2. Grundlagen der nicht-linearen konvexen Optimierung

2.1. **Die Karush-Kuhn-Tucker Bedingungen.** Unser Basisproblem (NLO) sei geben durch

$$\text{NLO} \begin{cases} \min f(x) & \\ g_i(x) & \leq 0,\ i \in I = \{1, ..., m\} \\ x \in \mathbb{R}^n & f, g_i \text{ stetig differenzierbar.} \end{cases}$$

Die Menge $X := \{x \in \mathbb{R}^n | g_i(x) \leq 0\, \forall i\}$ heißt Menge der *zulässigen Lösungen.* und wir setzten $X \neq \varnothing$ voraus.

Wir wollen nun Bedingungen für lokale Optima suchen. Unsere erste Idee könnte dabei sein, dass x^0 ein lokales Minimum ist, wenn in einer $\epsilon-$Umgebung geschnitten mit X kein besserer Punkt liegt. Die nicht-linearen Optimierer führen dafür eine Konkretisierung ein. Der Punkt x^0 ist nicht lokal optimal, wenn man entlang einer Kurve, die von x^0 ausgeht, zu einem besseren Punkt kommt. Eine stetig differenzierbare Kurve $\varphi : \mathbb{R}^1 \to \mathbb{R}^n$ heißt *zulässig*, wenn $\varphi(0) = x^0$ und $\varphi(\theta) \in X$ für kleines $\theta > 0$ gilt.

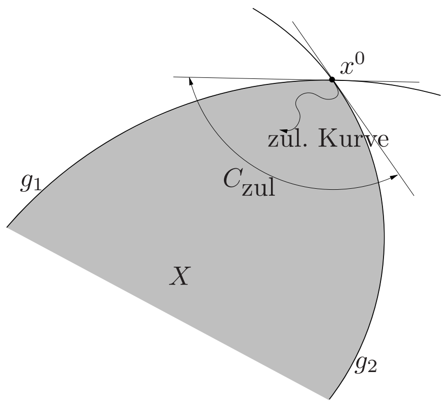

Abbildung 11. Zulässige Kurve in x^0.

Eine Tangente an einer zulässigen Kurve im Punkt 0 heißt *zulässige Richtung y*. Formaler ausgedrückt gilt

$$y = \frac{d\varphi}{d\theta}(0) = \begin{pmatrix} \frac{d\varphi_1}{d\theta}(0) \\ \vdots \\ \frac{d\varphi_n}{d\theta}(0) \end{pmatrix}.$$

Der Kegel, der von den zulässigen Richtungen in x_0 aufgespannt wird, bezeichnen wir mit $C_{\text{zul.}}$ (Abb. 11). Dieser Kegel spielt später eine wichtige Rolle für Optimalitätsbedingungen.

Betrachten wir zunächst notwendige Bedingungen dafür, dass $y \in C_{\text{zul.}}$ eine zulässige Richtung ist. Dazu definieren wir den Kegel

$$G := \{y \in \mathbb{R}^n | \nabla g_i^\top(x^0) y \leq 0\ \forall i \in I^0\}$$

mit $I^0 = \{i \in I | g_i(x^0) = 0\}$. Die Menge I^0 beschreibt die Indexmenge der *scharfen Nebenbedingungen* in x^0.

Lemma 2.1. *[notwendige Bedingungen für zulässige Richtungen] Sei y eine zulässige Richtung von x^0. Dann gilt*

$$\nabla g_i^\top(x^0) y \leq 0 \quad \forall i \in I^0,$$

d.h. $C_{zul.} \subset G$.

Beweis. Sei φ eine zulässige Kurve in x^0 und $y = \frac{d\varphi}{d\theta}(0)$. Dann gilt entweder $g_i(x^0) < 0$ für alle $i \notin I^0$, und daraus folgt $g_i(\varphi(\theta)) < 0$ für kleine θ. Oder es gilt $g_i(x^0) = 0$ für $i \in I^0$ und es folgt $g_i(\varphi(\theta)) \leq 0$ für kleines θ.

Betrachten wir nun die Taylorentwicklung von $g_i(\varphi(\theta))$. Die Taylorentwicklung von f in x^0 hat die Form

$$f(x) = \sum_{v=0}^{n} \frac{f^{(v)}(x^0)}{v!}(x - x^0)^v + (x - x^0)^{n+1} R_{n+1}.$$

Diese auf $g_i(\varphi(\theta))$ angewandt mit $x^0 = 0$, erhalten wir

$$g_i(\varphi(\theta)) = g_i(\varphi(0)) + \theta \nabla g_i^\top(\varphi(0)) \frac{d\varphi}{d\theta}(0) + \theta^2 R_2(\theta),$$

wobei $R_2(\theta) \to 0$ für $\theta \to 0$. Da $g_i(\varphi(0)) = g_i(x^0) = 0$ und $g_i(\varphi(\theta)) \leq 0$ gilt, muss $\theta \nabla g_i^\top(\varphi(0)) \frac{d}{d\theta}\varphi(0) \leq 0$ sein. Damit ist $\nabla g_i^\top(\varphi(0)) y \leq 0$ für $i \in I^0$. Diese Bedingung nennt man auch *Bedingung erster Ordnung.* □

Example 2.2. Leider sind die notwendigen Bedingungen noch nicht hinreichend und die Umkehrung gilt nicht. Betrachte das folgende Gebiet mit $g_1 : -x_1 \leq 0$, $g_2 : -x_2 \leq 0$ und $g_3 : -(1 - x_1)^3 + x_2 \leq 0$.

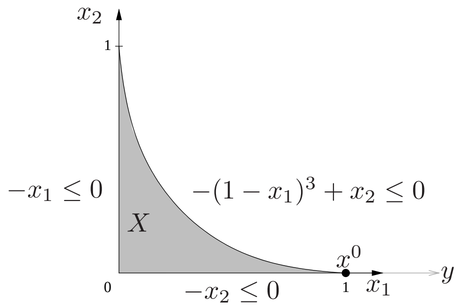

Abbildung 12. Gegenbeispiel zur Umkehrung von Lemma 2.1

Es gilt $\nabla g_2(x^0) = \binom{0}{-1}|_{x^0} = \binom{0}{-1}$ und $\nabla g_3(x^0) = \binom{3(1-x_1)^2}{x_2}|_{x^0} = \binom{0}{1}$ mit $x^0 = (1, 0)$. Für die Bedingungen erster Ordnung erhalten wir $-y_2 \leq 0$ und $y_2 \leq 0$, ohne dass eine Bedingung an y_1 gestellt wird. Daraus folgt, dass $y = \binom{1}{0}$ in G ist. Die Richtung y ist aber keine zulässige Richtung, da $x^0 + \theta\binom{1}{0} \notin X$ für beliebig kleines θ gilt.

Nach diesem Beispiel liegt es nahe noch einschränkende Bedingungen an X zu stellen. Wir fordern, dass X die constraint qualification Annahme (CQ) von Kuhn, Tucker '51 erfüllt. Die *constraint qualification Annahme* ist erfüllt, falls $\overline{C}_{\text{zul.}} = G$ für x^0 gilt. Diese Bedingung ist sehr stark und sichert die Umkehrung des Lemmas 2.1. Die CQ Bedingung ist allgemein schwer nachzuprüfen. Daher interessiert man sich für weitere hinreichende Bedingungen.

Lemma 2.3. *Jede der folgenden Bedingungen ist hinreichend für die Gültigkeit von (CQ):*

(1) *die Funktionen g_i sind linear (Korlin '59)*
(2) *alle Funktionen g_i sind konvex und das Innere von X ist nicht leer. (Slater '50)*
(3) *die Gradienten $\nabla g_i(x^0)$ mit $i \in I^0$ sind linear unabhängig.*

Note 2.4. Da die ersten beiden Bedingungen unabhängig von x^0 sind, sichern sie die Gültigkeit der Behauptung auf ganz X.

Beweis. zu (1): Sei $g_i(x) = a_i^\top x + b_i$. Dann ist X ein Durchschnitt von Halbräumen und jede Gerade von x^0 aus nach X ist eine zulässige Richtung. Es gilt $\nabla g_i(x^0) = a_i$. Damit entspricht $\nabla g_i(x^0)$ dem Vektor, der senkrecht auf der Hyperebene $a_i^\top x + b_i = 0$ steht. Die Gleichung $\nabla g_i(x^0)^\top y = a_i^\top y \leq 0$ wird von allen y erfüllt, die eine nicht-negative Projektion auf $-a_i$ haben. Diese y liegen im Halbraum, der durch $a_i^\top x + b_i = 0$ definiert wird und in dem X liegt. Also gilt $G = C_{\text{zul.}} = \overline{C}_{\text{zul.}}$.

Noch mal anders: Für ein $y \in G$ gilt$-a_i y \geq 0$, da $\nabla g_i^\top(x^0) = a_i$. Der Punkt x^0 ist zulässig, also ist $g_i(x^0) \leq 0$. Betrachten wir jetzt $g_i^\top(x^0 + \theta y) = a_i(x + \theta y) + b_i = \underbrace{a_i x + b_i}_{\leq 0} + \underbrace{a_i \theta y}_{\leq 0} \leq 0$, also ist $x + \theta y \in X$. Somit gilt $\varphi(\theta) = x^0 + y \cdot \theta$ ist eine zulässige Kurve und $y = \frac{d\varphi}{dx}(0)$ eine zulässige Richtung.

zu (2) und (3): Sei $R := \{y \in \mathbb{R}^n | \nabla g_i^T(x^0) y < 0, \ \forall i \in I^0\}$ der *Kegel der zulässigen Richtungen* in x^0. Für jedes $y \in R$ gilt, $\varphi(\theta) = x^0 + y\theta$ ist eine zulässige Kurve, weil $g_i(x^0 + y\theta) = \underbrace{g_i(x^0)}_{=0} + \theta \underbrace{\nabla g_i(x^0) y}_{<0} + \underbrace{R_2(\theta)}_{\to 0} \leq 0$ gilt. Also ist $R \subset C_{\text{zul.}}$ und damit $\overline{R} \subset \overline{C}_{\text{zul.}}$.

Betrachten wir nun folgende Behauptung:

Claim 2.5. Falls $R \neq \varnothing$, dann gilt $\overline{R} = G$, d.h. die (CQ) ist erfüllt.

Beweis. Zeige $G \subset R$. Sei $y \in G$ d.h. $\nabla g_i^\top(x^0) y \leq 0$ für alle $i \in I^0$ und $\overline{y} \in R$, d.h. $\nabla g_i^\top(x^0)\overline{y} < 0$ für $i \in I^0$. Dann gilt $\lambda y + (1 - \lambda)\overline{y} \in R$ für $0 \leq \lambda < 1$. Betrachte den Grenzübergang von $\lambda \uparrow 1$. Dies erzeugt eine Folge zulässiger Richtungen y^λ mit $\lim_{\lambda \uparrow 1} y^\lambda = y$. Also ist $y \in \overline{R}$. Da $\overline{R} \subset G$ nach der Definition von R und G gilt, folgt sofort $\overline{R} = G$. □

zu (2): Die Voraussetzung war, dass das Innere von X nicht leer ist. Daraus folgt, dass ein $\overline{x}$ mit $g_i(\overline{x}) < 0 \forall i$ existiert. Da alle g_i konvex sind, gilt

$$0 > g_i(\overline{x}) \geq g_i(x^0) + \nabla g_i^\top(x^0)(\overline{x} - x^0).$$

Für $i \in I^0$ gilt $g_i(x^0) = 0$ und somit $0 > g_i(\overline{x}) \geq \nabla g_i^\top(\overline{x} - x^0) \forall i \in I^0$. Also ist $(\overline{x} - x^0) \in R$ und damit ist $R \neq \varnothing$. Die Behauptung erhalten wir durch die Anwendung des Claim 2.5.

zu (3): Hier galt: die Vektoren $\nabla g_i(x^0)$ $i \in I^0$ sind linear unabhängig. Das ist äquivalent dazu, dass $\sum_{i \in I^0} \lambda_i \nabla g_i(x^0) = 0$ nur die triviale Lösung $\lambda_i = 0$ hat. Daraus folgt, dass $\lambda^\top \nabla g(x^0) = 0$ mit $g(x) = (g_1(x), ..., g_k(x))^\top$ für $\lambda \geq 0$ und $\lambda \neq 0$ keine Lösung hat.

Der Satz von Gordan 1873, eine Variante des Farkas-Lemmas, sagt aus: Sei $A \in \mathbb{R}^{n \times m}$.Dann gilt $\lambda^\top A = 0$ mit $\lambda \geq 0$ und $\lambda \neq 0$ hat genau dann keine Lösung, wenn ein y mit $Ay < 0$ existiert. Setzten wir $A = \left(\nabla g_1^\top(x^0), ..., \nabla g_k^\top(x^0)\right)^\top$ erhalten wir, dass ein y mit $\left(\nabla g_1^\top(x^0), ..., \nabla g_k^\top(x^0)\right) \cdot y < 0$ existiert. Komponentenweise betrachtet bedeutet dies, dass $\nabla g_i(x^0) y < 0$ für alle $i \in I^0$. Somit ist dieses $y \in R$ und wir erhalten die Behauptung mit Hilfe von Claim 2.5. □

Betrachten wir nun notwendige Bedingungen für die Existenz eines lokalen Minimums:

Theorem 2.6. *[notwendige Bedingungen für lokales Minimum, Kuhn-Tucker '51] Seien $f, g_i : \mathbb{R}^n \to \mathbb{R}$ stetig differenzierbare Funktionen. Außerdem sei die (CQ) in $x^0 \in X$ erfüllt. Ist x^0 ein lokales Minimum von NLO, so existieren Zahlen $\lambda_i \geq 0$ $(i \in I)$, die so genannten Kuhn-Tucker-Multiplikatoren, mit*

$$KT \begin{cases} \nabla f(x^0) + \sum_{i \in I} \lambda_i \nabla g_i(x^0) & = 0 \\ \lambda_i g_i(x^0) & = 0 \ \forall i. \end{cases}$$

Die beiden Bedingungen werden auch als Kuhn-Tucker Bedingungen bezeichnet.

Beweis. Sei x^0 ein lokales Minimum. Dann gilt $f(\varphi(\theta)) \geq f(\varphi(0)) = f(x^0)$ für kleine θ und jede zulässige Kurve φ . Die Taylor Entwicklung von $f(\varphi(\theta))$ in $\theta = 0$ ergibt

$$f(\varphi(\theta)) = f(\varphi(0)) + \theta \nabla f^\top(x^0) \frac{d\varphi}{d\theta}(0) + \underbrace{\theta^2 R_2(\theta)}_{\xrightarrow[\theta \to 0]{} 0}.$$

Mit $y := \frac{d\varphi}{d\theta}(0)$ erhalten wir dann nach Definition eine zulässige Richtung. Da x^0 ein lokales Minimum ist, gilt $\nabla f^\top(x^0) y \geq 0$ für alle zulässigen Richtungen $y \in C_{\text{zul.}}$.

Betrachten wir den Grenzübergang von $y^k \to y$. Dann erhalten wir aus der Stetigkeit der Funktionen, dass $\nabla f^\top(x^0) y \geq 0$ für alle $y \in \overline{C}_{\text{zul.}}$gilt. Mit der Constrain Qualification (CQ) ergibt sich $\nabla f(x^0)^\top y \geq 0$ für alle $y \in G = \{y | \nabla g_i^\top(x^0) y \leq 0\}$.

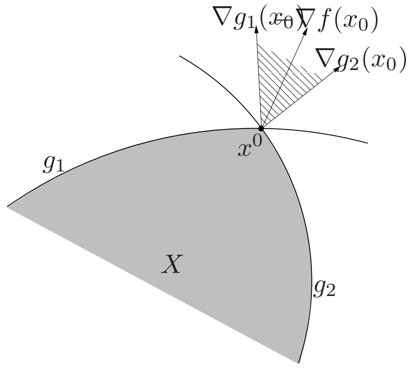

Abbildung 13. Der negative Gradient von f liegt im Kegel, der von $\nabla g_i(x^0)$ mit $i \in I_0$ aufgespannt wird.

Also ist notwendig dafür, dass x^0 ein lokales Minimum ist, dass $\nabla f^\top(x^0) y \geq 0$ für alle y mit $-\nabla g_i^\top(x^0) y \geq 0$ gilt. Dies hat wieder Ähnlichkeit mit dem Farkas Lemma [Sei $a_i \in \mathbb{R}^n$, $i = 1, ..., m$ und $c \in \mathbb{R}^n$. Folgt aus $y^\top a_i \geq 0$ für alle i, dass $y^T c \geq 0$ ist, dann ist dieses äquivalent dazu, dass $c \in C(a_1, ..., a_m)$] und wir erhalten $\nabla f^\top(x_0)$ ist ein Element des Kegels, der von den Vektoren $-\nabla g_i(x^0)$ mit $i \in I^0$ aufgespannt wird. Dies ist gleichbedeutend damit, dass $\lambda_i \geq 0$ für $i \in I^0$ mit $\nabla f(x^0) = \sum_{i \in I^0} \lambda_i(-\nabla g_i(x^0))$ existieren. Äquivalent dazu gilt mit $\lambda_i = 0$ für alle $i \in I \backslash I^0$, dass $\lambda_i \geq 0$ für $i \in I$ existierten, für die $\nabla f(x^0) + \sum_{i \in I} \lambda_i \nabla g_i(x^0) = 0$ gilt. Das kann man umformen zu: $\exists \lambda_i \geq 0$ mit $\lambda_i g_i(x^0) = 0$ für alle $i \in I$ und $\nabla f(x^0) + \sum_{i \in I} \lambda_i \nabla g_i(x^0) = 0$, die KT- Bedingungen. □

Note 2.7. Die λ_i sind eindeutig bestimmt, wenn die $\nabla g_i(x^0)$ linear unabhängig sind.

Betrachten wir nun die Verallgemeinerung auf " $\leq$ " und „=“ Restriktionen. Dazu erhalten wir das neue Problem (NLO$^+$)

$$\text{NLO}^+ \begin{cases} \min f(x) & \\ g_i(x) & \leq 0 \; i \in I \\ h_l(x) & = 0 \; l \in L \\ x & \in \mathbb{R}^n, \end{cases}$$

wobei $f, g_i, h_i : \mathbb{R}^n \to \mathbb{R}$ stetig differenzierbar sind. Die Begriffe Lösungsmenge, zulässige Richtungen und $C_{\text{zul.}}$ des NLO Problems lassen sich auf das NLO$^+$ übertragen. Die Menge G definieren wir jetzt als $G = \{y \in \mathbb{R}^n | \nabla g_i^\top(x^0) y \leq$

0 für alle $i \in I^0$, $\nabla h_l^\top(x^0)y = 0$ für alle $l \in L\}$. Die CQ- Bedingung bedeutet weiterhin $\overline{C}_{\text{zul.}} = G$. Für die Bedingungen $g_i(x) \leq x$ mit g_i stetig, gilt $C_{\text{zul.}} = \overline{C}_{\text{zul.}}$.

Theorem 2.8. *[Kuhn- Tucker '51] Seien $f, g_i, h_l : \mathbb{R}^n \to \mathbb{R}$ stetig differenzierbare Funktionen. Außerdem sei die (CQ) in $x^0 \in X$ erfüllt. Sei x^0 ein lokales Minimum. Dann gibt es Zahlen $\lambda_i \geq 0$ mit $i \in I$ und μ_l mit $l \in L$, so dass*

$$KT \begin{cases} \nabla f(x^0) + \sum_{i\in I} \lambda_i \nabla g_i(x^0) + \sum_{l\in L} \mu_l \nabla h_l(x^0) & = 0 \\ \lambda_i g_i(x^0) & = 0 \ \forall i \in I \end{cases}$$

gilt.

Beweis. Das Problem läßt sich auf den Fall aus Satz 2.6 zurückführen, indem man für $h(x) = 0$ die Bedingungen $h(x) \leq 0$ und $-h(x) \leq 0$ einführt. □

Note 2.9. Die (CQ) gilt in x^0, wenn eine der folgenden Bedingungen erfüllt sind:

(1) die Funktionen g_i sind konvex, die Funktionen h_l sind linear und es gibt ein $\overline{x} \in X$ mit $g_i(\overline{x}) < 0$ für alle i und $h_l(\overline{x}) = 0$ für alle l.
(2) Die Gradienten $\nabla g_i(x^0)$ und $\nabla h_l(x^0)$ sind (zusammen) linear unabhängig.

Wir wollen aber nicht nur notwendige sondern auch hinreichende Bedingungen für lokale Minima.

2.2. **Hinreichende Bedingungen für Optimalität: Sattelpunkte.** Betrachte wieder das nichtlineare Optimierungsproblem (NLO) mit

$$\text{NLO} \begin{cases} \min f(x) & \\ g_i(x) & \leq 0 \\ x \in & S \subset \mathbb{R}^n. \end{cases}$$

An die Funktionen $f, g_i : \mathbb{R}^n \to \mathbb{R}$ werden ansonsten keine Bedingung gestellt. Wir bringen die Nebenbedingungen $g_i(x) \leq 0$ mit den Lagrange-Multiplikatoren in die Zielfunktion und erhalten die *Lagrange-Funktion*

$$L(x, \lambda) := f(x) + \sum_{i\in I} \lambda_i g_i(x)$$

mit $\lambda_i \geq 0$. Der Punkt $(\overline{x}, \overline{\lambda})$ mit $\overline{x} \in S$ und $\overline{\lambda} \geq 0$ heißt *Sattelpunkt* (Abb. 14) von $L(x, \lambda)$, wenn

$$\begin{aligned} L(\overline{x}, \overline{\lambda}) &\leq L(x, \overline{\lambda}) \text{ für alle } x \in S \\ L(\overline{x}, \overline{\lambda}) &\geq L(\overline{x}, \lambda) \text{ für alle } \lambda \geq 0. \end{aligned}$$

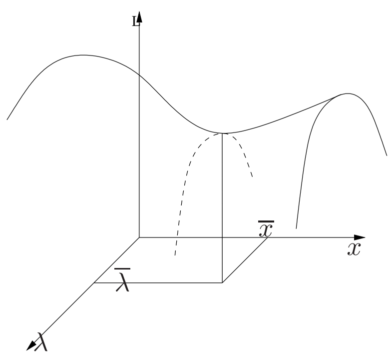

Abbildung 14. Sattelpunkt der Funktion $L(x, \lambda)$.

Wir werden zeigen, dass ein Sattelpunkt $(\overline{x}, \overline{\lambda})$ von $L(x, \lambda)$ ein globales Minimum $\overline{x}$ von f darstellt.

Theorem 2.10. *[charakteristische Eigenschaften von Sattelpunkten]*
Der Punkt $(\overline{x}, \overline{\lambda})$ mit $\overline{x} \in S$ und $\overline{\lambda} \geq 0$ ist genau dann ein Sattelpunkt von $L(x, \lambda)$ wenn

(1) $L(\overline{x}, \overline{\lambda}) = \min_{x \in S} L(x, \overline{\lambda})$
(2) $g_i(\overline{x}) \leq 0$ *für alle* $i \in I$
(3) $\overline{\lambda}_i g_i(\overline{x}) = 0$ *für alle* $i \in I$

gilt.

Beweis. " $\Rightarrow$ " zu (1): ist klar nach Definition.

zu (2): Es gilt $L(\overline{x}, \overline{\lambda}) \geq L(\overline{x}, \lambda)$ für alle $\lambda \geq 0$. Nach der Definition von $L(x, \lambda)$ erhalten wir $f(\overline{x}) + \sum_i \overline{\lambda}_i g_i(\overline{x}) \geq f(\overline{x}) + \sum_i \lambda_i g_i(\overline{x})$ für alle $\lambda \geq 0$. Umgeformt ergibt sich dann

$$\sum (\lambda_i - \overline{\lambda}_i) g_i(\overline{x}) \leq 0 \tag{1}$$

für alle $\lambda \geq 0$.

Annahme: (2) gilt nicht für i_0, d.h. $g_{i_0}(\overline{x}) > 0$. Dann kann λ_{i_0} so groß gewählt werden, dass $\sum_{i \in I} (\lambda_i - \overline{\lambda_i}) g_i(\overline{x}) > 0$. Dies ist ein Widerspruch zu (1).

zu (3): Betrachte den Vektor $\lambda = 0$. Für diesen muss auch (1) gelten. Damit gilt $\sum_{i \in I} -\overline{\lambda}_i g_i(\overline{x}) \leq 0$. Da für alle $\overline{\lambda} \geq 0$ und $g_i(\overline{x}) \leq 0$ die Ungleichung $\sum_{i \in I} \overline{\lambda_i} g_i(\overline{x}) \leq 0$ gilt (das haben wir gerade gezeigt), folgt nun, dass $\sum_{i \in I} \overline{\lambda}_i g_i(\overline{x}) = 0$ ergibt. Da $\lambda_i \geq 0$ ist und $g_i(\overline{x}) \leq 0$ muss für die einzelnen Summanden $\overline{\lambda}_i g_i(\overline{x}) = 0$ gelten.

" $\Leftarrow$ " (1),(2) und (3) gelten. Aus (1) folgt sofort, dass $L(\overline{x}, \overline{\lambda}) \leq L(x, \overline{\lambda})$ für alle $x \in S$ ist. Aus (3) erhalten wir

$$L(\overline{x}, \overline{\lambda}) = f(\overline{x}) + \sum \overline{\lambda}_i g_i(\overline{x}) = f(\overline{x}).$$

Außerdem wissen wir, dass

$$L(\overline{x}, \lambda) = f(\overline{x}) + \sum_i \lambda_i \underbrace{g_i(\overline{x})}_{\text{nach } (\textbf{??}(]\text{enu:(b)}2.5) \leq 0} \leq f(\overline{x}) = L(\overline{x}, \overline{\lambda})$$

für $\lambda \geq 0$ gilt. Die Bedingungen für eine Sattelpunkt sind also erfüllt. $\square$

Theorem 2.11. *Ist $(\overline{x}, \overline{\lambda})$ einen Sattelpunkt von $L(x, \lambda)$ so ist $\overline{x}$ ein globales Optimum von (NLO).*

Beweis. Nutze die drei Bedingungen von Satz 2.10. Für alle $x \in S \cap X$ gilt

$$\begin{array}{rrcl} & L(\overline{x}, \overline{\lambda}) & \leq & L(x, \overline{\lambda}) \\ \Leftrightarrow & f(\overline{x}) + \sum \underbrace{\overline{\lambda}_i g_i(\overline{x})}_{=0 \text{ nach } (3)} & \leq & f(x) + \sum \overline{\lambda}_i \underbrace{g_i(x)}_{\leq 0 \text{ für } x \in S \cap X} \\ & & \leq & f(x) \end{array}$$

Also ist $\overline{x}$ ein globales Minimum auf $S \cap X$, und $\overline{x} \in X$ wegen (2). $\square$

Note 2.12. Dieser Satz gilt ohne Voraussetzungen an f, g_i und S, also insbesondere auch für die ganzzahlige Optimierung.

Es stellt sich sofort die Frage, ob jedes globale Minimum von einem Sattelpunkt kommt? Dies muss im Allgemeinen nicht so sein.

Example 2.13. Wir wollen folgendes Problem betrachten:

$$\begin{aligned} \min f(x) &= -x^2 \\ 2x-1 &\leq 0 \\ -x &\leq 0 \\ x &\leq 1. \end{aligned}$$

Die letzten beiden Bedingungen definieren unsere Menge S (Abb. **??**).

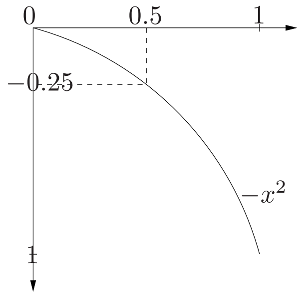

ABBILDUNG 15. Globales Minimum ohne Sattelpunkt.

Das Minimum wird bei $x = \frac{1}{2}$ mit $f(x) = -\frac{1}{4}$ angenommen. Die Lagrange-Funktion hat die Form

$$L(x,\lambda) = -x^2 + \lambda(2x-1).$$

Diese Funktion ist konkav in x für ein festes λ. Das bedeutet, dass $\min\limits_{x\in S} L(x,\lambda)$ für ein festes λ an einem der beiden Endpunkte angenommen wird. Also liegt das Minimum in $x = 0$ oder in $x = 1$. Kein $\overline{\lambda}$ liefert ein Minimum von $L(x,\overline{\lambda})$ bei $\frac{1}{2}$. Die Lagrangefunktion hat also keinen Sattelpunkt. Trotzdem existiert ein Minimum der Funktion f.

2.3. **Notwendige und hinreichende Bedingungen für die Existenz von Sattelpunkten: Stör-Funktionen.** Die Idee ist, das NLO in eine Familie von Problemen mit gestörter rechter Seite einzubetten. Wir betrachten also das NLO$_y$-Problem mit einem zusätzlichen Parameter:

$$\mathrm{NLO}_y \begin{cases} \min f(x) & \\ g_i(x) & \leq y_i,\ i\in I \\ x & \in S. \end{cases}$$

Die beiden Nebenbedingungen beschreiben die Menge $X(y)$ der zulässigen Lösungen. Die *Störfunktion* definieren wir als $\Phi(y) := \min_{x\in X(y)} f(x)$. Sie misst den Wert des Optimalwertes von f bei einer Variation von y. Es gilt $\Phi(0) = \min_{x\in X} f(x)$. Die Störfunktion ist antiton, d.h. für $y^{'} \leq y$ gilt $\Phi(y') \geq \Phi(y)$, da der Zulässigkeitsbereich von y' eine Teilmenge des Zulässigkeitsbereichs von y ist.

Theorem 2.14. *Das NLO habe ein globales Minimum $\overline{x}$ mit endlichem Wert $f(\overline{x})$. Dann gilt: der Punkt $(\overline{x},\overline{\lambda})$ ist genau dann ein Sattelpunkt, wenn die Hyperebene $z = \Phi(0) - \overline{\lambda}^\top y$ eine Stützhyperebene des Graphen der Störfunktion im Punkt $y = 0$ ist, d.h. $\forall y \in \mathbb{R}^m$ gilt $\Phi(y) \geq \Phi(0) - \overline{\lambda}^\top y$.*

Example 2.15. Wir wollen das Problem $\min x_1^2 + x_2^2$ unter $2x_1 + x_2 \leq -4$ lösen. Die Niveaumengen der Zielfunktion sind Kreise um den Nullpunkt. Der Berührpunkt mit der Gerade $2x_1 + x_2 = -4$ liefert dann das Optimum bei $\overline{x} = \binom{-\frac{8}{5}}{-\frac{4}{5}}$ mit dem Optimalwert $\frac{16}{5}$ (Abb. 16)

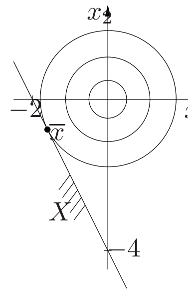

Abbildung 16. Das Optimum von $\min(x_1^2+x_2^2)$ unter $2x_1+x_2 \leq -4$ liegt bei $\overline{x} = (-\frac{8}{5}, -\frac{4}{5})$

Für das gestörte Problem gilt $\Phi(y) = \min(x_1^2+x_2^2)$ unter $2x_1+x_2+4 \leq y$. Damit erhalten wir mit $x_1 = \frac{2y-8}{5}$ und $x_2 = \frac{y-4}{5}$ für $y \leq 4$ und $x_1 = 0 = x_2$ für $y > 4$ das Optimum von NPO$_y$. Also gilt

$$\Phi(y) = \begin{cases} 0 & \text{für } y \geq 4 \\ (\frac{2y-8}{5})^2 + (\frac{y-4}{5})^2 = \frac{1}{5}y^2 - \frac{8}{5}y + \frac{16}{5} & \text{sonst} \end{cases}$$

Die Lagrangefunktion dieses Problemes hat die Form

$$L(x, \lambda) = x_1^2 + x_2^2 + \lambda(2x_1 + x_2 + 4).$$

Der Punkt $\overline{\lambda} = -\frac{8}{5}$ definiert den Sattelpunkt $(\overline{x}, \overline{\lambda})$ von $L(x, \lambda)$ und außerdem gilt $\Phi(y) \geq \Phi(0) - \overline{\lambda}y = \frac{16}{5} - \frac{8}{5}y$ (Abb. 17).

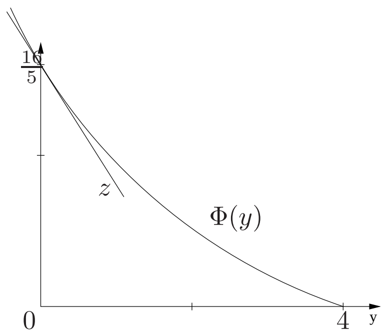

Abbildung 17. Die Funktion $\Phi(y)$ mit Stützhyperebene z in $y = 0$.

Beweis. „$\Rightarrow$“ Sei $g(x) := (g_1(x), ..., g_2(x))^\top$ und $(\overline{x}, \overline{\lambda})$ ein Sattelpunkt. Dann ist $f(x) + \overline{\lambda}^\top g(x) \geq f(\overline{x}) + \overline{\lambda}^\top g(\overline{x})$ für alle $x \in S \supset X(y)$ und $\overline{\lambda} \geq 0$. Aus den Bedingungen für eine Stattelpunkt erhalten wir $\overline{\lambda}^\top g(\overline{x}) = 0$ und $g(\overline{x}) \leq 0$. Also gilt

$$f(x) + \overline{\lambda}^\top g(x) \geq f(\overline{x}) = \Phi(0).$$

Betrachten wir ein y mit $X(y) \neq \varnothing$. Für ein solches y wissen wir aber, das $g(x) \leq y$ ist und $\overline{\lambda} \geq 0$. Also ist $\overline{\lambda}g(x) \leq \overline{\lambda} \cdot y$ und daraus folgt

$$f(x) \geq \Phi(0) - \overline{\lambda}^\top g(x) \geq \Phi(0) - \lambda^\top y.$$

Diese Gleichung gilt nun für jedes y, für das (NLO$_y$) einen Lösung hat und für alle $x \in X(y)$. Daraus folgt $\Phi(y) = \min\limits_{x \in X(y)} f(x) \geq \Phi(0) - \overline{\lambda}^\top y$

Falls das (NLO_y) keine Lösung hat, so ist $\Phi(y) = \infty$ und $\Phi(y) \geq \Phi(0) - \lambda^\top y$ automatisch erfüllt.

„$\Leftarrow$“: Es gibt ein $\overline{\lambda}$ mit $\Phi(y) \geq \Phi(0) - \overline{\lambda}^\top y$ für alle y. Wir wollen zeigen, dass $(\overline{x}, \overline{\lambda})$ ein Sattelpunkt ist.

(a): Wir zeigen, dass $\overline{\lambda_i} \geq 0$ gilt. Dazu nehmen wir an, dass ein $i \in I$ mit $\overline{\lambda}_i < 0$ existiert. Betrachte den Vektor y mit $y_i \to \infty$ und $y_j = 0$ für $j \in I \backslash i$. Dann gilt $\Phi(0) - \overline{\lambda} y \to \infty$. Dann muss nach Voraussetzung auch $\Phi(y)$ gegen ∞ gehen. Aber Φ ist antiton, d.h. für $y \geq 0$ gilt $\Phi(y) \leq \Phi(0)$. Damit kann Φ nicht beliebig wachsen und wir erhalten einen Widerspruch.

Wir wissen jetzt, dass die erste Bedingung für einen Sattelpunkt erfüllt ist.

(b): Wir wollen jetzt zeigen, dass $L(\overline{x}, \overline{\lambda}) \leq L(x, \overline{\lambda})$ für alle $x \in S$ gilt.

Sei $x \in S$. Es gilt $x \in X(g(x))$, da $g(x) \leq g(x)$ gilt. Dann erhalten wir $f(x) \geq \min\limits_{x' \in X(g(x))} f(x') = \Phi(g(x))$. Mit $\Phi(y) \geq \Phi(0) - \overline{\lambda}^\top y$ für alle y, ergibt sich

$$f(x) \geq \Phi(g(x)) \geq \Phi(0) - \overline{\lambda}^\top g(x).$$

Umgeformt gilt also

$$f(x) + \overline{\lambda}^\top g(x) \geq \Phi(0).$$

Betrachte nun $x = \overline{x}$. Dann erhält man

$$f(\overline{x}) + \overline{\lambda}^\top g(\overline{x}) \geq \Phi(0) = f(\overline{x}).$$

Daraus folgt, dass $\overline{\lambda}^\top g(\overline{x}) \geq 0$ ist. In Teil (a) hatten wir jedoch schon gezeigt, dass $\overline{\lambda} \geq 0$ ist und da $\overline{x}$ eine zulässige Lösung ist, gilt $g(\overline{x}) \leq 0$. Damit erhalten wir $\overline{\lambda}^\top g(\overline{x}) = 0$.

Da allgemein galt $f(x) + \overline{\lambda}^\top g(x) \geq \Phi(0)$, folgt

$$L(x, \overline{\lambda}) = f(x) + \overline{\lambda}^\top g(x) \geq f(\overline{x}) = f(\overline{x}) + \overline{\lambda}^\top g(\overline{x}) = L(\overline{x}, \overline{\lambda}).$$

Damit ist (b) bewiesen.

(c): $(\overline{x}, \overline{\lambda})$ ist ein Sattelpunkt.

Mit (b), $\overline{\lambda}^\top g(\overline{x}) = 0$ und $g(\overline{x}) \leq 0$ sind die Bedingungen von Satz 2.10 erfüllt. Also ist $(\overline{x}, \overline{\lambda})$ ein Sattelpunkt. □

Der Satz 2.14 kann auf das Problem (NLO^+) mit Gleichheitsbedingungen $h_l(x) = 0$ erweitert werden, wobei die zugehörigen μ_l dann nicht vorzeichenbeschränkt sind.

Im Beispiel 2.15 war $\Phi(y)$ konvex. Oft ist das NLO nicht konvex und Φ nicht konvex. Betrachten wir kurz vier unterschiedliche Fälle (Abb. 18)

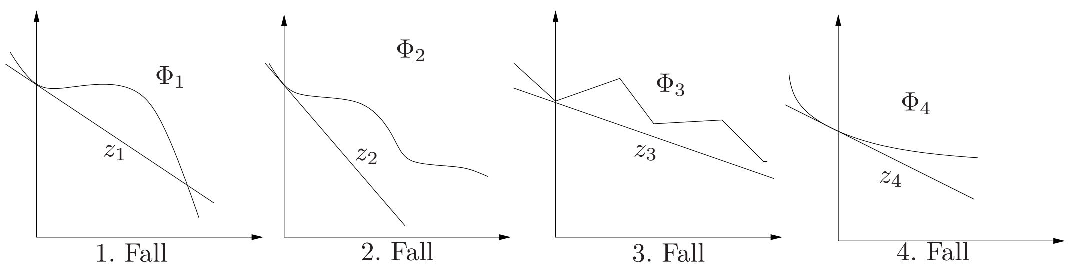

Abbildung 18. 1. Fall: Die Störfunktion ist nicht konvex und hat keine Stützhyperebene in $y = 0$. 2. Fall: Die Störfunktion ist nicht konvex, aber es existiert eine Stützhyperebene in $y = 0$. 3. Fall: Es existiert eine Stützhyperebene in $y = 0$. 4. Fall: Die Störfunktion ist konvex.

1. Fall: Es gibt keine Stützhyperebene in $y = 0$. Damit gibt es auch keinen Sattelpunkt.

2. Fall: Die Störfunktion Φ ist nicht konvex, aber dennoch existiert eine Stützhyperebene in $y = 0$. Dann gibt es auch einen Sattelpunkt und wir erhalten sofort ein Minimum von $f(x)$.

3. Fall: Die Störfunktion ist nicht differenzierbar und nicht konvex, aber es existiert ein Sattelpunkt. Dann gibt es auch eine Stützhyperebene.

4. Fall: Die Störfunktion ist konvex. Dann gibt es eine Stützhyperebene und damit auch einen Sattelpunkt.

2.4. **Sattelpunkte bei konvexen optimierungs Problemen.**

Theorem 2.16. *Seien die Funktionen f, g_i konvex, sei S konvex und es gebe ein $x \in S$ mit $g_i(x) < 0$ für alle i. Hat dann das (NLO) ein Minimum bei $\overline{x}$, so existiert ein $\overline{\lambda} \geq 0$ mit $(\overline{x}, \overline{\lambda})$ ist ein Sattelpunkt. D.h. bei konvexen Problemen kommen die Minima von einem Sattelpunkt.*

Beweis. Um den Satz zu beweisen, brauchen wir drei Eigenschaften:

(1): Die Störfunktion $\Phi(y)$ ist konvex.

Sei $\lambda \in [0, 1]$ und $y_1, y_2 \in \mathbb{R}$. Die Störfunktion ist definiert durch $\Phi(y) := \min\{f(x)|g(x) \leq y\}$. Sei $x_1 \in S$ mit $f(x_1) = \Phi(y_1)$ und $x_2 \in S$ mit $f(x_2) = \Phi(y_2)$. Dann gilt

$$\begin{aligned}(1-\lambda)\Phi(y_1) + \lambda\Phi(y_2) &= (1-\lambda)f(x_1) + \lambda f(x_2) \\ &\geq f((1-\lambda)x_1 + \lambda x_2),\end{aligned}$$

da f konvex ist. Der Punkt $x_3 := (1-\lambda)x_1 + \lambda x_2 \in S$, da S konvex ist. Außerdem gilt $g(x_3) \leq (1-\lambda)g(x_1) + \lambda g(x_2) \leq (1-\lambda)y_1 + \lambda y_2$, da auch g konvex ist. Also gilt $x_3 \in X((1-\lambda)y_1 + \lambda y_2)$ und wir können die Abschätzung von oben fortführen mit:

$$\begin{aligned}(1-\lambda)\Phi(y_1) + \lambda\Phi(y_2) &\geq f(x_3) \\ &\geq \min\{f(x)|g(x) \leq (1-\lambda)y_1 + \lambda y_2)\} \\ &= \Phi((1-\lambda)y_1 + \lambda y_2).\end{aligned}$$

(2): $\mathrm{dom}(\Phi) := \{y|\Phi(y) < \infty\} \neq \varnothing$

Da das NLO sein Optimum bei $\overline{x}$ annimmt, gilt $f(\overline{x}) < \infty$ und $g(\overline{x}) \leq 0$. Daraus folgt $0 \in \mathrm{dom}(\Phi)$.

(3): 0 ist aus dem Inneren von $\mathrm{dom}(\Phi)$.

Sei $y \geq 0$. Dann gilt $g_i(\overline{x}) \leq 0 < y$. Weiter erhalten wir $\Phi(y) \leq f(\overline{x}) < \infty$ und somit ist $y \in \mathrm{dom}(\Phi)$.

Sei jetzt $y < 0$. Nach Voraussetzung existiert ein $\tilde{x}$ mit $g_i(\tilde{x}) = y_i < 0$. Da immer noch $f(\tilde{x}) = \infty$ gelten kann, betrachten wir die nicht leere, offene Umgebung $U_{\tilde{x}}$ von $\tilde{x}$ mit $U_{\tilde{x}} := \{x|g_i(x) < 0 \text{ für alle } i \in I\}$. Eine solche Umgebung muss existieren, da g_i entweder stetig in $\tilde{x}$ sind (dann liegt $\tilde{x}$ im Inneren von S) oder die g_i machen an der Stelle $\tilde{x}$ einen Sprung nach unten, da die g_i konvex sind (dann liegt $\tilde{x}$ auf dem Rand von S). Da $f(\overline{x}) < \infty$ ist, muss auch aus der Umgebung $U_{\tilde{x}}$ ein x^*existieren, mit $f(x^*) < \infty$. Wähle $y = \max_{i \in I} g_i(x^*) < 0$. Dann gilt $\Phi(y) \leq f(x^*) < \infty$. Also ist $y < 0 \in \mathrm{dom}(\Phi)$. Damit liegt 0 im Inneren von $\mathrm{dom}(\Phi)$.

Ein Vektor γ für eine konvexe Funktion f heißt Subgradient in u, wenn $f(v) - f(u) \geq \gamma^\top(v-u)$ für alle v gilt. Da $y = 0$ im Inneren liegt, ist die konvexe Funktion Φ an der Stelle stetig. Daraus folgt, dass $\Phi(y)$ einen Subgradienten γ in $y = 0$ hat. Wenden wir die Definition des Subgradienten an, erhalten wir: $\exists\gamma$ mit $\Phi(y) - \Phi(0) \geq \gamma^\top(y-0)$. Umgeformt gilt: $\Phi(y) \geq \Phi(0) + \gamma^\top y$. Diese ist äquivalent zu: $\Phi(y) \geq \Phi(0) - \overline{\lambda}^\top y$ mit $\overline{\lambda} = -\gamma$.Nach Theorem 2.14 ist dann $(\overline{x}, \overline{\lambda})$ ein Sattelpunkt. $\square$

Conclusion 2.17. Im Fall, dass g_i und f konvex sind, gilt: Der Punkt $(\overline{x}, \overline{\lambda})$ ist genau dann ein Sattelpunkt, wenn $-\overline{\lambda}$ ein Subgradient der Störfunktion im Punkt $y = 0$ ist.

Eine entsprechende Aussage gilt für Gleichheitsbedingungen $h_l(x) = 0$.

2.5. **Kuhn-Tucker Bed. für den konvexen differenzierbaren Fall.**

Theorem 2.18. *Betrachte das nichtlineare Optimierungsproblem (NLO)*

$$NLO \begin{cases} \min f(x) & \\ g_i(x) \leq 0 & \forall i \\ f, g_i & \textit{konvex, differenzierbar} \end{cases}$$

Außerdem existiert ein $x \in \mathbb{R}^n$ mit $g_i(x) < 0$ für alle i. Dann ist $\overline{x} \in X$ genau dann ein globales Minimum von NLO, wenn die Kuhn-Tucker-Bedingungen in $\overline{x}$ erfüllt sind. D.h. $\exists \overline{\lambda} \geq 0$ mit $\nabla_x L(\overline{x}, \overline{\lambda}) = 0$ und $\overline{\lambda}_i g_i(\overline{x}) = 0$ für alle i.

Beweis. „$\Rightarrow$“ folgt aus Satz 2.6.
„$\Leftarrow$“ Aus der ersten KT- Bedingung $\nabla_x L(\overline{x}, \overline{\lambda}) = 0$ und da $L(x, \lambda)$ als Summe konvexer Funktionen konvex ist, folgt, dass $\overline{x}$ ein Minimum von $L(x, \overline{\lambda})$ ist. Da $\overline{x}$ zulässig ist, gilt $g_i(\overline{x}) \leq 0$. Zusammen mit der letzten KT- Bedingung, dass $\overline{\lambda} \geq 0$ und $\overline{\lambda}_i g_i(\overline{x}) = 0$, sind die Bedingungen für Satz 2.10 erfüllt. Damit ist $(\overline{x}, \overline{\lambda})$ ein Sattelpunkt und Satz 2.11 sagt aus, dass $\overline{x}$ dann ein globales Minimum ist. □

Wir machen jetzt noch eine Spezialisierung auf Nichtnegativitätsbedingungen und lineare Gleichungsbedingungen.
Betrachten wir als erstes die Nichtnegativitätsbedingungen $x_j \geq 0$.
1. Fall: Sei $n = 1$:

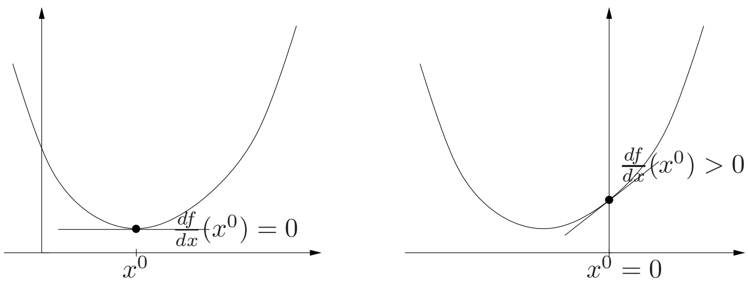

ABBILDUNG 19. Nichtnegativitätsbedingungen in Dimension 1

Dann erhalten wir für das Problem $\min f(x)$ unter $x \geq 0$ die Bedingung $\frac{df}{dx}(x^0) = 0$ für ein Optimales x^0 im ersten Fall (Abb. 19 links) und im zweiten Fall $\frac{df}{dx}(x^0) \geq 0$ und $x^0 = 0$ (Abb. 19 rechts). Also erhalten wir zusammengefasst die notwendigen Bedingung $x^0 \frac{df}{dx}(x^0) = 0$ und $\frac{df}{dx}(x^0) = 0$ für $x^0 \geq 0$. Entsprechend erhält man im mehrdimensionalen Fall $f : \mathbb{R}^n \to \mathbb{R}$ mit $\min f(x)$ unter $x \geq 0$, dass entweder für $x^0 > 0$ die Bedingung $\nabla f(x^0) = 0$ erfüllt sein muss, oder x^0 am Rand mit $x_j^0 = 0$ für ein oder mehrere j gilt. Komponentenweise ergeben sich die notwendigen Bedingungen:

$$x_j^0 \frac{\partial f}{\partial x_j}(x^0) = 0 \; \frac{\partial f}{\partial x_j}(x^0) \geq 0 \; x_j^0 \geq 0.$$

Diese Herleitung war jetzt sehr anschaulich. Man kann dies auch exakt aus den Kuhn-Tucker Bedingungen ablesen. Die Nebenbedingung hat die Form $g_j(x) =$

$-x_j \leq 0$. Kuhn-Tucker ist anwendbar, da die Nebenbedingungen konvex mit innerem Punkt sind und somit nach Lemma 2.3 die (CQ) erfüllt ist. Die KT-Bedingungen sagen jetzt aus, dass ein $\lambda_j \geq 0$ mit

$$\nabla f(x^0) + \sum_j \lambda_j \nabla g_j(x^0) = 0 \tag{2}$$

und

$$\lambda_j g_j(x^0) = 0 \;\forall j \tag{3}$$

existiert. Aus (2) erhalten wir

$$\frac{\partial f}{\partial x_j}(x^0) + \lambda_j(-1) = 0$$

für $j = 1, ..., n$. Also folgt $\lambda_j = \frac{\partial f}{\partial x_j}(x^0)$ und damit $\frac{\partial f}{\partial x_j}(x^0) \geq 0$, die zweite Bedingung. In (3) eingesetzt ergibt dies gerade

$$-x_j^0 \cdot \frac{\partial f}{\partial x_j}(x_0) = 0,$$

unsere erste Bedingung.

Betrachten wir nun lineare Gleichheitsbedingungen und Nichtnegativitätsbedingungen. Wir haben jetzt das Problem: minimiere $f(x)$ unter $h_i(x) = -\sum_j h_{ij}x_j + b_i = 0$ und $g_j(x) = -x_j \leq 0$. Die CQ ist erfüllt, da die Gleichheitsbedingungen linear und die Ungleichungsbedingungen konvex mit einem inneren Punkt sind. Also sind die KT-Bedingungen notwendig und wir erhalten: es gibt $\lambda_j \geq 0$ und μ_i beliebig mit

$$\nabla f(x^0) + \sum_j \lambda_j \nabla g(x^0) + \sum_i \mu_i \nabla h_i(x^0) = 0 \tag{4}$$

und

$$\lambda_j g_i(x^0) = 0 \forall j. \tag{5}$$

Betrachten wir Bedingungen (4) komponentenweise, erhalten wir

$$\frac{\partial f}{\partial x_j}(x^0) - \lambda_j - \sum_i \mu_i h_{ij} = 0$$

und damit

$$\lambda_j = \frac{\partial f}{\partial x_j}(x^0) - \sum_i \mu_i h_{ij}.$$

Da $\lambda_j \geq 0$ ist, erhalten wir

$$\frac{\partial f}{\partial x_j}(x^0) - \sum_i \mu_i h_{ij} \geq 0. \tag{6}$$

Setzten wir dieses in die Bedingung (5) ein, erhalten wir

$$x_j^0 (\frac{df}{dx_j}(x^0) + \sum_i \mu_i h_{ij}) = 0 \tag{7}$$

und wir erhalten für die Nebenbedingungen

$$\sum_j h_{ij} x_j = b_i \;\forall i \tag{8}$$

und

$$x_j \geq 0 \;\forall j. \tag{9}$$

Also: notwendige Bedingungen für Optimalität von x^0 ist die Existenz von Zahlen μ_i , $i = 1, ..., m$ mit (6)-(9).

Note 2.19. In beiden Fällen lässt sich über Sattelpunkte analog zu Satz 2.16 und 2.18 auch die Umkehrung zeigen.

2.6. **Anwendung auf das SO der Verkehrsoptimierung.** Das Verkehrsoptimierungsproblem hatte die folgende Form:

$$\min C(x(f)) = \sum_{a\in A} x_a \tau_a(x_a)$$
$$\sum_{p\in\mathcal{P}_k} f_p = d_k k \in C$$
$$f_p \geq 0 \forall p \in \mathcal{P}$$

wobei x der Kantenfluss und f der Pfadfluss ist, und $x_a = \sum_{p\ni a} f_p$ gilt. Betrachten wir die Kuhn-Tucker Bedingungen (6)-(9). Bedingung (7) sagt aus, dass

$$f_p(\frac{\partial C}{\partial f_p}(f) - \mu_k) = 0$$

für alle k und $p \in \mathcal{P}_k$ ist. Aus der Bedingung (6) ergibt sich dann

$$\frac{\partial C}{\partial f_p}(f) - \mu_k \geq 0.$$

Wir betrachten zunächst

$$\begin{aligned}
\frac{\partial C}{\partial f_p}(x(f)) &= \frac{\partial}{\partial f_p}(\sum_{a\in A} \underbrace{x_a(f)\tau_a(x_a(f))}_{=:c_a(x_a(f))}) \\
&= \sum_{a\in A} \frac{\partial}{\partial f_p} c_a(x_a(f)) \\
&= \sum_{a\in A} \frac{\partial c_a}{\partial x_a} \underbrace{(x_a)\frac{\partial x_a(f)}{\partial f_p}(f)}_{=:\delta_{p,a}} \\
&= [\qquad \underbrace{\sum_{a\in p} \tau_a(x_a) + x_a \frac{d}{dx_a}\tau(x_a)]}
\end{aligned}$$

marginale Kosten der Kante a, modfizierte Fahrzeifkt.

da $x_a = \sum_{p\ni a} f_p$ mit $\delta_{p,a} = \begin{cases} 1 & \text{falls } a \in p \\ 0 & \text{sonst} \end{cases}$. Die zwei Bedingungen sagen dann aus:

$$f_p(\sum_{a\in P} c'_a(x_a) - \mu_k) = 0 \; \forall k, \; \forall p \in \mathcal{P}_k \tag{10}$$

und

$$\sum_{a\in P} c'_a(x_a) - \mu_k \geq 0 \; \forall k, \; \forall p \in \mathcal{P}_k. \tag{11}$$

Man erhält also genau die Wardrop Bedingungen für das UE bzgl. $c'_a(x_a)$ als Fahrzeitfunktionen.

Weiter können wir aus der Gleichung (10) ableiten, dass alle Fluss führenden Wege zu Commodity k die selbe Länge bzgl. $c'_a(x_a)$ nämlich μ_k haben und (11) besagt, dass alle anderen Wege mindestens die Länge μ_k haben. Daraus können wir wieder folgern, dass das SO gleich dem UE bzgl. $c'(x_a)$ ist.