien $a \in \mathbb{R}^n \setminus \{0\}$ und $c \in \mathbb{R}$. Die Menge*

$$H = \{x \in \mathbb{R}^n : a^T x = c\}$$

ist eine Hyperebene.

Beweis. Aus der Linearität von $a^T x = c$ folgt sofort, dass H eine lineare Mannigfaltigkeit ist. Eine Verschiebung der Menge H um den Vektor x_1 , so dass die verschobene Menge $M = H - x_1$ alle Vektoren mit $a^T x = 0$ enthält, ergibt die Behauptung. $\square$

Lemma 1.1.7. *Sei H eine Hyperebene im $\mathbb{R}^n$. Dann existieren ein Vektor $a \in \mathbb{R}^n \setminus \{0\}$ und ein $c \in \mathbb{R}$, so dass $H = \{x \in \mathbb{R}^n : a^T x = c\}$ gilt.*

Bemerkung 1.1.8. *Wenn wir Lemma 1.1.6 und Lemma 1.1.7 kombinieren, dann erhalten wir, dass eine Hyperebene die Menge aller Lösungen einer linearen Gleichung ist.*

Definition 1.1.9. *Seien $a \in \mathbb{R}^n \setminus \{0\}$ und $c \in \mathbb{R}$. Zu der Hyperebene $H = \{x : a^T x = c\}$ definieren wir die positiven und negativen abgeschlossenen Halbräume*

$$\overline{H}_+ = \{x : a^T x \geq c\}, \quad \overline{H}_- = \{x : a^T x \leq c\}$$

und die positiven und negativen offenen Halbräume

$$H_+ = \{x : a^T x > c\}, \quad H_- = \{x : a^T x < c\}.$$

Definition 1.1.10 (Konvexes Polytop). *Eine Menge, die als Schnitt von endlich vielen abgeschlossenen Halbräumen ausgedrückt werden kann, heißt konvexes Polytop.*

Definition 1.1.11 (Polyhedron). *Ein nicht-leeres beschränktes Polytop heißt Polyhedron.*

Satz 1.1.12. *Seien $C \subset \mathbb{R}^n$ eine konvexe Menge und y ein Punkt außerhalb von $\text{cl}(C) = \{x : \text{es existiert eine Folge } \{x_n\}_{n \in \mathbb{N}} \text{ in } C \text{ mit } \lim_{n \rightarrow \infty} x_n = x\}$. Dann gibt es einen Vektor a mit*

$$a^T y < \inf_{x \in C} a^T x.$$

Bemerkung 1.1.13. *Der Satz 1.1.12 läßt sich geometrisch deuten: Es gibt eine Hyperebene durch y , die in einer ihrer offenen Halbräume die Menge C enthält.*

Satz 1.1.14. *Seien $C \subset \mathbb{R}^n$ eine konvexe Menge und y ein Randpunkt von C . Dann gibt es eine Hyperebene, die y enthält und für die C in einer der abgeschlossenen Halbräume liegt.*

Definition 1.1.15 (Tragende Halbebene). *Eine Hyperebene, die eine konvexe Menge C in einer ihrer abgeschlossenen Halbräume und einen Randpunkt von C enthält, wird tragende Halbebene genannt.*

Definition 1.1.16 (Extrempunkt). *Ein Punkt x in einer konvexen Menge C heißt Extrempunkt von C , wenn es keine zwei Punkte x_1 und x_2 aus C gibt, so dass $x = \alpha x_1 + (1 - \alpha)x_2$ für ein α mit $0 < \alpha < 1$ gilt.*

Satz 1.1.17. *Eine abgeschlossene, beschränkte und konvexe Menge $C \subset \mathbb{R}^n$ ist gleich der abgeschlossenen konvexen Hülle ihrer Extrempunkte.*

Bemerkung 1.1.18. Jedes konvexe Polytop ist die konvexe Hülle seiner Extrempunkte. Es kann gezeigt werden, dass jedes Polytop eine endliche Anzahl an Extrempunkten besitzt. Demnach ist ein konvexes Polyhedron gleich der konvexen Hülle von einer endlichen Anzahl von Extrempunkten.

Satz 1.1.19. Ein konvexes Polyhedron kann entweder durch einen beschränkten Schnitt von endlich vielen abgeschlossenen Halbräumen oder als konvexe Hülle von einer endlichen Anzahl an Punkten beschrieben werden.

1.2. Probleme ohne Nebenbedingungen. Nun betrachten wir unrestringierte Minimierungsprobleme. Die Problemstellung lautet

$$(1.1) \quad \text{minimiere } f(x) \quad \text{mit } x \in \Omega \subseteq \mathbb{R}^n.$$

In (1.1) ist f eine reellwertige Funktion, und Ω bezeichnet die zulässige Menge. Im folgenden sei $\|\cdot\|$ eine Norm auf $\mathbb{R}^n$. Wir wollen notwendige und hinreichende Optimalitätsbedingungen für (1.1) angeben.

Definition 1.2.1 (Relative Minimalstelle). Der Punkt $x^* \in \Omega$ heißt relative oder lokale Minimalstelle von f über Ω , wenn es ein $\epsilon > 0$ gibt, so dass $f(x) \geq f(x^*)$ für alle $x \in \Omega$ mit $\|x - x^*\| < \epsilon$ gilt. Wenn $f(x) > f(x^*)$ für alle $x \in \Omega$ mit $\|x - x^*\| < \epsilon$ gilt, so nennen wir x^* strikte relative Minimalstelle von f über Ω .

Definition 1.2.2 (Globale Minimalstelle). Der Punkt $x^* \in \Omega$ heißt globale Minimalstelle von f über Ω , wenn $f(x) \geq f(x^*)$ für alle $x \in \Omega$ gilt. Analog definieren wir eine strikte globale Minimalstelle.

Bemerkung 1.2.3. Sei $x \in \Omega$. Ein Vektor d heißt eine zulässige Richtung in x , wenn es ein $\bar{\alpha} > 0$ gibt, so dass $x + \alpha d \in \Omega$ für alle α mit $0 \leq \alpha \leq \bar{\alpha}$ gilt.

Satz 1.2.4 (Notwendige Bedingung 1. Ordnung). Seien $\Omega \subset \mathbb{R}^n$ und f eine Funktion aus der Menge $C^1(\Omega)$. Wenn x^* eine Stelle für eine relative Minimalstelle von f über Ω ist, dann folgt für jedes $d \in \mathbb{R}^n$, dass es eine zulässige Richtung in x^* angibt, die Ungleichung

$$\nabla f(x^*) d \geq 0,$$

wobei $\nabla f = \left(\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n}\right)$ den Gradienten von f bezeichnet.

Beweis. Für jedes α , $0 \leq \alpha \leq \bar{\alpha}$, gelten $x(\alpha) = x^* + \alpha d \in \Omega$ und $g(\alpha) = f(x(\alpha))$. Dann hat g an $x = 0$ ein relatives Minimum. Ferner folgt:

$$g(\alpha) - g(0) = g'(0)\alpha + o(\alpha).$$

Angenommen, es sei $g'(0) < 0$. Dann ergibt sich für genügend kleines $\alpha > 0$ die Ungleichung $g(0) > g(\alpha)$, was ein Widerspruch zum relativen Minimum von g an der Stelle $x = 0$ ist. Also: $0 \leq g'(0) = \nabla f(x^*) d$. $\square$

Bemerkung 1.2.5. Spezialfall $\Omega = \mathbb{R}^n$: Dann gilt $\nabla f(x^*) d \geq 0$ für alle $d \in \mathbb{R}^n$. Also folgt: $\nabla f(x^*) = 0$.

Korollar 1.2.6 (Unrestringierter Fall). Seien $\Omega \subseteq \mathbb{R}^n$ und $f \in C^1(\Omega)$. Wenn x^* ein innerer Punkt von Ω und eine relative Minimalstelle von f über Ω ist, dann folgt $\nabla f(x^*) = 0$.

Bemerkung 1.2.7. Notwendige Bedingungen führen im nicht-restringierten Fall auf n Gleichungen in n Unbekannten (n Komponenten von x^*).

Lemma 1.2.8 (Notwendige Bedingung 2. Ordnung). *Seien $f \in C^2(\Omega)$ und $\Omega \subseteq \mathbb{R}^n$. Wenn x^* eine relative Minimalstelle von f über Ω ist, dann gilt für jede zulässige Richtung $d \in \mathbb{R}^n$ in x^* :*

- a) $\nabla f(x^*) d \geq 0$
- b) Aus $\nabla f(x^*) d = 0$ folgt $d^T \nabla^2 f(x^*) d \geq 0$.

Beweis. $x(\alpha) = x^* + \alpha d$, $g(\alpha) = f(x(\alpha))$. Es ist nur b) zu beweisen. Zu a) verweisen wir auf Satz 1.2.4. Wegen $0 = \nabla f(x^*) d = g'(0)$ erhalten wir

$$g(\alpha) - g(0) = \frac{1}{2} g''(0) \alpha^2 + o(\alpha^2).$$

Angenommen, es sei $g''(0) < 0$. Dann folgt $g(\alpha) < g(0)$ für hinreichend kleines $\alpha > 0$, was einen Widerspruch zur Minimalität von $g(0)$ ergibt. Also: $0 \leq g''(0) = d^T \nabla^2 f(x^*) d$. $\square$

Korollar 1.2.9 (Notwendige Bedingung 2. Ordnung – unrestr. Fall). *Seien $x^* \in \Omega$ ein innerer Punkt und eine relative Minimalstelle von $f \in C^2(\Omega)$ über Ω . Dann gelten:*

- (1) $\nabla f(x^*) = 0$
- (2) Für alle $d \in \mathbb{R}^n$ folgt $d^T \nabla^2 f(x^*) d \geq 0$.

Bemerkung 1.2.10. *Wir schreiben im folgenden oft $F(x)$ für $\nabla^2 f(x)$. Die Matrix $F(x^*)$ hängt eng mit der Konvergenzrate von Algorithmen zusammen, die zur Minimierung von f dienen.*

Satz 1.2.11 (Hinreichende Bedingungen – unrestr. Fall). *Sei $f \in C^2(\Omega)$ eine Funktion mit $\Omega \subseteq \mathbb{R}^n$ und $x^* \in \text{int}(\Omega)$. Weiter gelten:*

- a) $\nabla f(x^*) = 0$ und
- b) $F(x^*)$ ist positiv definit.

Dann ist x^ eine strikte relative Minimalstelle von f .*

Beweis. Sei $F(x^*)$ positiv definit. Dann gibt es ein $\alpha > 0$ mit $d^T F(x^*) d \geq \alpha \|d\|^2$ für alle $d \in \mathbb{R}^n$. Die Taylor-Entwicklung liefert dann:

$$\begin{aligned} f(x^* + d) - f(x^*) &= \frac{1}{2} d^T F(x^*) d + o(\|d\|^2) \\ &\geq \frac{\alpha}{2} \|d\|^2 + o(\|d\|^2). \end{aligned}$$

Für kleines $\|d\|$ folgt $\alpha^2 \|d\|^2/2 \geq o(\|d\|^2)$. Also: $f(x^* + d) \geq f(x^*)$. $\square$

Im Falle von konvexen Funktionen können wir die Optimalitätsbedingungen noch besser charakterisieren.

Definition 1.2.12 (Strikt konvex). *Eine Funktion $f : \Omega \rightarrow \mathbb{R}$, $\Omega \subseteq \mathbb{R}^n$ konvex, wird strikt konvex genannt, wenn für jedes α mit $0 < \alpha < 1$ und für $x_1 \neq x_2$ gilt:*

$$f(\alpha x_1 + (1 - \alpha)x_2) < \alpha f(x_1) + (1 - \alpha)f(x_2).$$

Lemma 1.2.13. *Seien f_1 und f_2 konvexe Funktionen auf einer konvexen Menge Ω . Dann ist $f_1 + f_2$ konvex auf Ω .*

Beweis. Das Lemma folgt sofort aus der Definition der Konvexität und den Voraussetzungen. $\square$

Lemma 1.2.14. *Sei f eine konvexe Funktion auf einer konvexen Menge Ω . Dann ist auch αf für alle $\alpha \geq 0$ konvex auf Ω .*

Beweis. Das Lemma folgt ebenfalls sofort aus der Definition der Konvexität und den Voraussetzungen. $\square$

Bemerkung 1.2.15. Aus Lemma 1.2.13 und Lemma 1.2.14 folgt: $\alpha_1 f_1 + \dots + \alpha_m f_m$ ist konvex, falls $f_1, \dots, f_m$ konvex sind und $\alpha_i \geq 0$ gilt für $i = 1, \dots, m$.

Lemma 1.2.16. Sei f eine konvexe Funktion auf einer konvexen Menge Ω . Die Menge $\Gamma_c = \{x : x \in \Omega, f(x) \leq c\}$ ist dann auch konvex für alle $c \in \mathbb{R}$.

Beweis. Seien $x_1, x_2 \in \Gamma_c$, das heißt: $f(x_1) \leq c, f(x_2) \leq c$ für $0 < \alpha < 1$. Wir erhalten dann:

$$f(\alpha x_1 + (1 - \alpha)x_2) \leq \alpha f(x_1) + (1 - \alpha)f(x_2) \leq c.$$

Also folgt: $\alpha x_1 + (1 - \alpha)x_2 \in \Gamma_c$, so dass Γ_c konvex ist. $\square$

Bemerkung 1.2.17. Da der Schnitt von konvexen Mengen auch konvex ist, ist auch die Menge

$$\Gamma = \bigcap_{i=1}^m \{x : x \in \Omega, f_i(x) \leq c_i\}$$

für konvexe Funktionen $f_1, \dots, f_m$ und reelle Zahlen c_i konvex. Das ist wichtig für die mathematische Programmierung, weil die Menge, die die Restriktionen an die Zielfunktion f stellt, oft die Gestalt von Γ besitzt.

Wenn f differenzierbar ist, haben wir weitere Möglichkeiten, Konvexität zu charakterisieren.

Lemma 1.2.18. Seien $f \in C^1(\Omega)$ und $\Omega \subseteq \mathbb{R}^n$ konvex. f ist genau dann konvex, wenn

$$(1.2) \quad f(y) \geq f(x) + \nabla f(x)(y - x).$$

für alle $x, y \in \Omega$ gilt.

Bemerkung 1.2.19. Auf der rechten Seite der Ungleichung (1.2) steht die Gleichung der Tangente an f im Punkt x . Damit liegt der Wert von $f(y)$ immer oberhalb des Wertes der Tangente an dieser Stelle y .

Lemma 1.2.20. Sei $f \in C^2(\Omega)$, wobei die konvexe Menge $\Omega \subseteq \mathbb{R}^n$ einen inneren Punkt enthält. f ist genau dann konvex, wenn die Hessesche Matrix F von f positiv semidefinit auf Ω ist.

Bemerkung 1.2.21. Die Hessesche Matrix ist eine Verallgemeinerung des Konzeptes der Krümmung einer Funktion, das heißt, positive Definitheit der Hesseschen Matrix ist die Verallgemeinerung von positiver Krümmung. Konvexe Funktionen haben positive (oder zumindest nicht-negative) Krümmung in jeder Richtung. Diese Feststellungen veranlassen uns dazu, eine Funktion lokal konvex zu nennen, wenn die Hessesche Matrix positiv semidefinit in einer kleinen Umgebung ist (analog für eine lokal strikt konvexe Funktion). Damit ist f nach Lemma 1.2.11 am Minimalpunkt x^* lokal strikt konvex, und die lokale Untersuchung um eine Minimalstelle hängt eng mit der Konvexität zusammen. Ergebnisse für konvexe Funktionen können auf nicht-konvexe Funktionen angewandt werden, indem Umgebungen der Lösung untersucht werden, und umgekehrt.

Satz 1.2.22. Sei f eine konvexe Funktion, die auf der konvexen Menge Ω definiert ist. Dann ist die Menge Γ , auf der f sein Minimum erreicht, konvex. Jedes relative Minimum von f ist ein globales Minimum.

Beweis. Sei $\Gamma \neq \emptyset$. c_0 bezeichne das Minimum von f über Ω . Nach Lemma 1.2.16 ist die Menge $\Gamma = \{x : f(x) \leq c_0\} \subset \Omega$ konvex. $x^* \in \Omega$ sei eine relative Minimalstelle von f , $y \in \Omega$ mit $f(y) < f(x^*)$ und $0 < \alpha < 1$. Wir bekommen für die konvexe Funktion f :

$$f(\alpha y + (1 - \alpha)x^*) \leq \alpha f(y) + (1 - \alpha)f(x^*) < f(x^*).$$

Das ist aber ein Widerspruch zur Minimalität von f an x^* . $\square$

Bemerkung 1.2.23. *Bei konvexen Funktionen liegen alle Minimumstellen beieinander in einer konvexen Menge. Alle relativen Minima sind globale Minima. Der nächste Satz besagt, wenn f stetig differenzierbar und konvex ist, dann ist das Eintreten der notwendigen Bedingung erster Ordnung sowohl hinreichend als auch notwendig für ein globales Minimum.*

Satz 1.2.24. *Seien $\Omega \subset \mathbb{R}^n$ konvex und $f \in C^1(\Omega)$ konvex. Gibt es einen Punkt $x^* \in \Omega$, so dass für alle $y \in \Omega$ $\nabla f(x^*)(y - x^*) \geq 0$ gilt, dann ist x^* eine globale Minimalstelle von f auf Ω .*

Beweis. Der Vektor $y - x^*$ ist eine zulässige Richtung an x^* . Dann ist die Bedingung $\nabla f(x^*)(y - x^*) \geq 0$ identisch mit der notwendigen Bedingung erster Ordnung (vergleiche Lemma 1.2.4). Wegen Lemma 1.2.18 folgt sofort:

$$f(y) \geq f(x^*) + \underbrace{\nabla f(x^*)(y - x^*)}_{\geq 0} \geq f(x^*). \quad \square$$

Wir wenden uns jetzt noch dem Problem zu, eine konvexe Funktion auf einer konvexen Menge zu maximieren.

Satz 1.2.25. *Sei f eine konvexe Funktion, die auf einer beschränkten Menge Ω definiert sei. Wenn f ein Maximum auf Ω besitzt, so wird es an einem Extrempunkt von Ω erreicht.*

Wir beschäftigen uns jetzt mit dem Begriff der Konvergenzordnung, der in der Untersuchung von numerischen Verfahren eine wichtige Rolle spielt.

Definition 1.2.26 (Konvergenzordnung). *Sei $\{r_k\}_{k \in \mathbb{N}}$ eine Folge reeller Zahlen mit $\lim_{k \rightarrow \infty} r_k = r^*$. Die Konvergenzordnung von $\{r_k\}_{k \in \mathbb{N}}$ wird als Supremum der nicht-negativen Zahlen p definiert, die*

$$0 \leq \limsup_{k \rightarrow \infty} \frac{|r_{k+1} - r^*|}{|r_k - r^*|^p} < \infty$$

erfüllen.

Bemerkung 1.2.27. *Betrachte den Teil der Folge $\{r_k\}_{k \in \mathbb{N}}$ für $k \rightarrow \infty$, wo k sehr groß ist. Dann können wir sagen, dass die Konvergenzordnung ein Maß dafür ist, wie gut der schlechteste Teil der Folge (für große k) konvergiert. Große Werte von p bedeuten dann schnelle Konvergenz, da der Abstand zu r^* mindestens um das p -fache eines einzelnen Schrittes reduziert wird. Wenn die Folge die Ordnung p und den Grenzwert*

$$\beta = \lim_{k \rightarrow \infty} \frac{|r_{k+1} - r^*|}{|r_k - r^*|^p}$$

hat, dann gilt asymptotisch $|r_{k+1} - r^| = \beta |r_k - r^*|^p$.*

Definition 1.2.28 (Lineare Konvergenz). Wenn die Folge $\{r_k\}_{k \in \mathbb{N}}$ gegen r^* in der Form konvergiert, dass

$$\lim_{k \rightarrow \infty} \frac{|r_{k+1} - r^*|}{|r_k - r^*|} = \beta < 1$$

gilt, so nennen wir die Folge linear konvergent gegen r^* mit dem Konvergenzquotienten β .

Bemerkung 1.2.29. Eine linear konvergente Folge mit dem Konvergenzquotienten β kann interpretiert werden als eine Folge, die für große Werte von $k \in \mathbb{N}$ wie $c\beta^k$ konvergiert. Daher wird lineare Konvergenz oft auch als geometrische Konvergenz bezeichnet. Wenn wir zwei Algorithmen vergleichen, die beide linear konvergieren, müssen wir die entsprechenden Konvergenzquotienten vergleichen. Desto kleiner der Quotient ist, desto besser ist die Konvergenz. Der Fall $\beta = 0$ heißt superlineare Konvergenz. Die Konvergenz einer Folge mit der Konvergenzordnung größer als 1 ist superlinear:

$$\begin{aligned} 0 &\leq \lim_{k \rightarrow \infty} \frac{|r_{k+1} - r^*|}{|r_k - r^*|} = \lim_{k \rightarrow \infty} \left(\frac{|r_{k+1} - r^*|}{|r_k - r^*|^p} \frac{1}{|r_k - r^*|^{1-p}} \right) \\ &= \lim_{k \rightarrow \infty} \left(\frac{|r_{k+1} - r^*|}{|r_k - r^*|^p} |r_k - r^*|^{p-1} \right) \\ &\leq \underbrace{\overline{\lim}_{k \rightarrow \infty} \left(\frac{|r_{k+1} - r^*|}{|r_k - r^*|^p} \right)}_{< \infty} \lim_{k \rightarrow \infty} |r_k - r^*|^{p-1} \\ &= 0. \end{aligned}$$

Aber aus superlinearer Konvergenz kann auch folgen, dass die Konvergenzordnung gleich 1 ist.

1.3. Probleme mit Nebenbedingungen. Wir behandeln nun allgemeine nicht-lineare Probleme der Programmierung:

$$(1.3) \quad \left\{ \begin{array}{l} \text{minimiere} \quad f(x) \\ \text{mit} \quad \quad h_1(x) = 0, \quad g_1(x) \leq 0, \\ \quad \quad \quad \vdots \quad \quad \quad \vdots \\ \quad \quad \quad h_m(x) = 0, \quad g_p(x) \leq 0, \\ \quad \quad \quad x \in \Omega \subseteq \mathbb{R}^n. \end{array} \right.$$

In (1.3) setzen wir voraus, dass gelten: $m \leq n$; f , h_i mit $1 \leq i \leq m$ und g_j mit $1 \leq j \leq p$ sind stetig, oft auch zweimal stetig differenzierbar und $h = (h_1, \dots, h_m)$, $g = (g_1, \dots, g_p)$. Dann können wir (1.3) auch in der folgenden Form schreiben:

$$(1.4) \quad \text{minimiere } f(x) \quad \text{mit} \quad h(x) = 0, \quad g(x) \leq 0 \quad \text{und} \quad x \in \Omega \subseteq \mathbb{R}^n.$$

Die Restriktionen $h(x) = 0$ und $g(x) \leq 0$ bezeichnen wir als funktionale Nebenbedingungen. Ein Punkt $x \in \Omega$, der alle funktionalen Nebenbedingungen erfüllt, heißt zulässig. Eine Ungleichungs-Restriktion $g_i(x) \leq 0$ heißt aktiv an einem zulässigen Punkt x , wenn $g_i(x) = 0$ gilt, und inaktiv an x , wenn $g_i(x) < 0$ gilt. Als Konvention bezeichnen wir jede Gleichungs-Restriktion als aktiv an jedem zulässigen Punkt. Eine aktive Nebenbedingung an einem zulässigen Punkt x restringiert das Gebiet der zulässigen Punkte in der Nachbarschaft von x , während die inaktive Nebenbedingung lokal keinen Einfluß auf die zulässige Menge hat ($g_i(x) < 0 \Rightarrow g_i(x) \leq 0$).

und $g_i(\tilde{x}) < 0$ in einer Umgebung von x). Daher können wir uns bei der Untersuchung von Eigenschaften von lokalen Minimalstellen auf aktive Restriktionen beschränken.

Die Gleichungs-Restriktionen definieren eine Teilmenge auf $\mathbb{R}^n$: $h(x) = 0$. Diese Menge kann als Hyperfläche bezeichnet werden. Ist diese Hyperfläche überall regulär, dann hat sie die Dimension $n - m$. Wenn die Funktionen $h_i(x)$, $1 \leq i \leq m$, aus der Menge C^1 sind, dann ist die Hyperfläche auch glatt, und wir können zu jedem Punkt auf der Hyperfläche die Tangentialebene definieren.

In der restringierten Optimierung hat die Tangentialebene eine wesentliche Bedeutung.

Definition 1.3.1 (Regulärer Punkt). *Ein Punkt x^* mit $h(x^*) = 0$ heißt regulärer Punkt der Nebenbedingungen, wenn die Gradienten $\nabla h_1(x^*), \dots, \nabla h_m(x^*)$ linear unabhängig sind.*

Satz 1.3.2. *An einem regulären Punkt x^* der Fläche, die durch $h(x) = 0$ gegeben ist, ist die Tangentialebene gleich der Menge $M = \{y : \nabla h(x^*)y = 0\}$.*

Bemerkung 1.3.3. *Es ist wichtig, daran zu erinnern, dass die Bedingung, ein regulärer Punkt zu sein, keine Bedingung an die Fläche, sondern an ihre Repräsentation in Termen von h ist. Die Tangentialebene ist unabhängig von der Repräsentation definiert, während das für M nicht zutrifft.*

Zur Charakterisierung der notwendigen Optimalitätsbedingungen benötigen wir das folgende lemma.

Lemma 1.3.4. *Sei x^* ein regulärer Punkt der Nebenbedingung $h(x) = 0$ und eine lokale Extremstelle (Maximum oder Minimum an x^*). Dann gilt*

$$\nabla f(x^*)y = 0$$

für alle $y \in \mathbb{R}^n$ mit $\nabla h(x^*)y = 0$.

Beweis. Seien y ein Vektor in der Tangentialebene an x^* und $x(t)$ eine glatte Kurve auf der Fläche, die durch die Nebenbedingung beschrieben wird und die durch x^* geht: $x(0) = x^*$, $\dot{x}(0) = y$ und $h(x(t)) = 0$ für $-\alpha \leq t \leq \alpha$ für ein $\alpha > 0$. Da x^* regulär ist, ist die Tangentialebene identisch mit der Menge der y 's mit $\nabla h(x^*)y = 0$. Weil x^* lokale Extremstelle von f ist, gilt:

$$\left. \frac{d}{dt} f(x(t)) \right|_{t=0} = 0.$$

Also folgt $\nabla f(x^*)y = 0$. □

Bemerkung 1.3.5. *Das Lemma zeigt, dass $\nabla f(x^*)$ orthogonal zur Tangentialebene steht. Der nächste Satz sagt aus, dass $\nabla f(x^*)$ eine Linearkombination von Gradienten von h an x^* ist.*

Satz 1.3.6 (Notwendige Bedingung 1. Ordnung). *Sei x^* eine lokale Extremstelle von f unter der Nebenbedingung $h(x) = 0$. Weiter sei x^* regulär bezüglich dieser Nebenbedingungen. Dann gibt es ein $\lambda \in \mathbb{R}^m$, so dass*

$$\nabla f(x^*) + \lambda^T \nabla h(x^*) = 0$$

gilt.

Beweis. Das obige Lemma 1.3.4 ergibt, dass der Wert des linearen Programms

$$\text{maximiere } \nabla f(x^*) y \quad \text{mit} \quad \nabla h(x^*) y = 0$$

Null ist. Dann folgt aus dem Dualitätssatz (siehe [1, Seite 89]), dass das duale Problem zulässig ist. Speziell gibt es ein $\lambda \in \mathbb{R}^m$, so dass $\nabla f(x^*) + \lambda^T \nabla h(x^*) = 0$ erfüllt ist. $\square$

Bemerkung 1.3.7. Die notwendige Bedingung erster Ordnung

$$\nabla f(x^*) + \lambda^T \nabla h(x^*) = 0$$

ergibt zusammen mit der Nebenbedingung

$$h(x^*) = 0$$

ein Gleichungssystem mit $n + m$ (im allgemeinen) nichtlinearen Gleichungen in den $n + m$ Variablen $x^* \in \mathbb{R}^n$ und $\lambda \in \mathbb{R}^m$. Wir führen die Lagrange-Funktion ein, die mit dem restringierten Problem zusammenhängt:

$$\mathcal{L}(x, \lambda) = f(x) + \lambda^T h(x).$$

Die notwendigen Bedingungen können dann in der Form

$$\nabla_x \mathcal{L}(x, \lambda) = 0 \quad \text{und} \quad \nabla_\lambda \mathcal{L}(x, \lambda) = 0$$

geschrieben werden, wobei die letzte Gleichung eine Formulierung der Nebenbedingung darstellt.

Wir wollen nun Optimalitätsbedingungen zweiter Ordnung formulieren. Dazu nehmen wir an, dass f und h zweimal stetig differenzierbar sind.

Satz 1.3.8 (Notwendige Bedingungen 2. Ordnung). x^* sei eine lokale Minimalstelle von f mit $h(x^*) = 0$ und ein regulärer Punkt bezüglich der Nebenbedingungen $h(x) = 0$. Dann gibt es ein $\lambda \in \mathbb{R}^m$ mit

$$\nabla f(x^*) + \lambda^T \nabla h(x^*) = 0.$$

Wenn wir mit M die Tangentialebene $M = \{y : \nabla h(x^*) y = 0\}$ bezeichnen, dann ist die Matrix

$$L(x^*) = F(x^*) + \lambda^T H(x^*)$$

positiv semidefinit auf M , das heißt: $y^T L(x^*) y \geq 0$ für alle $y \in M$.

Beweis: Es ist klar, dass wir jede zweimal stetig differenzierbare Funktion auf der Fläche S , die durch die Nebenbedingungen gegeben ist und durch den Punkt x^* ($x(0) = x^*$) geht, die Ungleichung

$$(1.5) \quad \left. \frac{d^2}{dt^2} f(x(t)) \right|_{t=0} \geq 0$$

erhalten. Ferner gilt:

$$\left. \frac{d^2}{dt^2} f(x(t)) \right|_{t=0} = \dot{x}(0)^T F(x^*) \dot{x}(0) + \nabla f(x^*) \ddot{x}(0).$$

Wenn wir die Beziehung $\lambda^T h(x) = 0$ zweimal an $t = 0$ differenzieren, folgt

$$\dot{x}(0)^T H(x^*) \dot{x}(0) + \lambda^T \nabla h(x^*) \ddot{x}(0) = 0.$$

Wenn wir beide Gleichungen addieren und die Ungleichung (1.5) beachten, ergibt sich:

$$\left. \frac{d^2}{dt^2} f(x(t)) \right|_{t=0} = \dot{x}(0)^T L(x^*) \dot{x}(0) \geq 0.$$

Da $\dot{x}(0)$ in M liegt, folgt die Behauptung. $\square$

Bemerkung 1.3.9. Im Satz 1.3.8 haben wir zum ersten Mal die Matrix $L = F + \lambda^T H$ verwandt, welche die zweite Ableitung von der Lagrange-Funktion bezüglich x ist. Diese Matrix L bildet das Rückgrat für die Theorie von Algorithmen für das restringierte Problem.

Satz 1.3.10 (Hinreichende Bedingungen zweiter Ordnung). Seien x^* ein Punkt mit $h(x^*) = 0$ und $\lambda \in \mathbb{R}^m$, so dass

$$\nabla f(x^*) + \lambda^T \nabla h(x^*) = 0$$

gilt. Weiter sei die Matrix $L(x^*) = F(x^*) + \lambda^T H(x^*)$ positiv definit auf $M = \{y : \nabla h(x^*) y = 0\}$. Dann ist x^* eine strikte lokale Minimalstelle von f unter der Nebenbedingung $h(x) = 0$.

Die hinreichenden Bedingungen zweiter Ordnung können eingesetzt werden, um die Sensitivität des Optimierungsproblems zu untersuchen. Wir betrachten das Problem

$$(1.6) \quad \text{minimiere } f(x) \quad \text{mit } h(x) = 0.$$

Sei x^* eine Lösung von (1.6), die regulär bezüglich der Nebenbedingungen ist, und λ der zugehörige Lagrangesche Multiplikator. Dazu schauen wir uns das Problem

$$(1.7) \quad \text{minimiere } f(x) \quad \text{mit } h(x) = c$$

an, wobei $c \in \mathbb{R}^m$ gilt. Für hinreichend kleines $\|c\|$ hat das Problem (1.7) eine Lösung $x(c)$ nahe bei $x(0) \equiv x^*$. Für jede dieser Lösungen gibt es einen entsprechenden Wert $f(x(c))$, und dieser Wert kann als Funktion von c angesehen werden. Die Komponenten des Gradienten dieser Funktion können als Anstieg im Wert pro Einheitswechsel in den Nebenbedingungen betrachtet werden.

Satz 1.3.11 (Sensitivitätssatz). Seien $f, h \in C^2$ und $c \in \mathbb{R}^m$. Betrachte das Problem (1.7). Für $c = 0$ gebe es eine lokale Lösung, die regulärer Punkt ist und mit dem zugehörigen Lagrangeschen Multiplikator die hinreichende Bedingung zweiter Ordnung für eine strikte lokale Minimalstelle erfüllt. Dann gibt es für jedes $c \in \mathbb{R}^m$ in einer Umgebung, die 0 enthält, ein $x(c)$, das stetig von c abhängt, so dass $x(0) = x^*$ gilt und $x(c)$ lokale Lösung von (1.7) ist. Weiter folgt:

$$\nabla_c f(x(c))|_{c=0} = -\lambda^T.$$

Nun wollen wir Ungleichungs-Restriktionen betrachten. Dazu wenden wir uns nun dem Problem

$$(1.8) \quad \text{minimiere } f(x) \quad \text{mit } h(x) = 0 \text{ und } g(x) \leq 0$$

zu. Die Dimensionen von h und f seien wie bisher, g eine p -dimensionale Funktion. Zunächst gelte $f, g, h \in C^1$.

Definition 1.3.12 (Regulärer Punkt). Seien x^* ein Punkt, der die Nebenbedingungen $h(x^*) = 0$ und $g(x^*) \leq 0$ erfüllt, und J die Menge der Indizes j mit $g_j(x^*) = 0$. Dann heißt x^* regulärer Punkt bezüglich der Nebenbedingungen, wenn die Vektoren $\nabla h_1(x^*), \dots, \nabla h_m(x^*)$ und $\nabla g_j(x^*)$, $j \in J$, linear unabhängig sind.

Bemerkung 1.3.13. Ein Punkt x^* ist ein regulärer Punkt, wenn die Gradienten der an x^* aktiven Nebenbedingungen linear unabhängig sind.

Satz 1.3.14 (Kuhn-Tucker-Bedingungen). *Sei x^* eine relative Minimalstelle für das Problem (1.8) und ein regulärer Punkt bezüglich der Nebenbedingungen. Dann gibt es ein $\lambda \in \mathbb{R}^m$ und ein $\mu \in \mathbb{R}^p$ mit $\mu \geq 0$, so dass*

$$(1.9a) \quad \nabla f(x^*) + \lambda^T \nabla h(x^*) + \mu^T \nabla g(x^*) = 0,$$

$$(1.9b) \quad \mu^T g(x^*) = 0$$

gelten.

Beweis. Da $\mu \geq 0$ und $g(x^*) \leq 0$ gelten, ist die Gleichung $\mu^T g(x^*)$ äquivalent zu der Aussage, dass eine Komponente von $\mu^T g(x^*) = 0$ nur dann ungleich Null ist, wenn die entsprechende Nebenbedingung aktiv ist.

Weil x^* eine relative Minimalstelle auf der Restriktionsmenge Ω ist, ist x^* auch eine relative Minimalstelle bezüglich der aktiven Nebenbedingungen. Für die aktiven Nebenbedingungen gibt es Lagrangesche Multiplikatoren. Wir erhalten die Gleichung (1.9a) mit $\mu_j = 0$ für alle j mit $g_j(x^*) \neq 0$. Daher gilt offensichtlich (1.9b). Wir müssen nur noch $\mu \geq 0$ zeigen. Angenommen, es gebe ein $\mu_k < 0$ mit $k \in J$. Seien S die Fläche und M die Tangentialebene an x^* , die beide durch die anderen Nebenbedingungen definiert seien:

$$S = \{x \in \mathbb{R}^n : h(x^*) = 0 \wedge g_j(x^*) \leq 0 \quad \forall j \in J \setminus \{k\}\},$$

$$M = \{y \in \mathbb{R}^n : \nabla h(x^*) y = 0 \wedge \nabla g_j(x^*) y = 0 \quad \forall j \in J \setminus \{k\}\}.$$

Da x^* regulär ist, gibt es ein $y \in M$ mit $\nabla g_k(x^*) y < 0$: $\nabla g_k(x^*)$ kann nicht orthogonal auf M stehen, da wir es sonst wegen $\nabla h(x^*) y = 0$ durch eine Linearkombination der Zeilenvektoren von $\nabla h(x^*)$ ausdrücken könnten. $x(t)$ bezeichne eine Kurve auf S mit $x(0) = x^*$ und $\dot{x} = y$. Für kleines $t \geq 0$ ist $x(t)$ zulässig, denn die Ableitung von $g_k(x^*)$ in Richtung y ist negativ und damit $g_k(x) \leq 0$ lokal erfüllt, und wir erhalten

$$\left. \frac{df}{dt}(x(t)) \right|_{t=0} = \nabla f(x^*) y.$$

Wegen $y \in M$ folgt aus (1.9a):

$$\nabla f(x^*) y = - \underbrace{\nabla h(x^*) y}_{=0} - \underbrace{\mu_k \nabla g_k(x^*) y}_{>0} - \sum_{\substack{j \in J \\ j \neq k}} \mu_j \underbrace{\nabla g_j(x^*) y}_{=0} < 0.$$

Also gilt $\nabla f(x^*) y < 0$, was einen Widerspruch zu Eigenschaft von x^* darstellt, eine relative Minimalstelle zu sein. $\square$

Satz 1.3.15 (Notwendige Bedingungen 2. Ordnung). *Seien $f, g, h \in C^2$ und x^* ein regulärer Punkt der Nebenbedingungen $h(x^*) = 0$ und $g(x^*) \leq 0$. Wenn x^* eine relative Minimalstelle für (1.8) ist, dann gibt es $\lambda \in \mathbb{R}^m$, $\mu \in \mathbb{R}^p$ mit $\mu \geq 0$, so dass*

$$\begin{aligned} \nabla f(x^*) + \lambda^T \nabla h(x^*) + \mu^T \nabla g(x^*) &= 0, \\ \mu^T g(x^*) &= 0 \end{aligned}$$

gelten und

$$L(x^*) = F(x^*) + \lambda^T H(x^*) + \mu^T G(x^*)$$

positiv semidefinit auf der Tangentialebene der Nebenbedingungen an x^* ist.

Bemerkung 1.3.16. Es ist auch wieder möglich, hinreichende Bedingungen zweiter Ordnung anzugeben. Wir könnten vermuten, dass die erforderliche Bedingung lautet: $L(x^*)$ ist positiv definit auf M . Das ist auch in den meisten Fällen so. Wenn es aber degenerierte Ungleichungs-Restriktionen gibt (das heißt, die aktiven Ungleichungs-Nebenbedingungen haben Null als zugehörigen Lagrange-Multiplikator), so müssen wir die positive Definitheit auf einem Teilraum fordern, der größer als M ist.

Satz 1.3.17 (Hinreichende Bedingungen 2. Ordnung). Seien f, g und h aus C^2 . x^* mit $h(x^*) = 0$ und $g(x^*) \leq 0$ ist eine strikte relative Minimalstelle des Problems (1.8), wenn ein $\lambda \in \mathbb{R}^m$ und ein $\mu \in \mathbb{R}^p$ existieren, so dass

$$\begin{cases} \mu \geq 0 \\ \mu^T g(x^*) = 0, \\ \nabla f(x^*) + \lambda^T \nabla h(x^*) + \mu^T \nabla g(x^*) = 0 \end{cases}$$

gelten und die Hessematrix

$$L(x^*) = F(x^*) + \lambda^T H(x^*) + \mu^T G(x^*)$$

positiv definit auf dem Teilraum

$$M' = \{y : \nabla h(x^*) y = 0, \nabla g_j(x^*) y = 0 \quad \forall j \in J\}$$

mit $J = \{j : g_j(x^*) = 0, \mu_j \geq 0\}$ ist.

Bemerkung 1.3.18. Wir bemerken, wenn alle aktiven Ungleichungs-Restriktionen strikt positive Lagrange-Multiplikatoren haben (also keine degenerierten Ungleichungen), dass dann die Menge J alle aktiven Ungleichungen enthält. Dann ist die hinreichende Bedingung: $L(x^*)$ ist positiv definit auf M , der Tangentialebene der aktiven Restriktionen.

Satz 1.3.19 (Sensitivitätssatz). Seien f, g und $h \in C^2$. Betrachte dann das Problem

$$(1.10) \quad \text{minimiere } f(x) \quad \text{mit } h(x) = c \text{ und } g(x) \leq d.$$

Für $c = 0$ und $d = 0$ existiere lokal eine eindeutige Lösung x^* , die regulär ist. Mit den Lagrangeschen-Multiplikatoren λ und $\mu \geq 0$ gelten die hinreichenden Bedingungen zweiter Ordnung für eine strikte lokale Minimalstelle. Keine aktive Ungleichungs-Restriktion sei degeneriert. Dann gibt es für alle $(c, d) \in \mathbb{R}^{m+p}$ eine Umgebung von $(0, 0)$, in der eine Lösung $x(c, d)$ existiert, die stetig von (c, d) abhängt, so dass $x(0, 0) = x^*$ und $x(c, d)$ eine relative Minimalstelle von (1.10) sind. Weiter gelten:

$$\begin{aligned} \nabla_c f(x(c, d))|_{(0,0)} &= -\lambda^T, \\ \nabla_d f(x(c, d))|_{(0,0)} &= -\mu^T. \end{aligned}$$

1.4. Verfahren. In diesem Abschnitt beschäftigen wir uns kurz mit Penalty-, Barriere- und augmentierten Lagrange-Verfahren. Beginnen werden wir mit der Penalty-Methode. Wir betrachten das Problem:

$$(1.11) \quad \text{minimiere } f(x) \quad \text{mit } x \in S.$$

In (1.11) sind $f : \mathbb{R}^n \rightarrow \mathbb{R}$ stetig und S eine Restriktionsmenge in $\mathbb{R}^n$. In den meisten Anwendungen wird S implizit durch eine Zahl von funktionalen Nebenbedingungen bestimmt. Hier wollen wir aber allgemein Probleme der Form (1.11) behandeln. Wir ersetzen (1.11) durch ein unrestringiertes Problem

$$(1.12) \quad \text{minimiere } f(x) + cP(x),$$

wobei c eine Konstante und P eine reellwertige Funktion auf $\mathbb{R}^n$ sind mit

- a) P ist stetig,
- b) $P(x) \geq 0$ für alle $x \in \mathbb{R}^n$ und
- c) $P(x) = 0$ genau dann, wenn $x \in S$.

Beispiel 1.4.1. Für $S = \{x : g_i(x) \leq 0, i = 1, \dots, p\}$ definieren wir

$$P(x) = \frac{1}{2} \sum_{i=1}^p \left(\max(0, g_i(x)) \right)^2.$$

Für große Werte von c ist klar, dass die Minimalstelle des Problems (1.12) in einer Umgebung sein wird, in der $P(x)$ klein ist. Die entsprechenden Lösungen nähern sich für $c \rightarrow \infty$ der zulässigen Menge S an. Wenn S abgeschlossen ist, wird f durch (1.12) minimiert werden.

Sei $\{c_k\}_{k \in \mathbb{N}}$ eine Folge, die gegen unendlich konvergiert, so dass für jedes k sowohl $c_k \geq 0$ als auch $c_{k+1} > c_k$ gelten. Wir definieren:

$$q(c, x) = f(x) + cP(x)$$

und lösen für jedes k das Problem

$$(1.13) \quad \text{minimiere } q(c_k, x).$$

Wir erhalten einen Lösungspunkt x_k . Angenommen, (1.13) hat für jedes k eine eindeutig bestimmte Lösung.

Das folgende Lemma gibt einige Ungleichungen an, die direkt aus der Definition der x_k und der Ungleichung $c_{k+1} > c_k$ folgen.

Lemma 1.4.2. a) Es gelten die drei Ungleichungen:

$$\begin{aligned} q(c_k, x_k) &\leq q(c_{k+1}, x_{k+1}), \\ P(x_k) &\geq P(x_{k+1}), \\ f(x_k) &\leq f(x_{k+1}). \end{aligned}$$

- b) Sei x^* eine Lösung von (1.12). Dann gilt für jedes k : $f(x^*) \geq q(c_k, x_k) \geq f(x_k)$.

Satz 1.4.3. Sei $\{x_k\}_{k \in \mathbb{N}}$ eine Folge, die durch die Penalty-Methode generiert wird. Dann ist jeder Grenzwert der Folge eine Lösung von (1.12).

Beweis. Sei $\{x_{k_j}\}_{j \in \mathbb{N}}$ eine konvergente Teilfolge von $\{x_k\}_{k \in \mathbb{N}}$ mit Grenzwert $\bar{x}$. Aus der Stetigkeit von f folgt dann

$$(1.14) \quad \lim_{j \rightarrow \infty} f(x_{k_j}) = f(\bar{x}).$$

Sei x^* die optimale Lösung von (1.12). Dann ergibt sich aus dem Lemma 1.4.2, dass die Folge $\{q(c_{k_j}, x_{k_j})\}_{j \in \mathbb{N}}$ nichtfallend und durch $f(x^*) = f^*$ nach unten beschränkt ist:

$$(1.15) \quad \lim_{j \rightarrow \infty} q(c_{k_j}, x_{k_j}) = q^* \leq f^*.$$

Aus den Gleichungen (1.14) und (1.15) erhalten wir:

$$\lim_{j \rightarrow \infty} c_{k_j} P(x_{k_j}) = q^* - f(\bar{x}).$$

Wegen $P(x_k) \geq 0$ und $\lim_{k \rightarrow \infty} c_k = \infty$ folgt

$$\lim_{j \rightarrow \infty} P(x_{k_j}) = 0.$$

Weiter ist P stetig, das bedeutet: $\lim_{j \rightarrow \infty} P(x_{k_j}) = P(\bar{x}) = 0$. Damit ist $\bar{x}$ zulässig für das Problem (1.12). Es bleibt zu zeigen, dass $\bar{x}$ optimal für (1.12) ist. Dieses ergibt sich aber sofort aus dem Lemma 1.4.2-b):

$$f^* \geq \lim_{j \rightarrow \infty} f(x_{k_j}) = f(\bar{x}). \quad \square$$

Nun kommen wir zu den Barriere-Verfahren. Diese Methoden sind für Probleme der Form:

$$(1.16) \quad \text{minimiere } f(x) \quad \text{mit } x \in S.$$

Die Restriktionsmenge S habe ein nicht-leeres Inneres, und es sei möglich, jeden Randpunkt durch Punkte im Innern zu approximieren. Eine solche Menge nennen wir **Robust**. Diese Art von Mengen begegnet uns oft im Zusammenhang mit den Ungleichungs-Restriktionen, wo S die Form $S = \{x : g_i(x) \leq 0, i = 1, \dots, p\}$ hat. Barriere-Methoden arbeiten so, dass am Rand der zulässigen Menge eine Barriere errichtet wird, die dafür sorgt, dass eine Suchprozedur die Menge S nicht verläßt. Eine Barriere-Funktion ist eine Funktion B , die auf dem Innern von S definiert ist, so dass gelten:

- a) B ist stetig,
- b) $B(x) \geq 0$ und
- c) $B(x) \rightarrow \infty$, wenn x sich dem Rand von S nähert.

Beispiel 1.4.4. Seien $g_i, i = 1, \dots, p$ stetige Funktionen auf $\mathbb{R}^n$. Angenommen, die Menge $S = \{x : g_i(x) \leq 0, i = 1, \dots, p\}$ ist ein Robust und das Innere von S ist die Menge aller x , für die $g_i(x) < 0, i = 1, \dots, p$, gilt. Dann ist die Funktion

$$B(x) = - \sum_{i=1}^p \frac{1}{g_i(x)},$$

die im Innern von S definiert ist, eine Barriere-Funktion.

Wir betrachten das restringierte Problem mit veränderten Nebenbedingungen:

$$\text{minimiere } f(x) + \frac{1}{c} B(x) \quad \text{mit } S \in \text{int}(S).$$

Der Vorteil dieses Problems ist aber, dass es mittels einer Suchtechnik für nicht-restringierte Probleme gelöst werden kann. Dabei müssen wir in $\text{int}(S)$ starten. Mit Hilfe von Abstiegsmethoden kann ein neuer Punkt gesucht werden. Da der Wert der Zielfunktion an ∂S unendlich wird, bleibt die Suche in $\text{int}(S)$. Die Nebenbedingungen brauchen also nicht explizit geprüft werden.

Sei $\{c_k\}_{k \in \mathbb{N}}$ eine Folge mit $\lim_{k \rightarrow \infty} c_k = \infty, c_k \geq 0$ für alle $k \in \mathbb{N}$ und $c_{k+1} > c_k$. Definiere

$$r(c, x) = f(x) + \frac{1}{c} B(x).$$

Für alle $k \in \mathbb{N}$ löse das Problem:

$$(1.17) \quad \text{minimiere } r(c_k, x) \quad \text{mit } S \in \text{int}(S).$$

Die Lösung des Problems (1.17) bezeichnen wir mit x_k .

Satz 1.4.5. *Jeder Grenzwert der Folge $\{x_k\}_{k \in \mathbb{N}}$, die durch die Barriere-Methode generiert wird, ist eine Lösung des Problems (1.16).*

Penalty- und Barriere-Methoden sind anwendbar für nichtlineare Probleme, die eine sehr allgemeine Form der Restriktionsmenge S haben. In den meisten Situationen ist aber die Menge S nicht explizit gegeben, sondern implizit durch eine Anzahl von funktionalen Nebenbedingungen. In diesen Fällen ist die Penalty- oder Barriere-Funktion in Termen der Restriktionsfunktionen selbst definiert.

Betrachte Probleme der Form:

$$(1.18) \quad \text{minimiere } f(x) \quad \text{mit } g_i(x) \leq 0 \text{ für } i = 1, \dots, p.$$

Eine Penalty-Funktion für dieses Problem wird meistens ausgedrückt in den Hilfsfunktionen

$$g_i^+(x) \equiv \max(0, g_i(x)), \quad i = 1, \dots, p.$$

Im Innern der Nebenbedingungsmenge gilt $P(x) \equiv 0$, und somit ist P nur eine Funktion der verletzten Restriktionen. Seien $g^+(x) = (g_1^+(x), \dots, g_p^+(x))^T$ und

$$P(x) = \gamma(g^+(x)),$$

wobei $\gamma : \mathbb{R}^p \rightarrow \mathbb{R}$ stetig ist und so gewählt wird, dass P die Eigenschaften der Penalty-Funktion erfüllt.

Beispiel 1.4.6. a) $\gamma(y) = \frac{1}{2} \sum_{i=1}^p y_i^2$,

b) $\gamma(y) = y^T \Gamma y$, $\Gamma = \Gamma^T$ positiv definite $(p \times p)$ -Matrix und

c) $\gamma(y) = \sum_{i=1}^p y_i^\epsilon$ mit $\epsilon > 0$.

In der Penalty-Methode lösen wir für verschiedene c_k das nicht-restringierte Problem

$$(1.19) \quad \text{minimiere } f(x) + c_k P(x).$$

Die meisten Algorithmen erfordern, dass die Zielfunktion stetige erste Ableitungen besitzt. Wir nehmen daher $f, g, P \in C^1$ an.

$$\nabla g_i^+(x) = \begin{cases} \nabla g_i(x), & \text{wenn } g_i(x) \geq 0, \\ 0, & \text{wenn } g_i(x) < 0, \end{cases}$$

$$\nabla g^+(x) = \begin{pmatrix} \nabla g_1^+(x) \\ \vdots \\ \nabla g_p^+(x) \end{pmatrix} \in \mathbb{R}^{p \times n}.$$

∇g^+ ist an den Punkten x mit $g_i^+(x) = 0$ nicht stetig für einige $i = 1, \dots, p$. Daher müssen einige Restriktionen an γ gestellt werden, um $P \in C^1$ zu garantieren. Sei $\gamma \in C^1(\mathbb{R}^p)$ und, wenn $y = (y_1, \dots, y_p)^T$, $\nabla \gamma(y) = (\nabla \gamma_1(y), \dots, \nabla \gamma_n(y))$ ist, gelte

$$(1.20) \quad y_i = 0 \implies \nabla \gamma_i(y) = 0.$$

Dann ist die Ableitung $\nabla\gamma(g^+(x)) \nabla g(x)$ von $\gamma(g^+(x))$ stetig bezüglich x . Diese Betrachtungen sind nur für Ungleichungs-Restriktionen notwendig. Problem (1.19) wird dann eine Lösung x_k haben, für die gilt:

$$\nabla f(x_k) + c_k \nabla\gamma(g^+(x_k)) \nabla g(x_k) = 0.$$

Oder anders geschrieben:

$$\nabla f(x_k) + \lambda_k^T \nabla g(x_k) = 0,$$

wobei

$$\lambda_k^T \equiv c_k \nabla\gamma(g^+(x_k))$$

für jedes c_k ein Lagrange-Multiplikator ist. Wenn x^* ein regulärer Punkt von (1.18) ist, dann hängt der eindeutige Lagrange-Multiplikator λ^* mit der Lösung zusammen.

Lemma 1.4.7. *Wir wenden die Penalty-Methode auf das Problem (1.18) an, indem wir eine Penalty-Funktion der Form*

$$P(x) = \gamma(g^+(x))$$

mit $\gamma \in C^1$ und Gültigkeit von (1.20) verwenden. Zu der generierten Folge $(x_k)_{k \in \mathbb{N}}$ definieren wir $\lambda_k^T := c_k \nabla\gamma(g^+(x_k))$. Wenn $x_k \rightarrow x^*$ für $k \rightarrow \infty$ konvergiert und x^* Lösung von (1.18) ist, die regulärer Punkt ist, dann folgt

$$\lambda_k \xrightarrow{k \rightarrow \infty} \lambda^*.$$

λ^* ist der Lagrange-Multiplikator von dem Problem (1.18).

Bemerkung 1.4.8. *Die Hessesche Matrix des Problems ist allerdings schlecht konditioniert.*

Genau das gleiche Resultat erhalten wir für die Penalty-Methoden bei Barriere-Funktionen. Wenn wir für das Problem (1.18) die Barriere-Funktion der Form

$$B(x) = \eta(g(x))$$

betrachten, so bekommen wir die Lagrangeschen Multiplikatoren und wieder eine schlecht konditionierte Hessesche Matrix.

Im folgenden wenden wir uns dem Thema Dualität zu. Wir betrachten wieder das Problem

$$(1.21) \quad \text{minimiere } f(x) \quad \text{mit } h(x) = 0,$$

wobei $x \in \mathbb{R}^n$, $h(x) \in \mathbb{R}^m$, $f, h \in C^2$ gelten. Alles, was wir jetzt tun, kann leicht auf Probleme übertragen werden, die Ungleichungs-Restriktionen haben. Sei x^* eine lokale Lösung von (1.21), die regulärer Punkt bezüglich der Nebenbedingungen ist. Dann gibt es einen Lagrangeschen Multiplikator, so dass

$$\nabla f(x^*) + (\lambda^*)^T \nabla h(x^*) = 0$$

gilt und die Hessesche Matrix der Lagrange-Funktion

$$L(x^*) = F(x^*) + (\lambda^*)^T H(x^*)$$

auf dem Tangentialraum

$$M = \{x : \nabla h(x^*) x = 0\}$$

positiv semidefinit sein muß.

Wir führen nun die spezielle lokale Konvexitätsannahme ein, die notwendig für die Entwicklung der lokalen Dualitätstheorie ist: Die Matrix $L(x^*)$ sei positiv definit, und zwar nicht nur auf M , sondern auf dem ganzen Raum $\mathbb{R}^n$. Dann ist die Lagrange-Funktion $\mathcal{L}(x, \lambda) = f(x) + \lambda^T h(x)$ lokal konvex an x^* . Mit dieser Annahme folgt, dass x^* nicht nur eine lokale Lösung von (1.21), sondern auch eine lokale Lösung des nicht-restringierten Problems

$$\text{minimiere } f(x) + (\lambda^*)^T h(x)$$

ist, da x^* die hinreichenden Kriterien erster und zweiter Ordnung für ein lokales Minimum erfüllt. Weiter hat die Funktion $f(x) + \lambda^T h(x)$ für jedes λ nahe bei λ^* eine lokale Minimalstelle x in der Nähe von x^* . Das folgt aus dem Satz über implizite Funktionen, da $L(x^*)$ regulär ist. Dann ist nämlich auch die Hessesche Matrix $F(x) + \lambda^T H(x)$ lokal positiv definit. Es gibt also lokal eine eindeutige Zuordnung von x und λ durch die Lösung des nicht-restringierten Problems

$$(1.22) \quad \text{minimiere } f(x) + \lambda^T h(x).$$

Ferner ist diese Beziehung stetig differenzierbar. In einer Umgebung von λ^* definieren wir die duale Funktion ϕ durch die Gleichung

$$(1.23) \quad \phi(\lambda) = \min_x (f(x) + \lambda^T h(x)),$$

wobei das Minimum lokal hinsichtlich x nahe bei x^* gesucht ist. Wir sind dann in der Lage zu zeigen, dass das ursprüngliche restringierte Problem (1.21) lokal äquivalent mit der nicht-restringierten Maximierung der dualen Funktion ϕ bezüglich λ . Mit $x(\lambda)$ bezeichnen wir die eindeutige Lösung von (1.22) in einer Umgebung von x^* .

Lemma 1.4.9. *Die duale Funktion ϕ hat den Gradienten $\nabla\phi(\lambda) = h(x(\lambda))^T$.*

Beweis. Von der Gleichung (1.23) erhalten wir:

$$\phi(\lambda) = f(x(\lambda)) + \lambda^T h(x(\lambda)).$$

Also gilt:

$$\nabla\phi(\lambda) = (\nabla f(x(\lambda)) + \lambda^T \nabla h(x(\lambda))) \nabla x(\lambda) + h(x(\lambda))^T.$$

Da der erste Term auf der ersten Seite aufgrund der Definition von $x(\lambda)$ verschwindet, folgt: $\nabla\phi(\lambda) = h(x(\lambda))^T$. $\square$

Bemerkung 1.4.10. *Der Gradient der dualen Funktion ist also einfach zu berechnen. Die Hessesche Matrix der dualen Funktion kann in Termen der Hesseschen Matrix der Lagrange-Funktion ausgedrückt werden. Wir verwenden die Notation*

$$L(x, \lambda) = F(x) + \lambda^T H(x),$$

um die Abhängigkeit von λ darzustellen.

Lemma 1.4.11. *Die Hessesche Matrix der dualen Funktion ist durch*

$$\Phi(\lambda) = -\nabla h(x(\lambda)) L^{-1}(x(\lambda), \lambda) \nabla h(x(\lambda))^T$$

gegeben.

Satz 1.4.12 (Lokaler Dualitätssatz). *Betrachte das Problem*

$$\text{minimiere } f(x) \quad \text{mit } h(x) = 0.$$

x^* sei eine lokale Lösung mit dem entsprechenden Wert r^* und dem Lagrange-Multiplikator λ^* . Weiter seien x^* ein regulärer Punkt und die Hessesche Matrix der Lagrange-Funktion $L^* = L(x^*)$ positiv definit. Dann hat das duale Problem

$$\text{maximiere } \phi(\lambda)$$

eine lokale Lösung an λ^* mit dem entsprechendem Wert r^* und x^* als entsprechendem Wert in der Definition von ϕ zu λ^* .

Beweis. Offensichtlich folgt aus der Definition von ϕ , dass x^* mit λ^* korrespondiert. An λ^* gilt nach dem Lemma 1.4.9:

$$\nabla\phi(\lambda^*) = h(x^*)^T = 0$$

und wegen Lemma 1.4.11 folgt, dass die Hessesche Matrix negativ definit ist. Also erfüllt λ^* die hinreichenden Kriterien erster und zweiter Ordnung für ein nichtrestringiertes Maximum von ϕ . Der entsprechende Wert folgt aus der Definition:

$$\phi(\lambda^*) = f(x^*) + (\lambda^*)^T \underbrace{h(x^*)}_{=0} = f(x^*) = r^*. \quad \square$$

Betrachte

$$(1.24) \quad \text{minimiere } f(x) \quad \text{mit } h(x) = 0 \text{ und } g(x) \leq 0,$$

wobei $g \in C^2$, $g(x) \in \mathbb{R}^p$. Sei x^* eine lokale Lösung von (1.24), die regulär bezüglich der Nebenbedingungen ist. Dann gibt es Lagrange-Multiplikatoren λ^* und $\mu^* \geq 0$, so dass

$$\nabla f(x^*) + (\lambda^*)^T \nabla h(x^*) + (\mu^*)^T \nabla g(x^*) = 0$$

und

$$(\mu^*)^T g(x^*) = 0$$

gelten. Wir fordern lokale Konvexität: die Hessesche Matrix der Lagrange-Funktion

$$L(x^*) = F(x^*) + (\lambda^*)^T H(x^*) + (\mu^*)^T G(x^*)$$

ist positiv definit auf dem ganzen Raum. Für alle λ und $\mu \geq 0$ in der Umgebung von λ^* und μ^* definieren wir die duale Funktion

$$\phi(\lambda, \mu) = \min_x (f(x) + \lambda^T h(x) + \mu^T g(x)),$$

wobei das Minimum in einer Umgebung von x^* angenommen wird. Man kann analog zum obigen Fall zeigen, dass ϕ ein lokales Maximum hinsichtlich λ und $\mu \geq 0$ in einer Umgebung von λ^* und μ^* annimmt.

Wenn wir annehmen, dass die Funktionen f und g konvex und h affinlinear sind, so ist (1.24) ein konvexes Optimierungsproblem. Dann ist x^* nicht nur eine lokale, sondern auch eine globale Lösung. Ferner ist die Lagrange-Funktion $f(x) + \lambda^T h(x) + \mu^T g(x)$ konvex für alle λ und $\mu \geq 0$; denn für $h = c + r$ mit r linear und

c konstant gilt:

$$\begin{aligned} & f(\alpha x_1 + (1 - \alpha)x_2) + \lambda^T h(\alpha x_1 + (1 - \alpha)x_2) + \mu^T g(\alpha x_1 + (1 - \alpha)x_2) \\ & \leq \alpha f(x_1) + (1 - \alpha)f(x_2) + \lambda^T c + \alpha \lambda^T r(x_1) + (1 - \alpha)\lambda^T r(x_2) \\ & \quad + \alpha \mu^T g(x_1) + (1 - \alpha)\mu^T g(x_2) \\ & = \alpha f(x_1) + \alpha \lambda^T h(x_1) + \alpha \mu^T g(x_1) + (1 - \alpha)f(x_2) \\ & \quad + (1 - \alpha)\lambda^T h(x_2) + (1 - \alpha)\mu^T g(x_2). \end{aligned}$$

Die lokale Minimierung kann dann durch eine globale ersetzt werden. Es kann gezeigt werden, dass ϕ konkav ist. Dann ist jedes lokale Maximum ein globales. Alles zusammen bedeutet, dass die Dualitäts-Resultate alle gültig bleiben, wenn die lokalen Operationen durch globale ersetzt werden.

Es ist nicht notwendig, die Lagrange-Multiplikatoren von allen Nebenbedingungen des Problems in die Definition der dualen Funktion aufzunehmen. Wenn die lokale Konvexitätsannahme — L^* positiv definit — gilt, kann lokale Dualität im allgemeinen im Hinblick auf eine Teilmenge der funktionalen Nebenbedingungen definiert werden. Für

$$\text{minimiere } f(x) \quad \text{mit } h(x) = 0 \text{ und } g(x) \leq 0$$

können wir die duale Funktion nur im Hinblick auf die Gleichungs-Restriktionen definieren:

$$\phi(\lambda) = \min_{g(x) \leq 0} (f(x) + \lambda^T h(x)),$$

wobei das Minimum lokal bei der Lösung x^* liegt, aber durch $g(x) \leq 0$ restringiert ist. Restringierte Probleme können also dadurch gelöst werden, indem das dazugehörige Problem gelöst wird, sofern die lokale Konvexitätsannahme gilt.

Das augmentierte Lagrange-Verfahren kann als eine Kombination von Penalty-Verfahren und lokalen Dualitätsmethoden gesehen werden. Diese beiden Konzepte arbeiten hier zusammen, um Nachteile zu eliminieren, die bei jedem einzelnen Verfahren entstehen. Das augmentierte Lagrange-Verfahren für das durch Gleichungen restringierte Problem:

$$(1.25) \quad \text{minimiere } f(x) \quad \text{mit } h(x) = 0$$

benutzt die augmentierte Lagrange-Funktion, die wie folgt lautet:

$$\mathcal{L}_c(x, \lambda) = f(x) + \lambda^T h(x) + \frac{c}{2} \|h(x)\|^2$$

mit $c > 0$. Von der Sichtweise der Penalty-Verfahren ist die augmentierte Lagrange-Funktion für ein fixes λ einfach die quadratische Penalty-Funktion für das Problem:

$$(1.26) \quad \text{minimiere } f(x) + \lambda^T h(x) \quad \text{mit } h(x) = 0.$$

Das Problem ist offensichtlich äquivalent mit dem Problem (1.25), da Kombinationen mit der Nebenbedingung den Minimalpunkt von $f(x)$ nicht beeinträchtigen. Wenn der Multiplikationsvektor gleich λ^* ist, dann verschwindet der Gradient von $\mathcal{L}_c(x, \lambda^*)$ an der Lösung x^* ; denn

$$\nabla \mathcal{L}_c(x, \lambda^*) = 0$$

impliziert

$$\nabla f(x) + (\lambda^*)^T \nabla h(x) + ch(x) \nabla h(x) = 0,$$

was durch $\nabla f(x) + (\lambda^*)^T \nabla h(x) = 0$ und $h(x) = 0$ erfüllt wird. So kann die augmentierte Lagrange-Funktion genau als eine Penalty-Funktion betrachtet werden, sofern der passende Wert λ^* verwandt wird.

Ein typischer Schritt der augmentierten Lagrange-Methode beginnt mit einem Vektor λ_k . Der Punkt x_k wird dann als Minimalstelle von

$$(1.27) \quad f(x) + \lambda_k^T h(x) + \frac{c}{2} \|h(x)\|^2$$

bestimmt. Dann wird aus λ_k ein neues λ_{k+1} berechnet. Eine Standardmethode ist durch:

$$\lambda_{k+1} = \lambda_k + c h(x_k)$$

gegeben. Um dieses zu motivieren, betrachten wir das restringierte Problem (1.25) mit $\lambda = \lambda_k$. Der Lagrangesche Multiplikator des Problems ist $\lambda^* - \lambda_k$, wobei λ^* der Lagrangesche Multiplikator von (1.25) ist. Da auf der anderen Seite (1.27) die Penalty-Funktion von (1.26) ist, kann der Term $c h(x_k)$ approximativ gleich dem Lagrangeschen Multiplikator von (1.26) gesetzt werden: $c h(x_k) \simeq \lambda^* - \lambda_k$. Damit liefert $\lambda_{k+1} = \lambda_k + c h(x_k)$ eine gute Approximation für die Unbekannte λ^* .

Obwohl die Hauptiteration in augmentierten Lagrange-Methoden bezüglich λ stattfindet, muß der Penalty-Parameter während der Iterationen auch modifiziert werden. Wie in den eigentlichen Penalty-Verfahren wird c entweder festgehalten oder bis zu einem endlichen Wert erhöht oder (langsam) gegen unendlich laufen gelassen. Hier ist es nicht notwendig, dass c gegen unendlich konvergiert. Tatsächlich kann c nämlich auf einem relativ geringen Wert bleiben, so dass die schlechte Konditionierung des Penalty-Verfahrens vermieden wird.

Von der Sichtweise der Dualitätstheorie kann die augmentierte Lagrange-Funktion einfach als Lagrange-Funktion des Problems:

$$\text{minimiere } f(x) + \frac{c}{2} \|h(x)\|^2 \quad \text{und} \quad h(x) = 0$$

betrachtet werden. Dieses Problem ist äquivalent zu dem Ausgangsproblem (1.25), da eine Addition des Termes $c \|h(x)\|^2/2$ zur Zielfunktion weder den optimalen Wert noch die Optimalstelle noch die ursprüngliche Lagrange-Funktion ändert. Wohingegen die ursprüngliche Lagrange-Funktion aber nahe der Lösung nicht konvex ist und damit die Dualitätstheorie nicht anwendbar ist, führt der Term $c \|h(x)\|^2/2$ dazu, die Lagrange-Funktion konvex zu machen.

Wenn der Penalty-Parameter c hinreichend groß ist, dann hat die augmentierte Lagrange-Funktion eine lokale Minimalstelle in einer Umgebung nahe beim optimalen Punkt. Zum Beweis benötigen wir ein Hilfsresultat aus der Linearen Algebra.

Lemma 1.4.13. *Seien A und B zwei $(n \times n)$ -Matrizen. B sei positiv semidefinit und A sei positiv definit auf dem Teilraum $Bx = 0$. Dann gibt es ein c^* , so dass für alle $c \geq c^*$ die Matrix $A + cB$ positiv definit ist.*

Bemerkung 1.4.14. *Dieses Lemma wenden wir direkt auf die Hessesche Matrix der Lagrange-Funktion, ausgewertet an dem optimalen Lösungspaar x^* und λ^* , an. Wir setzen noch voraus, dass die hinreichende Bedingung zweiter Ordnung für ein restringiertes Minimum an x^* und λ^* erfüllt ist. Die Hessesche Matrix an dem optimalen Paar x^* und λ^* ergibt sich zu:*

$$\begin{aligned} L_c(x^*) &= F(x^*) + (\lambda^*)^T H(x^*) + c \nabla h(x^*)^T \nabla h(x^*) \\ &= L(x^*) + c \nabla h(x^*)^T \nabla h(x^*). \end{aligned}$$

Der erste Term $L(x^*)$ — die Hessesche Matrix der eigentlichen Lagrange-Funktion — ist positiv definit auf dem Teilraum $\nabla h(x^*)x = 0$. Ihr entspricht die Matrix A aus Lemma 1.4.13. Die Matrix $\nabla h(x^*)^T \nabla h(x^*)$ ist positiv semidefinit und entspricht der Matrix B im Lemma 1.4.13. Es gibt also ein c^* , so dass für alle $c > c^*$ die Matrix $L_c(x^*)$ positiv definit ist.

Satz 1.4.15. *Es gelten die hinreichenden Bedingungen zweiter Ordnung für ein lokales Minimum an x^* und λ^* . Dann existiert ein c^* , so dass für alle $c \geq c^*$ die augmentierte Lagrange-Funktion ein lokales Minimum an x^* hat.*

Bemerkung 1.4.16. *Unter der Ausnutzung der Stetigkeit kann das Resultat auf eine Umgebung von x^* und λ^* ausgeweitet werden: Für jedes λ in der Nähe von λ^* hat die augmentierte Lagrange-Funktion einen eindeutigen Minimalpunkt nahe bei x^* . Diese Beziehung definiert eine stetige Funktion. Wenn ein Wert λ mit $h(x(\lambda)) = 0$ gefunden werden kann, so muß dieses λ tatsächlich λ^* sein, denn $x(\lambda)$ erfüllt die notwendigen Bedingungen des Originalproblems. Deshalb kann ein geeigneter Wert für λ derart gefunden werden, indem die Gleichung $h(x(\lambda)) = 0$ gelöst wird.*

2. Grundlagen der infiniten Optimierung

Nun kommen wir zur unendlich-dimensionalen Optimierung. Als Referenz für die präsentierten Resultate und deren Beweise verweisen wir auf das Buch [2] von D. G. Luenberger.

2.1. Ableitungsbegriffe. Seien im folgenden X ein Vektorraum, Y ein normierter Raum und T eine (möglicherweise nichtlineare) Transformation, die auf einem Gebiet $D \subset X$ definiert ist und deren Bild R eine Teilmenge von Y ist.

Definition 2.1.1 (Gateaux-Differential). *Seien $x \in D$ und $h \in X$ geeignet. Wenn der Grenzwert*

$$(2.28) \quad \delta T(x; h) = \lim_{\alpha \rightarrow 0} \frac{1}{\alpha} [T(x + \alpha h) - T(x)]$$

existiert, so heißt er Gateaux-Differential von T an x mit Steigung h . Wenn (2.28) für jedes $h \in X$ existiert, dann nennen wir die Transformation T an x Gateaux-differenzierbar.

Bemerkung 2.1.2. *Wir bemerken, dass es nur Sinn macht, den Grenzwert (2.28) für $x + \alpha h \in D$ zu betrachten, also für hinreichend kleine α . Der Grenzwert (2.28) wird im gewöhnlichen Sinn der Normkonvergenz in Y verstanden. Für festes $x \in D$ bei variablem h definiert das Gateaux-Differential eine Transformation von X nach Y . Im besonderen Fall, wenn T linear ist, so erhalten wir $\delta T(x; h) = T(h)$.*

Die meisten Anwendungen dieser Definition beschränken sich auf den Fall, dass Y die reelle Zahlenachse ist, und damit reduziert sich die Transformation T zu einem reellwertigen Funktional auf X . Wenn f ein Funktional auf X ist, so lautet das Gateaux-Differential im Falle der Existenz:

$$\delta f(x; h) = \left. \frac{d}{d\alpha} f(x + \alpha h) \right|_{\alpha=0}.$$

Für festes $x \in X$ ist $\delta f(x; h)$ ein Funktional bezüglich der Variable $h \in X$.

Das Gateaux-Differential verallgemeinert das Konzept der Richtungsableitung von finiten auf infinite Räumen. Die Existenz des Gateaux-Differentials ist eine

weitaus schwächere Forderung, da die Definition keine Norm auf X erfordert. Eigenschaften des Gateaux-Differentials hängen zum Beispiel nicht einfach mit der Stetigkeit zusammen, insbesondere folgt aus der Existenz des Gateaux-Differential an einer Stelle $x \in X$ nicht sofort die Stetigkeit in x . Wenn X normiert ist, so wird eine zufriedenstellendere Definition durch den Begriff des Fréchet-Differentials gegeben.

Definition 2.1.3 (Fréchet-Differential). *Sei T eine Transformation, die auf einem offenen Gebiet $D \subset X$ definiert ist und deren Bild R eine Teilmenge von Y ist. Wenn für jedes $x \in D$ und jedes $h \in X$ das Gateaux-Differential $\delta T(x; h) \in Y$ existiert, das linear und stetig bezüglich h ist, so dass*

$$\lim_{\|h\|_X \rightarrow 0} \frac{\|T(x+h) - T(x) - \delta T(x; h)\|_Y}{\|h\|_X} = 0$$

gilt, dann heißt T Fréchet-differenzierbar an x , und $\delta T(x; h)$ heißt das Fréchet-Differential von T an x mit der Steigung h .

Bemerkung 2.1.4. *Wir benutzen für das Fréchet- und Gateaux-Differential das gleiche Symbol, da im allgemeinen aus dem Textzusammenhang klar wird, welches von beiden jeweils gemeint ist.*

Satz 2.1.5 (Eindeutigkeit des Fréchet-Differential). *Wenn die Transformation T an einer Stelle ein Fréchet-Differential besitzt, so ist dieses eindeutig bestimmt.*

Beweis. Angenommen, sowohl $\delta T(x; h)$ als auch $\delta' T(x; h)$ erfüllen die Voraussetzungen der letzten Definition. Dann folgt:

$$\begin{aligned} \|\delta T(x; h) - \delta' T(x; h)\|_Y &\leq \|T(x+h) - T(x) - \delta T(x; h)\|_Y + \\ &\quad \|T(x+h) - T(x) - \delta' T(x; h)\|_Y, \end{aligned}$$

das bedeutet: $\|\delta T(x; h) - \delta' T(x; h)\|_Y = o(\|h\|_X)$. Da $\delta T(x; h) - \delta' T(x; h)$ nach Definition des Fréchet-Differentials beschränkt und linear ist, muß $\delta T(x; h) - \delta' T(x; h) = 0$ gelten. $\square$

Satz 2.1.6. *Wenn das Fréchet-Differential von T in x existiert, dann ist T in x auch Gateaux-differenzierbar und beide Differentiale haben den gleichen Wert.*

Beweis. Wir bezeichnen mit $\delta T(x; h)$ das Fréchet-Differential. Dann gilt nach Definition für jedes h :

$$\frac{1}{\alpha} \|T(x + \alpha h) - T(x) - \delta T(x; \alpha h)\|_Y \xrightarrow{\alpha \rightarrow 0} 0.$$

Mit der Linearität von $\delta T(x; \alpha h)$ bezüglich α bekommen wir

$$\lim_{\alpha \rightarrow 0} \frac{T(x + \alpha h) - T(x)}{\alpha} = \delta T(x; h). \quad \square$$

Satz 2.1.7. *Wenn die Transformation T , die auf einer offenen Menge $D \subset X$ definiert sei, in x ein Fréchet-Differential besitzt, dann ist T stetig in x .*

Beweis. Sei $\epsilon > 0$. Dann gibt es eine Kugel um x , so dass $x + h$ in dieser Kugel enthalten ist und

$$\|T(x+h) - T(x) - \delta T(x; h)\|_Y < \epsilon \|h\|_X$$

gilt. Daraus folgt:

$$\|T(x+h) - T(x)\|_Y < \epsilon \|h\|_X + \|\delta T(x; h)\|_Y < M \|h\|_X,$$

woraus wir die Stetigkeit in x erhalten. $\square$

Wir nehmen an, dass die Transformation $T : D \rightarrow Y$, $D \subset X$ offen, in ganz D Fréchet-differenzierbar ist. An einem festen Punkt $x \in D$ ist dann das Fréchet-Differential von der Form $\delta T(x; h) = A_x h$, wobei $A_x : X \rightarrow Y$ ein beschränkter linearer Operator ist. Wenn x aus der Menge D gewählt wird, so wird durch die Zuordnung $x \rightarrow A_x$ eine Transformation von D in den normierten Vektorraum $B(X, Y)$ definiert. Diese Abbildung nennen wir Fréchet-Ableitung T' von T . Nach Definition erhalten wir: $\delta T(x; h) = T'(x) h$.

Wenn die Abbildung $x \rightarrow A_x$ stetig am Punkt x_0 ist, so nennen wir die Fréchet-Ableitung von T stetig an x_0 . Das sollte nicht damit durcheinander gebracht werden, dass die Abbildung $T'(x_0)$ eine stetige Abbildung von X nach Y ist, was bereits aus der Definition der Fréchet-Ableitung folgt.

In dem besonderen Fall, wo die ursprüngliche Transformation ein Funktional f auf X ist, so bekommen wir $\delta f(x; h) = f'(x) h$ mit $f'(x) \in X^*$ für alle x . Das Element $f'(x)$ heißt Gradient von f in x , meist auch mit $\nabla f(x)$ bezeichnet. Wir schreiben oft $\langle h, f'(x) \rangle$ für $\delta f(x; h)$, da $f'(x) \in X^*$ gilt, werden aber die Notation $f'(x) h$ vorziehen, die mit der von Differentialen von geeigneten Transformationen konsistent ist.

Aus der Definition folgt sofort, wenn T_1 und T_2 in $x \in D$ Fréchet-differenzierbar sind, dass dann $\alpha_1 T_1 + \alpha_2 T_2$ in x Fréchet-differenzierbar ist und $(\alpha_1 T_1 + \alpha_2 T_2)'(x) = \alpha_1 T_1'(x) + \alpha_2 T_2'(x)$ gilt.

Satz 2.1.8 (Kettenregel). *Seien $S : D \rightarrow E$. $D \subset X$ und $E \subset Y$ offen, und $P : E \rightarrow Z$, Z normierter Raum, zwei Transformationen. Es gelte $T = P S$, wobei S in $x \in D$ und P in $y = S(x) \in E$ Fréchet-differenzierbar sind. Dann ist T in x Fréchet-differenzierbar mit $T'(x) = P'(y) S'(x)$.*

Beweis. Für $h \in X$, $x + h \in D$ erhalten wir

$$T(x + h) - T(x) = P[S(x + h)] - P[S(x)] = P(y + g) - P(y),$$

wobei $g = S(x + h) - S(x) \in Y$ gilt. Also bekommen wir

$$\|T(x + h) - T(x) - P'(y) g\|_Z = o(\|g\|_Y).$$

Wegen $\|g - S'(x) h\|_Y = o(\|h\|_X)$ gilt

$$\begin{aligned} & \|T(x + h) - T(x) - P'(y) S'(x) h\|_Z \\ &= \|T(x + h) - T(x) - P'(y) g + P'(y) (g - S'(x) h)\|_Z \\ &\leq \|T(x + h) - T(x) - P'(y) g\|_Z + \|P'(y)\|_{L(Y, Z)} \|g - S'(x) h\|_Y \\ &= o(\|g\|_Y) + o(\|h\|_X). \end{aligned}$$

Hier bezeichnet $L(Y, Z)$ den Banachraum aller beschränkten linearen Operatoren von Y nach Z . Wegen Satz 2.1.7 ist S stetig in x , und wir bekommen $\|g\|_Y = O(\|h\|_X)$ und damit:

$$T'(x) h = P'(y) S'(x) h. \quad \square$$

Lemma 2.1.9. *Seien $T : D \rightarrow Y$ Fréchet-differenzierbar auf einem offenen Gebiet $D \subset X$ und $x \in D$ ein Punkt, so dass $x + \alpha h \in D$ für alle α , $0 \leq \alpha \leq 1$, gilt. Dann folgt die Abschätzung:*

$$\|T(x + h) - T(x)\|_Y \leq \|h\|_X \sup_{0 < \alpha < 1} \|T'(x + \alpha h)\|_Y.$$

Wenn $T : D \rightarrow Y$ Fréchet-differenzierbar auf einem offenen Gebiet $D \subset X$ ist, dann bildet die Ableitung T' D in die Menge $L(X, Y)$ ab und ist eventuell selbst Fréchet-differenzierbar auf einer Teilmenge $D_1 \subset D$. In diesem Fall wird die Fréchet-Ableitung von T' die zweite Fréchet-Ableitung von T genannt und mit T'' bezeichnet.

Lemma 2.1.10. *Seien T zweimal Fréchet-differenzierbar auf einem offenen Gebiet D und $x \in D$ ein Punkt, so dass $x + \alpha h \in D$ für alle α , $0 \leq \alpha \leq 1$, gilt. Dann ergibt sich:*

$$\|T(x+h) - T(x) - T'(x)h\|_Y \leq \frac{1}{2} \|h\|_X^2 \sup_{0 < \alpha < 1} \|T''(x + \alpha h)\|_{L(X, L(X, Y))}.$$

Definition 2.1.11 (Relative/Lokale Minimalstelle). *Sei f ein reellwertiges Funktional auf einer Teilmenge Ω des normierten Raumes X . Ein Punkt $x_0 \in \Omega$ heißt eine relative (oder lokale) Minimalstelle von f auf Ω , wenn es eine offene Kugel K mit $x_0 \in K$ gibt, so dass $f(x_0) \leq f(x)$ für alle $x \in \Omega \cap K$. Der Punkt x_0 wird eine strikte relative Minimalstelle von f auf Ω genannt, wenn $f(x_0) < f(x)$ für alle $x \neq x_0$, $x \in \Omega \cap K$, erfüllt ist.*

Satz 2.1.12 (Notwendige Bedingung 1. Ordnung). *Das Funktional $f : X \rightarrow \mathbb{R}$ besitze eine Gateaux-Ableitung auf einem Vektorraum X . Eine notwendige Bedingung für f , an $x_0 \in X$ ein Extremum zu haben, ist $\delta f(x_0; h) = 0$ für alle $h \in X$.*

Beweis. Für jedes $h \in X$ nimmt die Funktion $f(x_0 + \alpha h)$ der reellen Variablen α an $\alpha = 0$ ihr Extremum an. Daher folgt:

$$\left. \frac{d}{d\alpha} f(x_0 + \alpha h) \right|_{\alpha=0} = 0. \quad \square$$

Ein Punkt, an dem $\delta f(x_0; h) = 0$ für alle h gilt, heißt stationärer Punkt. Damit ist die Aussage des vorherigen Satzes, dass ein Extremum von f nur an einem stationären Punkt vorliegen kann. Ein analoges Resultat erhalten wir für ein lokales Extremum von einem Funktional f , welches auf einer offenen Teilmenge des normierten Raumes definiert ist, und der Beweis entspricht dem Beweis oben.

Satz 2.1.13. *Sei f ein reellwertiges Funktional, definiert auf einem Vektorraum X . An x_0 werde f auf der konvexen Menge $\Omega \subset X$ minimiert und an x_0 sei f Gateaux-differenzierbar. Dann gilt*

$$\delta f(x_0; x - x_0) \geq 0$$

für alle $x \in \Omega$.

Beweis. Da Ω konvex ist, bekommen wir $x_0 + \alpha(x - x_0) \in \Omega$ für $0 \leq \alpha \leq 1$ und

$$\left. \frac{d}{d\alpha} f(x_0 + \alpha(x - x_0)) \right|_{\alpha=0} \geq 0$$

für ein Minimum an x_0 . □

Wir betrachten das Problem, ein Funktional f unter n nichtlinearen, implizit gegebenen Bedingungen

$$(2.29) \quad g_i(x) = 0 \quad \text{für } i = 1, \dots, n$$

zu optimieren. Diese n Gleichungen definieren eine Menge Ω in dem Raum X , in der der optimale Vektor x_0 liegen muß. Die Funktionale $f, g_1, \dots, g_n$ seien stetig und Fréchet-differenzierbar auf dem normierten Raum X .

Wenn x_0 optimal ist, dann hat das Funktional f ein Extremum an x_0 bezüglich kleiner Variationen entlang Ω . Unter hinreichend glatten Voraussetzungen sollte f ein Extremum an x_0 bezüglich kleiner Variationen entlang von T , der Tangentialebene an Ω in x_0 , haben. Der Nutzen dieser Beobachtung ist, dass die exakte Form der Oberfläche Ω nahe bei x_0 durch die einfache Beschreibung der Tangentialfläche bei der Formulierung der notwendigen Bedingungen ersetzt werden kann.

Definition 2.1.14 (Regulärer Punkt). *Ein Punkt, der die Restriktionen (2.29) erfüllt, wird ein regulärer Punkt bezüglich dieser Nebenbedingungen genannt, wenn die n linearen Funktionale $g'_1(x_0), \dots, g'_n(x_0)$ linear unabhängig sind.*

Satz 2.1.15 (Notwendige Bedingung 1. Ordnung). *Wenn x_0 eine Extremstelle des Funktionals f unter den Nebenbedingungen $g_i(x) = 0, i = 1, \dots, n$, und ein regulärer Punkt bezüglich dieser Nebenbedingungen sind, dann gilt*

$$\delta f(x_0; h) = 0$$

für alle h mit $\delta g_i(x_0; h) = 0, i = 1, \dots, n$.

2.2. Konvexe Theorie der restringierten Optimierung. Indem wir einen Kegel einführen, der die positiven Vektoren in einem gegebenen Raum definiert, ist es möglich, Ungleichungs-Restriktionen in abstrakten Vektorräumen zu betrachten.

Definition 2.2.1 (Positiver/Negativer Kegel). *Sei P ein konvexer Kegel in einem Vektorraum X . Für $x, y \in X$ schreiben wir $x \geq y$ bezüglich P , wenn $x - y \in P$ gilt. Der Kegel P , der durch diese Relation definiert wird, heißt positiver Kegel in X . Der Kegel $N = -P$ wird entsprechend negativer Kegel in X genannt, und wir schreiben $y \leq x$ für $y - x \in N$.*

Zum Beispiel definiert im $\mathbb{R}^n$ der konvexe Kegel

$$(2.30) \quad P = \{x \in \mathbb{R}^n : x = (\xi_1, \dots, \xi_n) \text{ mit } \xi_i \geq 0 \text{ für alle } i\}$$

den üblichen positiven Orthanten. In einem Raum von Funktionen, die auf der reellen Achse definiert sind, sagen wir auf dem Intervall $[t_1, t_2]$, ist es natürlich, den positiven Kegel so zu definieren, dass er aus den Funktionen besteht, die auf $[t_1, t_2]$ nichtnegativ sind.

Wir können leicht verifizieren, dass aus $x \geq y$ und $y \geq z$ folgt: $x \geq z$. Da $\theta \in P$ gilt, erhalten wir $x \geq x$ für alle $x \in X$. Hier haben wir mit θ das Nullelement in dem Vektorraum X bezeichnet.

Im Fall des normierten Raumes schreiben wir $x > \theta$, wenn x ein innerer Punkt des positiven Kegels P ist. Für viele Anwendungen ist es wesentlich, dass P einen inneren Punkt besitzt, so dass der Satz von der tragenden Hyperebene verwendet werden kann.

Gegeben sei ein normierter Raum X mit einem positiven konvexen Kegel $P \subset X$. Es ist üblich, einen entsprechenden positiven Kegel $P^\ominus$ in dem Dualraum X^* durch

$$(2.31) \quad P^\ominus = \{x^* \in X^* : \langle x, x^* \rangle_{X, X^*} \geq 0 \text{ für alle } x \in P\}$$

zu definieren.

Beispiel 2.2.2. Wenn P wie in (2.30) gewählt wird, dann besteht $P^\oplus$ aus allen linearen Funktionalen, die durch Elemente aus dem $\mathbb{R}^n$ mit nichtnegativen Komponenten repräsentiert werden.

Wir werden in diesem Skript den Begriff der Variation einer Funktion f keinen Gebrauch machen. Da die Variation aber in der Optimierung an vielen Stellen verwendet wird, wollen wir hier die Definition und einige Eigenschaften bereitstellen.

Definition 2.2.3 (Variation). *Es sei Z eine Zerlegung des Intervalls $[a, b]$ mit den Teilpunkten $a = x_0 < x_1 < \dots < x_n = b$. Ist f eine Funktion mit $[a, b] \subseteq D(f)$, so heißt die Zahl*

$$V(f, Z) = \sum_{i=1}^n |f(x_i) - f(x_{i-1})|$$

die Variation (Schwankung) von f bezüglich Z . Existiert die obere Grenze

$$V(f, [a, b]) = \sup_Z V(f, Z)$$

der Menge aller Variationen bezüglich aller Zerlegungen Z von $[a, b]$, dann wird diese totale Variation von f über $[a, b]$ genannt. f heißt in diesem Fall von beschränkter Variation.

Bemerkung 2.2.4. Funktionen von beschränkter Variation brauchen nicht stetig zu sein. Es gibt auch stetige Funktionen, die nicht von beschränkter Variation sind.

- Satz 2.2.5.**
- a) Jede in $[a, b]$ monotone Funktion ist von beschränkter Variation.
 - b) Jede in $[a, b]$ differenzierbare Funktion, deren Ableitung in $[a, b]$ beschränkt ist, ist von beschränkter Variation.
 - c) Eine in $[a, b]$ definierte Funktion ist genau dann von beschränkter Variation, wenn sie als Differenz zweier monoton wachsender Funktionen darstellbar ist.
 - d) Jede Funktion von beschränkter Variation ist im Riemannschen Sinne integrierbar.

Beispiel 2.2.6. a) Ist f in $[a, b]$ monoton wachsend, dann gilt: $V(f, [a, b]) = f(b) - f(a)$.

b) Die Funktion

$$f(x) = \begin{cases} x \cos\left(\frac{\pi}{2x}\right) & \text{für } x \in (0, 1], \\ 0 & \text{für } x = 0 \end{cases}$$

ist in $[0, 1]$ nicht von beschränkter Variation, denn für die spezielle Zerlegung Z :

$$0 < \frac{1}{2n} < \frac{1}{2n-1} < \dots < \frac{1}{3} < \frac{1}{2} < 1$$

erhalten wir $V(f, Z) = \sum_{k=1}^n 1/(2k)$, woraus wegen der Divergenz der harmonischen Reihe die Behauptung folgt.

Beispiel 2.2.7. Wenn in dem Raum $C[t_1, t_2]$ der Kegel P als die Menge aller nichtnegativen stetigen Funktionen auf $[t_1, t_2]$ gewählt wird, dann besteht $P^\oplus$ aus allen linearen Funktionalen auf $C[t_1, t_2]$, die durch Funktionen von beschränkter Variation und wachsender Monotonie auf $[t_1, t_2]$ repräsentiert werden.

Wir können zeigen, dass $P^\oplus$ abgeschlossen ist, gerade wenn P es nicht ist.

Lemma 2.2.8. *Sei der positive Kegel P in den normierten Raum X abgeschlossen. Wenn $x \in X$*

$$\langle x, x^* \rangle_{X, X^*} \geq 0 \quad \text{für alle } x^* \geq \theta$$

erfüllt, dann folgt: $x \geq \theta$. Hier bezeichnet $\langle \cdot, \cdot \rangle_{X, X^}$ die duale Paarung eines Elementes in X mit einem Funktional aus X^* bezeichnet.*

Da wir die Ungleichung zwischen Vektoren eingeführt haben, ist es möglich, auch eine allgemeine Definition von konvexen Abbildungen zu geben.

Definition 2.2.9 (Konvexe Abbildung im Vektorraum). *Seien X und Z Vektorräume, wobei Z einen positiven Kegel P enthält. Eine Abbildung $G : X \rightarrow Z$ heißt konvex, wenn das Gebiet $\Omega \subset X$ von G eine konvexe Menge ist und*

$$G(\alpha x_1 + (1 - \alpha)x_2) \leq \alpha G(x_1) + (1 - \alpha)G(x_2)$$

für alle $x_1, x_2 \in \Omega$ und alle α mit $0 < \alpha < 1$ gilt.

Beispiel 2.2.10. *Beispiele von konvexen Funktionen in einer Dimension sind durch $f(x) = x^2$ und $f(x) = e^x$ für $x > 0$. Das Funktional*

$$f(x) = \int_0^1 (x^2(t) + |x(t)|) dt$$

für $x \in L^2[0, 1]$ ist auch konvex und stetig.

Lemma 2.2.11. *Seien f ein konvexes Funktional auf einer konvexen Teilmenge C des normierten Raumes und $\mu = \inf_{x \in C} f(x)$. Dann gelten:*

- a) *Die Teilmenge $\Omega \subset C$, auf der $f(x) = \mu$ gilt, ist konvex.*
- b) *Wenn x_0 ist eine lokale Minimalstelle von f ist, so folgt $f(x_0) = \mu$ und somit ist x_0 eine globale Minimalstelle.*

Wir bemerken, dass Konvexität keine der Abbildung innewohnende Eigenschaft, sondern abhängig von dem speziellen positiven Kegel im Bildraum ist.

Lemma 2.2.12. *Sei G eine konvexe Abbildung wie in der letzten Definition. Dann ist für jedes $z \in Z$ die Menge*

$$\{x \in \Omega : G(x) \leq z\}$$

konvex.

Wir betrachten das Problem:

$$(2.32) \quad \text{minimiere } f(x) \quad \text{mit } G(x) \leq \theta \text{ und } x \in \Omega,$$

wobei Ω eine konvexe Teilmenge des Vektorraumes X , f ein reellwertiges Funktional auf Ω , G eine konvexe Abbildung von Ω in den Vektorraum Z und $P \subset Z$ ein positiver Kegel sind. Das Problem (2.32) wird ein allgemeines konvexes Programmierungsproblem genannt.

Wir analysieren das Problem (2.32) und entwickeln den Satz über die Lagrange-Multiplikatoren im wesentlichen durch die Einbettung in die Familie von Problemen

$$\text{minimiere } f(x) \quad \text{mit } G(x) \leq z \text{ und } x \in \Omega,$$

wobei z ein passender Vektor in Z ist. Die Lösungen dieses Problems hängen von z ab, und diese Abhängigkeit wollen wir untersuchen.

Dazu definieren wir die Menge $\Gamma \subset Z$ durch

$$\Gamma = \{z \in Z : \text{es existiert ein } x \in \Omega \text{ mit } G(x) \leq z\}.$$

Die Menge Γ ist konvex: $z_1, z_2 \in \Gamma$ implizieren die Existenz von $x_1, x_2 \in \Omega$ mit $G(x_1) \leq z_1, G(x_2) \leq z_2$; also erhalten wir für jedes α mit $0 < \alpha < 1$ die Ungleichung $G(\alpha x_1 + (1 - \alpha)x_2) \leq \alpha z_1 + (1 - \alpha)z_2$ und daher $\alpha x_1 + (1 - \alpha)x_2 \in \Gamma$.

Auf der Menge Γ definieren wir das primale Funktional, das nicht endlich sein muß, durch

$$\omega(z) = \inf\{f(x) : G(x) \leq z \text{ und } x \in \Omega\}.$$

Das Originalproblem (2.32) kann als Bestimmung des Wertes $\omega(\theta)$ betrachtet werden. Wir untersuchen daher das Funktional ω .

Lemma 2.2.13. *Das Funktional ω ist konvex.*

Beweis. Wir schätzen wie folgt ab:

$$\begin{aligned} \omega(\alpha z_1 + (1 - \alpha)z_2) &= \inf\{f(x) : x \in \Omega, G(x) \leq \alpha z_1 + (1 - \alpha)z_2\} \\ &\leq \inf\{f(x) : x = \alpha x_1 + (1 - \alpha)x_2, G(x_1) \leq z_1, G(x_2) \leq z_2\} \\ &\leq \alpha \inf\{f(x) : x_1 \in \Omega, G(x_1) \leq z_1\} + (1 - \alpha) \inf\{f(x) : x_2 \in \Omega, G(x_2) \leq z_2\} \\ &\leq \alpha \omega(z_1) + (1 - \alpha)\omega(z_2). \quad \square \end{aligned}$$

Lemma 2.2.14. *Das Funktional ω ist monoton fallend.*

Konzeptionell folgt der Satz über die Lagrange-Multiplikatoren aus der einfachen Beobachtung, dass aufgrund der Konvexität von ω eine Tangentialebene für ω an $z = \theta$ existiert, die überall im Definitionsbereich unterhalb von ω liegt. Wenn man seinen Kopf neigt, so dass die Tangentialebene der neue Horizont wird, dann scheint es, dass ω an θ minimiert wird. In einer anderen Sprechweise bedeutet dieses, indem wir ein passendes lineares Funktional $\langle z, z_0^* \rangle_{Z, Z^*}$ zu $\omega(z)$ dazuzugieren, dass die resultierende Summe an $z = \theta$ minimiert wird. Das Funktional z_0^* ist der Lagrange-Multiplikator für das Problem. Die Tangentialebene im obigen Bild entspricht dem Element $(1, z_0^*) \in \mathbb{R} \times Z^*$.

Satz 2.2.15 (Lagrange-Multiplikatoren). *Seien X ein Vektorraum, Z ein linearer Raum, Ω eine konvexe Teilmenge von X und P ein positiver Kegel in Z , der einen inneren Punkt besitzt. Seien f ein reellwertiges Funktional auf Ω und $G : \Omega \rightarrow Z$ eine konvexe Abbildung. Es gebe einen Punkt $x_1 \in \Omega$ für den $G(x_1) < \theta$ gilt, das bedeutet: $G(x_1)$ ist ein innerer Punkt von $N = -P$. Sei*

$$(2.33) \quad \mu_0 = \inf_{\substack{x \in \Omega \\ G(x) \leq \theta}} f(x),$$

und es gelte $\mu_0 < \infty$. Dann gibt es ein Element $z_0^* \geq \theta$ in Z^* , so dass gilt:

$$(2.34) \quad \mu_0 = \inf_{x \in \Omega} \{f(x) + \langle G(x), z_0^* \rangle_{Z, Z^*}\}.$$

Wenn weiter das Infimum in (2.33) an $x_0 \in \Omega, G(x_0) \leq \theta$, angenommen wird, dann wird es auch an x_0 in (2.34) angenommen, und es folgt:

$$(2.35) \quad \langle G(x_0), z_0^* \rangle_{Z, Z^*} = 0.$$

Bemerkung 2.2.16. Zu zwei Voraussetzungen in dem Satz 2.2.15 wollen wir etwas sagen: die Annahme, dass der positive Kegel P einen inneren Punkt besitzt und $G(x_1) < \theta$ für ein $x_1 \in \Omega$ erfüllt ist, sichert die Existenz einer nichtvertikalen tragenden Hyperebene. Die Bedingung, der positive Kegel P besitze einen inneren Punkt, kann nicht weggelassen. In vielen Anwendungen wird dadurch die Wahl des Raumes vorgegeben, in dem das Problem formuliert wird. Die Voraussetzung $G(x_1) < \theta$ wird Regularitätsbedingung genannt und ist typisch für Sätze über Lagrange-Multiplikatoren. Sie sichert, dass die tragende Hyperebene nicht vertikal steht.

Wir haben nur konvexe Nebenbedingungen der Form $G(x) \leq \theta$ betrachtet. Eine Gleichungsrestriktion $H(x) = \theta$ mit $H(x) = Ax - b$, wobei A linear ist, ist äquivalent zu den zwei konvexen Ungleichungen $H(x) \leq \theta$ und $-H(x) \leq \theta$, kann also in die Form $G(x) \leq \theta$ gebracht werden.

Korollar 2.2.17. Es gelten die gleichen Voraussetzungen wie in dem Satz 2.2.15, und an x_0 werde das restringierte Minimum angenommen. Dann existiert ein $z_0^* \geq \theta$, so dass die Lagrange-Funktion

$$\mathcal{L}(x, z^*) = f(x) + \langle G(x), z^* \rangle_{Z, Z^*}$$

an (x_0, z_0^*) einen Sattelpunkt besitzt, das heißt:

$$\mathcal{L}(x_0, z^*) \leq \mathcal{L}(x_0, z_0^*) \leq \mathcal{L}(x, z_0^*)$$

für alle $x \in \Omega$ und $z^* \geq \theta$.

Beweis. Sei z_0^* wie im Satz 2.2.15 definiert. Aus (2.34) erhalten wir: $\mathcal{L}(x_0, z_0^*) \leq \mathcal{L}(x, z_0^*)$. Mit der Gleichung (2.35) folgern wir

$$\begin{aligned} \mathcal{L}(x_0, z^*) - \mathcal{L}(x_0, z_0^*) &= \langle G(x_0), z^* \rangle_{Z, Z^*} - \langle G(x_0), z_0^* \rangle_{Z, Z^*} \\ &= \langle G(x_0), z^* \rangle_{Z, Z^*} \leq 0, \end{aligned} \quad \square$$

Die Bedingungen der Konvexität und der Existenz von inneren Punkten kann nicht weggelassen werden, wenn wir die Existenz einer tragenden Hyperebene in dem Raum $\mathbb{R} \times Z$ garantieren wollen. Wenn aus irgendeinem Grund die geeignete Hyperebene ohne diese Bedingungen existiert, kann die Technik der Lagrange-Multiplikatoren zur Lokalisation des Optimums immer noch angewandt werden. Wenn eine geeignete Hyperfläche existiert, dann ist geradezu klar, dass $f(x) + \langle G(x), z_0^* \rangle_{Z, Z^*}$ an x_0 sein Minimum annimmt.

Satz 2.2.18 (Hinreichende Bedingung 1. Ordnung). Seien $f : \Omega \rightarrow \mathbb{R}$ ein Funktional auf einer Teilmenge Ω eines linearen Raumes X und G eine Abbildung von Ω in den normierten Raum Z , der einen nicht-leeren positiven Kegel P enthält. Es gebe Elemente $z_0^* \in Z^*$, $z_0^* \geq \theta$, und $x_0 \in \Omega$, so dass

$$f(x_0) + \langle G(x_0), z_0^* \rangle_{Z, Z^*} \leq f(x) + \langle G(x), z_0^* \rangle_{Z, Z^*}$$

für alle $x \in \Omega$ erfüllt ist. Dann löst x_0 das Problem:

$$\text{minimiere } f(x) \quad \text{mit } G(x) \leq G(x_0) \text{ und } x \in \Omega.$$

Beweis. Angenommen, es gebe ein $x_1 \in \Omega$ mit $f(x_1) < f(x_0)$ und $G(x_1) \leq G(x_0)$. Für $z_0^* \geq \theta$ ergibt sich dann

$$\langle G(x_1), z_0^* \rangle_{Z, Z^*} < \langle G(x_0), z_0^* \rangle_{Z, Z^*}$$

und daher

$$f(x_1) + \langle G(x_1), z_0^* \rangle_{Z, Z^*} < f(x_0) + \langle G(x_0), z_0^* \rangle_{Z, Z^*},$$

was der Voraussetzung des Satzes widerspricht. $\square$

Satz 2.2.19. *Seien X, Z, Ω, P, f, G wie im vorigen Satz und zusätzlich P abgeschlossen. Es gebe ein $z_0^* \in Z^*$, $z_0^* \geq \theta$, und ein $x_0 \in \Omega$, so dass die Lagrange-Funktion $\mathcal{L}(x, z^*) = f(x) + \langle G(x), z^* \rangle_{Z, Z^*}$ an x_0, z_0^* einen Sattelpunkt besitzt, das bedeutet:*

$$\mathcal{L}(x_0, z^*) \leq \mathcal{L}(x_0, z_0^*) \leq \mathcal{L}(x, z_0^*)$$

für alle $x \in \Omega, z^* \geq \theta$. Dann löst x_0 das Problem:

$$\text{minimiere } f(x) \quad \text{mit } G(x) \leq \theta \text{ und } x \in \Omega.$$

Die Sattelpunkteigenschaft bietet eine angenehme kompakte Beschreibung der wesentlichen Elemente von den Resultaten von Lagrange-Multiplikatoren für die konvexe Programmierung. Wenn f und G konvex sind, der positive Kegel $P \subset Z$ abgeschlossen ist und ein nicht-leeres Inneres besitzt und die Regularitätsbedingung erfüllt ist, dann ist die Sattelpunkteigenschaft sowohl notwendig als auch hinreichend für die Optimalität von x_0 .

2.3. Sensitivität. Wir wenden uns nun der Untersuchung der Sensitivität des Optimierungsproblems zu. Die Sätze über Lagrange-Multiplikatoren der vorhergehenden Abschnitte nutzen nicht alle geometrischen Eigenschaften aus, die sich aus der Darstellung des Problems im $\mathbb{R} \times Z$ ergeben. Es fehlen die Sensitivität und die Dualität, die beide wichtig für die Theorie und Anwendung sind.

Für alle z_0 ist die Hyperebene, die durch den Lagrange-Multiplikator des Problems

$$\text{minimiere } f(x) \quad \text{mit } G(x) \leq z_0 \text{ und } x \in \Omega$$

gegeben ist, eine tragende Hyperebene an $\omega(z_0)$, und diese Hyperebene dient als untere Schranke für ω .

Satz 2.3.1 (Sensitivitätssatz). *Seien f und G konvex. Es existieren die Lösungen x_0 und x_1 der Probleme*

$$\text{minimiere } f(x) \quad \text{mit } G(x) \leq z_0 \text{ beziehungsweise } G(x) \leq z_1 \text{ und } x \in \Omega.$$

z_0^* und z_1^* seien die dazugehörigen Lagrange-Multiplikatoren. Dann folgt:

$$\langle z_1 - z_0, z_1^* \rangle_{Z, Z^*} \leq f(x_0) - f(x_1) \leq \langle z_1 - z_0, z_0^* \rangle_{Z, Z^*}.$$

Beweis. Mit dem Lagrange-Multiplikator z_0^* erhalten wir:

$$f(x_0) + \langle G(x_0) - z_0, z_0^* \rangle_{Z, Z^*} \leq f(x) + \langle G(x) - z_0, z_0^* \rangle_{Z, Z^*}$$

für alle $x \in \Omega$. Insbesondere für $x = x_1$ bekommen wir mit $\langle G(x_0) - z_0, z_0^* \rangle_{Z, Z^*} = 0$:

$$f(x_0) - f(x_1) \leq \langle G(x_0) - z_0, z_0^* \rangle_{Z, Z^*} \leq \langle z_1 - z_0, z_0^* \rangle_{Z, Z^*}.$$

Eine ähnliche Argumentation, angewandt auf x_1 und z_1^* , ergibt die andere Ungleichung. $\square$

Eine zu dem Satz 2.3.1 äquivalente Aussage ist

$$\omega(z) - \omega(z_0) \geq \langle z_0 - z, z_0^* \rangle_{Z, Z^*}.$$

Wenn das Funktional ω an z_0 Fréchet-differenzierbar ist, folgt:

$$\omega'(z_0) = -z_0^*.$$

Daher ist der Lagrange-Multiplikator z_0^* das Negative der Sensitivität erster Ordnung von dem Optimum der Zielfunktion hinsichtlich des Restriktionstermes z_0 .

2.4. Dualität. Wir betrachten wieder das konvexe Programmierungsproblem:

$$\text{minimiere } f(x) \quad \text{mit } G(x) \leq \theta \text{ und } x \in \Omega,$$

wobei f , G und Ω konvex sind. Das allgemeine Dualitätsprinzip für dieses Problem basiert auf einfachen geometrischen Eigenschaften des in $\mathbb{R} \times Z$ betrachteten Problems. Wir das primale Funktional auf der Menge Γ :

$$(2.36) \quad \omega(z) = \inf\{f(x) : G(x) \leq z \text{ und } x \in \Omega\}$$

und setzen $\mu_0 = \omega(\theta)$. Das Dualitätsprinzip basiert auf der Beobachtung, dass μ_0 gleich dem maximalen Schnittpunkt der vertikalen Achse mit allen Hyperebenen, die unterhalb von ω liegen. Der maximale Schnittpunkt wird natürlich durch die Hyperebene bestimmt, die durch den Lagrange-Multiplikator des Problems gegeben ist. Um das obige Dualitätsprinzip analytisch auszudrücken, führen wir das **duale Funktional** φ zu (2.36) ein, welches auf dem positiven Kegel in Z^* definiert ist:

$$(2.37) \quad \varphi(z^*) = \inf_{x \in \Omega} \{f(x) + \langle G(x), z^* \rangle_{Z, Z^*}\}.$$

Im allgemeinen ist φ nicht endlich auf dem positiven Kegel in Z^* , aber die Menge, auf der φ endlich ist, ist konvex.

Lemma 2.4.1. *Das duale Funktional ist konkav und kann durch*

$$(2.38) \quad \varphi(z^*) = \inf_{z \in \Gamma} \{\omega(z) + \langle z, z^* \rangle_{Z, Z^*}\}$$

ausgedrückt werden.

Beweis. Die Konkavität von φ ist einfach zu zeigen. Für alle $z^* \geq \theta$ und alle $z \in \Gamma$ gilt:

$$\begin{aligned} \varphi(z^*) &= \inf_{x \in \Omega} \{f(x) + \langle G(x), z^* \rangle_{Z, Z^*}\} \\ &\leq \inf \{f(x) + \langle z, z^* \rangle_{Z, Z^*} : G(x) \leq z \text{ und } x \in \Omega\} \\ &= \omega(z) + \langle z, z^* \rangle_{Z, Z^*}. \end{aligned}$$

Auf der anderen Seite erhalten wir für jedes $x_1 \in \Omega$ mit $z_1 = G(x_1)$:

$$\begin{aligned} f(x_1) + \langle G(x_1), z^* \rangle_{Z, Z^*} &\geq \inf \{f(x) + \langle z_1, z^* \rangle_{Z, Z^*} : G(x) \leq z_1 \text{ und } x \in \Omega\} \\ &= \omega(z_1) + \langle z_1, z^* \rangle_{Z, Z^*} \end{aligned}$$

und damit:

$$\varphi(z^*) \geq \inf_{z \in \Gamma} \{\omega(z) + \langle z, z^* \rangle_{Z, Z^*}\}.$$

Also gilt die Gleichheit in (2.38). $\square$

Satz 2.4.2 (Dualitätssatz). *Seien f ein reellwertiges konvexes Funktional auf einer konvexen Menge $\Omega \subset X$, X ein Vektorraum, und G eine konvexe Abbildung von X in den normierten Raum Z . Es existiere ein x_1 , so dass $G(x_1) < \theta$ gilt und $\mu_0 = \inf \{f(x) : G(x) \leq \theta \text{ und } x \in \Omega\}$ endlich ist. Dann folgt:*

$$(2.39) \quad \inf_{\substack{G(x) \leq \theta \\ x \in \Omega}} f(x) = \max_{z^* \geq \theta} \varphi(z^*),$$

und das Maximum auf der rechten Seite wird an einem $z_0^ \geq \theta$ angenommen.*

Wenn das Infimum auf der linken Seite an einem $x_0 \in \Omega$ angenommen wird, so gilt

$$\langle G(x_0), z^* \rangle_{Z, Z^*} = 0,$$

und x_0 minimiert $f(x) + \langle G(x), z_0^* \rangle_{Z, Z^*}$ auf Ω .

Obwohl die Theorie der konvexen Programmierung nicht die Stetigkeit der konvexen Funktionale erfordert, sind bei den meisten Problemen die Funktionale nicht nur stetig, sondern auch Fréchet-differenzierbar. In diesem Fall ist es geeignet, die notwendigen und hinreichenden Bedingungen in differentieller Form auszudrücken. Das folgende Lemma verallgemeinert die Tatsache, dass eine auf dem Intervall $[0, \infty)$ konvexe Funktion, die ihr Minimum an einem inneren Punkt x_0 annimmt, die Gleichung $f'(x_0) = 0$ erfüllt. Wenn sie aber ihr Minimum an $x_0 = 0$ annimmt, so folgt: $f'(x_0) \geq 0$. In beiden Fällen gilt: $x_0 f'(x_0) = 0$.

Lemma 2.4.3. *Seien f ein Fréchet-differenzierbares konvexes Funktional auf einem reellen normierten Raum X und P ein konvexer Kegel in X . Eine notwendige und hinreichende Bedingung, dass $x_0 \in P$ das Funktional f minimiert, ist durch*

$$\delta f(x_0; x) \geq 0 \text{ für alle } x \in P \quad \text{und} \quad \delta f(x_0; x_0) = 0$$

gegeben.

Bemerkung 2.4.4. *In dem Lemma können die Fréchet-Differentiale überall durch die Gateaux-Differentiale ersetzt werden, vorausgesetzt, die Gateaux-Differentiale sind linear.*

2.5. Lokale Theorie der restringierten Optimierung.

Definition 2.5.1 (Regulärer Punkt). *Sei $T : D \rightarrow X$ eine stetig Fréchet-differenzierbare Transformation von der offenen Menge D in einen Banachraum X in den Banachraum Y . Wenn für $x_0 \in D$ die Abbildung $T'(x_0) : X \rightarrow Y$ surjektiv auf Y abbildet, so wird x_0 ein regulärer Punkt der Transformation T genannt.*

Beispiel 2.5.2. *Für $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ ist ein Punkt $x_0 \in \mathbb{R}^n$ ein regulärer Punkt, wenn die Jacobi-Matrix von T den Rang m hat.*

Satz 2.5.3. *Sei x_0 ein regulärer Punkt der Transformation $T : X \rightarrow Y$, wobei X und Y zwei Banachräume sind. Dann gibt es eine Umgebung $N(y_0)$ des Punktes $y_0 = T(x_0)$ und eine Konstante K , so dass die Gleichung $T(x) = y$ für alle $y \in N(y_0)$ eine Lösung besitzt, die die Abschätzung $\|x - x_0\| \leq K \|y - y_0\|$ erfüllt.*

Wir wollen nun notwendige Bedingungen für ein Extremum von f unter der Restriktion $H(x) = \theta$ angeben, wobei f ein reellwertiges Funktional auf dem Banachraum X und H eine Abbildung von X in den Banachraum Z sind.

Lemma 2.5.4. *f nehme an x_0 ein lokales Extremum unter der Nebenbedingung $H(x) = \theta$ an. Weiter seien f und H stetig Fréchet-differenzierbar in einer offenen Menge, die x_0 enthält, und x_0 ein regulärer Punkt von H . Dann folgt $f'(x_0)h = 0$ für alle h mit $H'(x_0)h = \theta$.*

Das obige Resultat kann geometrisch in dem Raum X mit Hilfe des Tangentialraumes der Restriktionen visualisiert werden. Mit dem Tangentialraum an x_0 meinen wir die Menge der Vektoren h , für die $H'(x_0)h = 0$ gilt, also den Nullraum von $H'(x_0)$. Dieses ist eine Teilmenge von X , die, wenn sie hinsichtlich x_0 gedeutet wird, als Tangente an den Teilraum $N = \{x : H(x) = \theta\}$ bei x_0 betrachtet werden

kann. Eine äquivalente Aussage zu dem Lemma 2.5.4 lautet, dass f an x_0 stationär hinsichtlich Variationen entlang der Tangentialebene ist.

Satz 2.5.5 (Lagrange-Multiplikatoren). *Wenn das stetig Fréchet-differenzierbare Funktional f an dem regulären Punkt x_0 ein lokales Minimum unter der Restriktion $H(x) = \theta$ besitzt, dann gibt es ein Element $z_0^* \in Z^*$, so dass das Lagrange-Funktional*

$$L(x) = f(x) + z_0^* H(x)$$

an x_0 stationär ist, das bedeutet: $f'(x_0) + z_0^* H'(x_0) = \theta$.

Wir wollen nun notwendige Bedingungen für das Problem

$$(2.40) \quad \text{minimiere } f(x) \quad \text{mit } G(x) \leq \theta,$$

wobei f auf dem Vektorraum X definiert und G eine Abbildung von X in den normierten Raum Z , der einen positiven Kegel hat, ist.

Um zu sehen, wie die Technik der Lagrange-Multiplikatoren auf Probleme dieser Art ausgedehnt werden kann, betrachten wir das Problem in zwei Dimensionen mit drei skalaren Ungleichungen $g_i(x) \leq 0$ als Nebenbedingungen. Der Rand der Restriktionsmenge besteht aus drei Teilrändern, wo jeweils ein g_i gleich Null ist. Wenn das Minimum an einem inneren Punkt x_0 der Restriktionsmenge angenommen wird, folgt offensichtlich sofort $f'(x_0) = 0$. Wenn allerdings das Minimum an einer Stelle am Rand $g_1(x) = 0$ angenommen wird, dann muß $f'(x_0)$ orthogonal zum Rand sein und nach innen zeigen (Richtung des stärksten Anstieges). In diesem Fall gilt $f'(x_0) + \lambda_1 g_1'(x_0) = \theta$ für ein $\lambda_1 \geq 0$, da auch $g_1'(x_0)$ orthogonal auf dem Rand steht. Wenn das Minimum an einer Stelle des Randes mit $g_1(x) = 0$ und $g_2(x) = 0$ (Schnittpunkt beider Teilränder) angenommen wird, so erhalten wir $f'(x_0) + \lambda_1 g_1'(x_0) + \lambda_2 g_2'(x_0) = \theta$ mit $\lambda_1 \geq 0$, $\lambda_2 \geq 0$. Alle diese Fälle können wir zusammenfassen durch die Gleichung

$$f'(x_0) + \lambda^* G'(x_0) = 0,$$

wobei $\lambda^* \geq \theta$ und $\lambda_i g_i(x_0) = 0$, $i = 1, 2, 3$, gelten. Die Gleichung $\lambda_i g_i(x_0) = 0$ besagt bloß, dass im Falle von $g_i(x_0) < 0$ der entsprechende Lagrange-Multiplikator aus der notwendigen Bedingung fehlt (x_0 ist innerer Punkt).

Indem wir verschiedene positive Kegel oder Restriktionen der Form $H(x) \leq \theta$, $-H(x) \leq \theta$ betrachten, erhalten wir, dass das Problem (2.40) Minimierungsprobleme mit Gleichungsrestriktionen enthält. Damit ist die allgemeine Behandlung von (2.40) mindestens genauso schwer wie die Behandlung von Problemen, die bloß Gleichungsrestriktionen enthalten. Daher werden wir hier nur die Nebenbedingungen mit Ungleichungen berücksichtigen.

Definition 2.5.6 (Regulärer Punkt). *Seien X ein Vektorraum und Z ein normierter Raum mit einem positiven Kegel P , der ein nicht-leeres Inneres besitze. Sei $G : X \rightarrow Z$ eine Abbildung, die in ihren Argumenten lineares Gateaux-Differential besitze. Ein Punkt $x_0 \in X$ heißt ein regulärer Punkt der Ungleichung $G(x) \leq \theta$, wenn $G(x_0) \leq \theta$ gilt und es ein $h \in X$ mit $G(x_0) + \delta G(x_0; h) < \theta$ gibt.*

Satz 2.5.7 (Verallgem. Kuhn-Tucker). *Seien X ein Vektorraum und Z ein normierter Raum mit einem positiven Kegel P , der ein nicht-leeres Inneres besitze. Seien f ein Gateaux-differenzierbares reellwertiges Funktional auf X und $G : X \rightarrow Z$ eine Gateaux-differenzierbare Abbildung. Die Gateaux-Differentiale seien linear in ihren Argumenten. An dem regulären Punkt x_0 der Nebenbedingung $G(x) \leq \theta$*

werde f minimiert unter dieser Nebenbedingung. Dann gibt es ein $z_0^* \in Z^*$, $z_0^* \geq \theta$, so dass die Lagrange-Funktion

$$f(x) + \langle G(x), z_0^* \rangle_{Z, Z^*}$$

an x_0 stationär ist. Weiter folgt: $\langle G(x_0), z_0^* \rangle_{Z, Z^*} = 0$.

Beispiel 2.5.8. Wenn X ein normierter Raum — also nicht nur ein Vektorraum — ist, f und G Fréchet-differenzierbar und die Lösung ein regulärer Punkt sind, so bedeutet die Folgerung des Satzes 2.5.7:

$$f'(x_0) + z_0^* G'(x_0) = \theta \quad \text{und} \quad \langle G(x_0), z_0^* \rangle_{Z, Z^*} = 0.$$

REFERENCES

- [1] D. G. Luenberger. *Linear and Nonlinear Programming*. Addison-Wesley Publishing Company, Reading, 1989.
- [2] D. G. Luenberger. *Optimization by Vector Space Methods*. John Wiley & Sons, New York, 1969.

S. VOLKWEIN, INSTITUT FÜR MATHEMATIK, KARL-FRANZENS-UNIVERSITÄT GRAZ, HEINRICH-STRASSE 36, A-8010 GRAZ, AUSTRIA

E-mail address: `stefan.volkwein@uni-graz.at`