§2 Grundlagen der nichtlinearen (konvexen) Optimierung

2.1 Karush-Kuhn-Tucker Bedingungen

Basisproblem (NLO)

$$\min \; f(x)$$
$$\text{unter } g_i(x) \le 0 \qquad i = 1,\dots,m$$
$$x \in \mathbb{R}^n, \quad f, g_i : \mathbb{R}^n \to \mathbb{R} \quad \text{stetig, diff'bar}$$

$X := \{x \in \mathbb{R}^n \mid g_i(x) \le 0 \; \forall i\}$ heißt Menge der <u>zulässigen Lösungen</u>,

$X \neq \emptyset$ vorausgesetzt

Suchen Bedingungen für lokales Optimum

Idee: x^0 ist lokales Optimum, wenn in ε-Umgebung kein besserer Punkt ist

Konkretisierung:

Entlang einer <u>Kurve ausgehend von x^0</u> kommt man nicht zu einem besseren Punkt

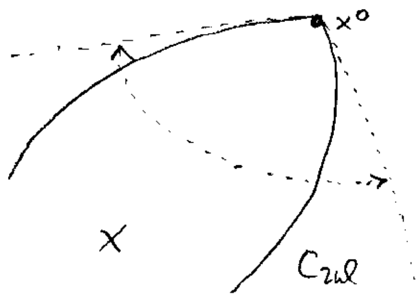
<u>zulässige Kurve</u>: $\varphi : \mathbb{R}^1 \to \mathbb{R}^n$ stetig, diff'bar
mit $\varphi(0) = x^0$, $\varphi(\Theta) \in X$ für kleines $\Theta > 0$

<u>zulässige Richtung in x_0</u>
= jede Tangente einer zulässigen Kurve φ in x^0

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

$C_{zul} :=$ Kegel (cone) aufgespannt durch alle zulässigen Richtungen in x^0

C_{zul} spielt wichtige Rolle für Optimalitätsbedingungen



zunächst: notwendige Bedingungen dafür, dass $y \in C_{zul}$

Sei $I^0 := \{ i = 1,\dots,m \mid g_i(x^0) = 0 \}$ scharfe Bedingungen in x^0

Sei $G := \{ y \in \mathbb{R}^n \mid \nabla g_i^T(x^0) \cdot y \leq 0 \quad \forall i \in I^0 \}$

G ist ebenfalls ein Kegel

2.1 LEMMA: (Notwendige Bedingung für zulässige Richtung)
Sei y eine zulässige Richtung in x^0. Dann gilt

$$\nabla g_i^T(x^0) \cdot y \leq 0 \qquad \forall i \in I^0$$

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

Beweis: Sei φ eine zulässige Kurve in x^0 und $y = \frac{d\varphi}{d\theta}(0)$ eine zulässige Richtung in x^0

a) $g_i(x^0) < 0$ für $i \notin I^0 \Rightarrow g_i(\varphi(\theta)) < 0$ für kleines θ

$g_i(x^0) = 0$ für $i \in I^0 \Rightarrow g_i(\varphi(\theta)) \leq 0$ für kleines θ

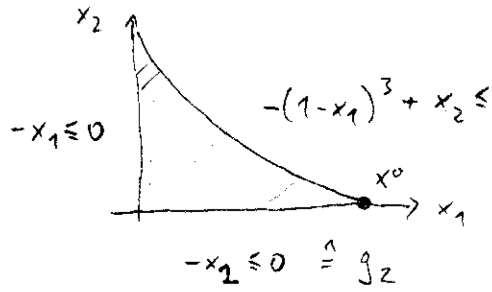
b) Taylor Entwicklung von $g_i(\varphi(\theta))$ in Umgebung von $\theta = 0$ und θ ergibt

$$\underbrace{g_i(\varphi(0))}_{g_i(x^0)} + \theta \nabla g_i^T(x^0) \frac{d\varphi}{d\theta}(0) + \underbrace{\theta^2 R_2(\theta)}_{\to 0 \text{ für } \theta \to 0} \leq 0$$

[Taylor: $f(x) = \sum_{\nu=0}^{n} \frac{f^{(\nu)}(x_0)}{\nu!}(x-x_0)^\nu + (x-x_0)^{n+1} \cdot R_n(x)$]

$g_i(x^0) = 0$ für $i \in I^0 \Rightarrow \nabla g_i^T(x^0) \cdot y \leq 0$ für $i \in I^0$ $\square$

Beachte: Die Umkehrung gilt i. A. nicht



$I^0 = \{2,3\}$

$$\nabla g_2(x_0) = \begin{pmatrix} 0 \\ -1 \end{pmatrix}\Bigg|_{x^0} = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$$

$$\nabla g_3(x_0) = \begin{pmatrix} 3(1-x_1)^2 \\ 1 \end{pmatrix}\Bigg|_{x^0} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

Also: $\nabla g_i^T(x^0) \cdot y \leq 0 \;\hat{=}\; \underbrace{\begin{pmatrix} 0 \\ -1 \end{pmatrix}^T \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \leq 0}_{-y_2 \leq 0}$ und $\underbrace{\begin{pmatrix} 0 \\ 1 \end{pmatrix}^T \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \leq 0}_{y_2 \leq 0}$

erfüllt von $y = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, aber $x^0 + \theta \begin{pmatrix} 1 \\ 0 \end{pmatrix} \notin X$

Daher einschränkende Bedingungen an X:

X erfüllt die constraint qualification Annahme (Kuhn & Tucker 51) (CQ)

$:\Leftrightarrow$ abgeschlossene Hülle $\bar{C}_{zul} = G$

ist sehr starke Bedingung, sichert Umkehrung in Lemma 2.1

(zumindest in dem Sinne, dass jedes $y \in G$ Limes einer Folge von zulässigen Richtungen ist)

((CQ) ist schwer zu überprüfen. Daher ist man an hinreichenden Bedingungen interessiert

Hinreichende Bedingungen für die Umkehrung:

2.2 LEMMA: Jede der folgenden Bedingungen ist hinreichend für die Gültigkeit von (CQ)

(a) alle Funktionen g_i sind linear (Karlin 59)

(b) alle Funktionen g_i sind konvex und das Innere von X ist nicht leer (Slater 50)

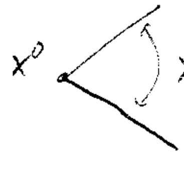
(c) die Gradienten $\nabla g_i(x^0)$ $i \in I^0$ sind linear unabhängig (Fiacco & McCormick 68)

Bemerkung: da (a) und (b) unabhängig von x^0 sind, sichern sie die Gültigkeit der Behauptung auf ganz X.

Beweis: <u>zu (a):</u> Sei $g_i(x) = a_i^T x + b_i$,

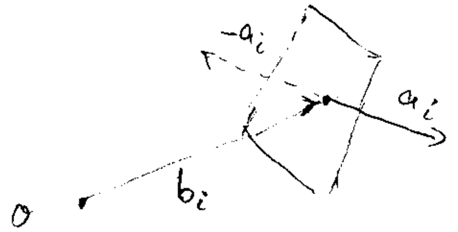
$\Rightarrow$ X ist Durchschnitt von Halbräumen. O.B.d.A sei $x^0 = 0$

$\Rightarrow$ jede Gerade von x^0 aus in X ist zulässige Richtung



$\nabla g_i(x^0) = a_i$: Vektor senkrecht auf der Hyperebene $a_i^T x + b_i = 0$

(Normalenvektor)



$\nabla g_i(x^0)^T y = a_i^T y \le 0$

erfüllt für alle y, die nicht-negative

Projektion auf $-a_i$ haben

$\Rightarrow$ alle y im Halbraum der durch $a_i^T x + b_i = 0$ definiert wird

und in dem X liegt

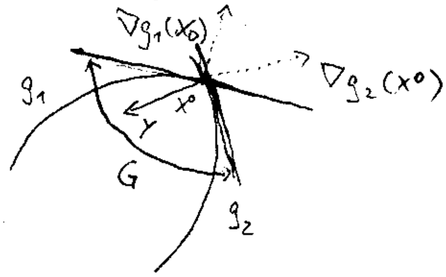
$\Rightarrow G = C_{zul} = \overline{C}_{zul}$

zu (b) und (c)

Sei $R := \{ y \mid \nabla g_i^T(x^0) \cdot y < 0 \}$ Kegel der zulässigen Richtungen in x^0

(Def)

klar: $R \subseteq C_{zul} \Rightarrow \overline{R} \subseteq \overline{C}_{zul}$



Claim: $R \neq \emptyset \Rightarrow \overline{R} = G$ (d.h. (CQ) erfüllt)

denn: Sei $\bar{y} \in R$, d.h. $\nabla g_i^T(x^0) \cdot \bar{y} < 0 \quad \forall i \in I^0$

Sei $y \in G$ beliebig, d.h. $\nabla g_i^T(x^0) \cdot y \le 0 \quad \forall i \in I^0$

$\Rightarrow \lambda y + (1-\lambda)\bar{y} \in R$ für alle $\lambda \in (0,1)$

Grenzübergang $\lambda \nearrow 1$ erzeugt Folge zulässiger Richtungen $y^\lambda \in R$

mit $\lim\limits_{\lambda \nearrow 1} y^\lambda = y \Rightarrow \overline{R} = G$

zu (b): $\operatorname{int}(X) \neq \emptyset \Rightarrow \exists\, \bar{x}$ mit $g_i(\bar{x}) < 0 \quad \forall i$

Konvexität der $g_i \Rightarrow 0 > g_i(\bar{x}) \geqslant \underbrace{g_i(x^0)}_{\uparrow} + \nabla g_i^T(x^0)(\bar{x} - x_0)$

beliebiges x^0

$\geqslant \nabla g_i^T(x^0)(\bar{x} - x_0)$

da $g_i(x^0) = 0$ für $i \in I^0$

$\Rightarrow \bar{x} - x_0 \in R$

$\Rightarrow$ Behauptung mit Claim

zu (c): $\nabla g_i(x^0)$ für $i \in I^0$ linear unabhängig

$\Leftrightarrow \sum_{i \in I^0} \lambda_i \nabla g_i(x^0) = 0$ hat nur die triviale Lösung $\lambda = 0$

$\Rightarrow \lambda^T \begin{pmatrix} \nabla g_1^T(x^0) \\ \vdots \\ \nabla g_k^T(x^0) \end{pmatrix} = 0 \quad \lambda \geqslant 0,\ \lambda \neq 0$ hat keine Lösung

$\overset{\text{Farkas Variante}}{\underset{(*)}{\Rightarrow}} \quad \exists\, y$ mit $\begin{pmatrix} \nabla g_1^T(x^0) \\ \vdots \\ \nabla g_k^T(x^0) \end{pmatrix} \cdot y < 0$

$\Leftrightarrow \nabla g_i^T(x^0) \cdot y < 0 \quad \forall i \in I^0$

$\Leftrightarrow y \in R$

zu $(*)$: Sei $A = \begin{pmatrix} \nabla g_1^T(x^0) \\ \vdots \end{pmatrix}$. Dann sagt $(*)$

$\underbrace{\lambda^T A = 0,\ \lambda \geqslant 0,\ \lambda \neq 0 \text{ hat keine Lösung}}_{(1)} \Leftrightarrow \underbrace{\exists\, y \text{ mit } Ay < 0}_{(2)}$

(2) $\Leftrightarrow Ay \leq -\mathbb{1}$ (Skalierung)

Lemma 7.17 in ADM II: $\{y \mid Ay \leq -\mathbb{1}\} = \emptyset \Leftrightarrow \exists\, \lambda \geqslant 0 \quad \lambda^T A = 0,\ \underbrace{\lambda^T(-\mathbb{1}) < 0}_{\Rightarrow \lambda \neq 0}$

also: Lemma 7.17 sichert $\neg(2) \Rightarrow \neg(1)$, Widerspruch $\square$

Jetzt notwendige Bedingungen für Existenz eines lokalen Minimums:

2.3 SATZ (Kuhn & Tucker 51)

$f, g_i : \mathbb{R}^n \to \mathbb{R}^1$ stetig diffbar, (CQ) erfüllt in $x^0 \in X$.

Ist x^0 lokales Minimum von (NLP): $\min f(x)$, $x \in X = \{x \in \mathbb{R}^n \mid g_i(x) \leq 0 \ \forall i\}$

so existieren Zahlen $\lambda_i \geq 0$ $(i \in I)$ (die Kuhn-Tucker Multiplikatoren)

mit

$$(KT) \begin{cases} \nabla f(x^0) + \sum\limits_{i \in I} \lambda_i \nabla g_i(x^0) = 0 \\ \lambda_i g_i(x^0) = 0 \quad \text{für alle } i \end{cases}$$

Beweis:

x^0 lokales Minimum $\Rightarrow$ $f(\varphi(\theta)) \geq f(\varphi(0)) = f(x^0)$

für jede zulässige Kurve φ und kleines $\theta > 0$

Taylor Entwicklung von $f(\varphi(\theta))$ in $\theta = 0$ ergibt

$$f(\varphi(\theta)) = f(x^0) + \theta \nabla f^T(x^0) \underbrace{\frac{d\varphi}{d\theta}(0)}_{=: y} + \underbrace{\theta^2 R_2(\theta)}_{\to 0 \text{ für } \theta \to 0}$$

zulässige Richtung nach Def, d.h. $y \in C_{zul}$

x^0 lokales Minimum $\Rightarrow$ $\nabla f^T(x^0) \cdot y \geq 0$ für alle $y \in C_{zul}$

Stetigkeit des Fkt, Grenzübergang $y^k \to y$ ergibt

$$\nabla f^T(x^0) \cdot y \geq 0 \quad \text{für alle } y \in \overline{C_{zul}}$$

Constraint Qualification (CQ) ergibt

$$\nabla f^T(x^0) \cdot y \geq 0 \quad \text{für alle } y \in G$$

Also: notwendig dafür, dass x^0 ein lokales Minimum ist ist

$$\nabla f^T(x^0) \cdot y \geq 0 \quad \forall y \text{ mit } -\nabla g_i^T(x^0) \cdot y \geq 0$$

$\Leftrightarrow$ (Farkas Lemma) $\quad \nabla f(x^0) \in \underbrace{C(-\nabla g_i(x^0),\ i \in I^0)}_{\text{Kegel von den } -\nabla g_i(x^0) \text{ aufgespannt}}$

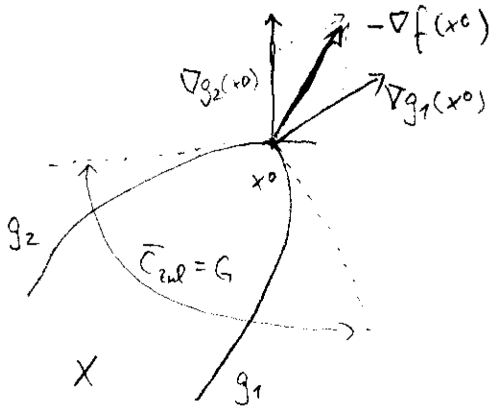
$\Leftrightarrow \quad \exists\, \lambda_i \geq 0,\ i \in I^0,$ mit $\nabla f(x^0) = \sum_{i \in I^0} \lambda_i \cdot (-\nabla g_i(x^0))$

$\Leftrightarrow$ (mit $\lambda_i = 0$ für alle $i \in I \setminus I^0$)

$\exists\, \lambda_i \geq 0,\ i \in I$ mit $\nabla f(x^0) = -\sum_{i \in I^0} \lambda_i \nabla g_i(x^0)$

$\Leftrightarrow \quad \exists\, \lambda_i \geq 0$ mit $\lambda_i g_i(x^0) = 0 \quad \forall i \in I$

und $\nabla f(x^0) + \sum_{i \in I} \lambda_i \nabla g_i(x^0) = 0 \qquad \square$



Bemerkung: Sind die $\nabla g_i(x^0)$ linear unabhängig, so sind die λ_i eindeutig bestimmt

Verallgemeinerung auf "$\leq$" und "$=$" Restriktionen

$$(NLO^+) \quad \begin{cases} \min\ f(x) \text{ unter} \\ g_i(x) \leq 0 \quad i \in I = \{1,..,m\} \\ h_\ell(x) = 0 \quad \ell \in L = \{1,..,p\} \\ x \in \mathbb{R}^n \end{cases}$$

alle Fkt stetig diffbar

Begriffe von NLO übertragen:

$$X = \{ x \mid g_i(x) \leq 0,\ i \in I,\ h_\ell(x) = 0\ \ \ell \in L \}$$

zulässige Richtung analog definiert, C_{zul} ebenso

$$G = \{ y \in \mathbb{R}^n \mid \nabla g_i^T(x^0)^T y \leq 0,\ \underset{\substack{\uparrow \\ g_i(x^0)=0}}{i \in I^0};\ \nabla h_\ell^T(x^0) \cdot y = 0 \ \ \ell \in L \}$$

Constraint Qualification (CQ) $\quad \bar{C}_{zul} = G$

2.4 SATZ (Kuhn & Tucker 1951)

Betrachte (NLO$^+$). x^0 ist lokal optimal, (CQ) gilt in x^0 $\Rightarrow$

$\exists$ Zahlen $\lambda_i \geq 0$ $(i \in I)$ und μ_ℓ $(\ell \in L)$ nicht vorzeichenbeschränkt

mit

$$\text{KT} \left\{ \begin{array}{ll} \nabla f(x^0) + \sum_{i \in I} \lambda_i \nabla g_i(x^0) + \sum_{\ell \in L} \mu_\ell \nabla h_\ell(x^0) = 0 & \\ \lambda_i \cdot g_i(x^0) = 0 & i \in I \end{array} \right.$$

Beweis: | Aufgabe 3 |

Bemerkung: (CQ) gilt in x^0, wenn eine der folgenden Bedingungen erfüllt ist:

(a) g_i sind konvex, h_ℓ sind linear
$\exists\ \bar{x} \in X$ mit $g_i(\bar{x}) < 0\ \forall i$ und $h_\ell(\bar{x}) = 0\ \forall \ell$

(b) Die Gradienten $\nabla g_i(x^0)$ und $\nabla h_\ell(x^0)$ sind (zusammen) linear unabhängig

2.2 Hinreichende Bedingungen für Optimalität: Sattelpunkte

Betrachte wieder (NLO): $\min f(x)$

$g_i(x) \le 0 \quad i \in I$ ← keine Annahmen

$x \in S \subseteq \mathbb{R}^n$ ← neue Bedingung

z.B. $S = \mathbb{Z}^n$

Bringe Nebenbedingung durch Lagrange-Multiplikatoren in die Zielfunktion

$$L(x,\lambda) := f(x) + \sum_{i \in I} \lambda_i g_i(x) \qquad \lambda_i \geq 0$$

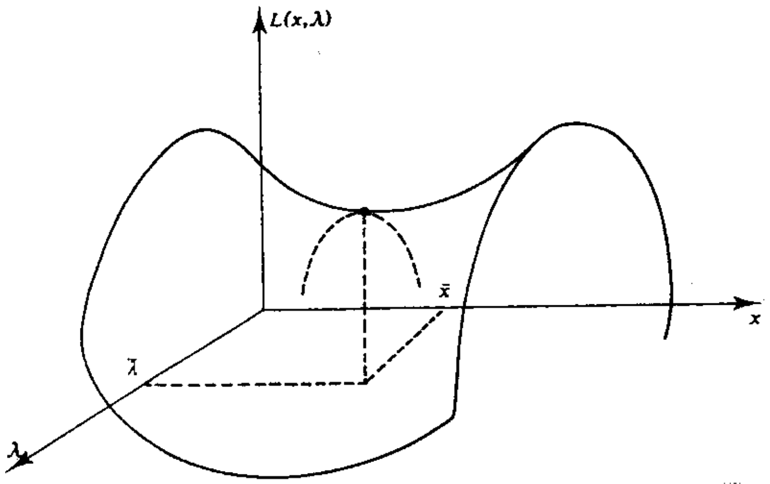
↑

Lagrange Funktion

$(\bar{x},\bar{\lambda})$ mit $\bar{x} \in S$ und $\bar{\lambda} \geq 0$ heißt Sattelpunkt von $L(x,\lambda)$

$:\Leftrightarrow \quad L(\bar{x},\bar{\lambda}) \le L(x,\bar{\lambda}) \quad \forall x \in S$

$\quad\quad L(\bar{x},\bar{\lambda}) \geq L(\bar{x},\lambda) \quad \forall \lambda \geq 0$



2.5 SATZ: (Charakteristische Eigenschaft von Sattelpunkten)

$(\bar{x},\bar{\lambda})$ mit $\bar{x} \in S$ und $\bar{\lambda} \geq 0$ ist Sattelpunkt von $L(x,\lambda) \Leftrightarrow$

(a) $L(\bar{x},\bar{\lambda}) = \min_{x \in S} L(x,\bar{\lambda})$

(b) $g_i(\bar{x}) \le 0 \quad \forall i \in I$

(c) $\bar{\lambda}_i \cdot g_i(\bar{x}) = 0 \quad \forall i \in I$

Beweis: "$\Rightarrow$"

(a) ist trivial

zu (b): Def $\Rightarrow \quad L(\bar{x},\bar{\lambda}) \geq L(\bar{x},\lambda) \quad \forall \lambda \geq 0$

$$\overset{\text{Def } L}{\Leftrightarrow} \quad f(\bar{x}) + \sum_i \bar{\lambda}_i g_i(\bar{x}) \geq f(\bar{x}) + \sum_i \lambda_i g_i(\bar{x}) \qquad \forall \lambda \geq 0$$

$$\Leftrightarrow \quad \sum_{i \in I} (\lambda_i - \bar{\lambda}_i) g_i(\bar{x}) \leq 0 \qquad \forall \lambda \geq 0 \qquad (*)$$

Annahme: (b) gilt nicht für i_0, d.h. $g_{i_0}(\bar{x}) > 0$

$\Rightarrow$ mache λ_{i_0} so groß, dass $\sum_{i \in I} (\lambda_i - \bar{\lambda}_i) g_i(\bar{x}) > 0$

$\Rightarrow$ Widerspruch

zu (c) $\lambda = 0 \quad \overset{(*)}{\Rightarrow} \quad \sum_i -\bar{\lambda}_i g_i(\bar{x}) \leq 0$

$\bar{\lambda} \geq 0, \; g_i(\bar{x}) \leq 0 \quad \Rightarrow \quad \sum_i \bar{\lambda}_i g_i(\bar{x}) \leq 0$

$\Rightarrow \sum_i \underbrace{\bar{\lambda}_i g_i(\bar{x})}_{\leq 0} = 0$

$\Rightarrow \bar{\lambda}_i g_i(\bar{x}) = 0 \quad \forall i$

"$\Leftarrow$" (a), (b), (c) gelten

(a) $\Rightarrow \quad L(\bar{x},\bar{\lambda}) \leq L(x,\bar{\lambda}) \qquad \forall x \in S$

(c) $\Rightarrow \quad L(\bar{x},\bar{\lambda}) = f(\bar{x})$

$$L(\bar{x},\lambda) \overset{\text{Def}}{=} f(\bar{x}) + \sum_i \underbrace{\lambda_i}_{\geq 0} \underbrace{g_i(\bar{x})}_{\leq 0 \text{ nach (b)}} \leq f(\bar{x}) = L(\bar{x},\bar{\lambda}) \qquad \forall \lambda \geq 0$$

$$\Rightarrow L(\bar{x},\lambda) \leq L(\bar{x},\bar{\lambda}) \leq L(x,\bar{\lambda}) \qquad \forall x \in S, \; \lambda \geq 0 \qquad \square$$

2.6 SATZ. Ist $(\bar{x},\bar{\lambda})$ ein Sattelpunkt von $L(x,\lambda)$ so ist $\bar{x}$ ein globales Optimum von (NLP)

Beweis: Nutze Satz 2.5. (a) ⇒ (c)

$$f(\bar{x}) + \sum_i \underbrace{\bar{\lambda}_i\, g_i(\bar{x})}_{=0 \text{ nach (c)}} \le f(x) + \sum_i \bar{\lambda}_i\, g_i(x) \qquad \forall x \in S$$

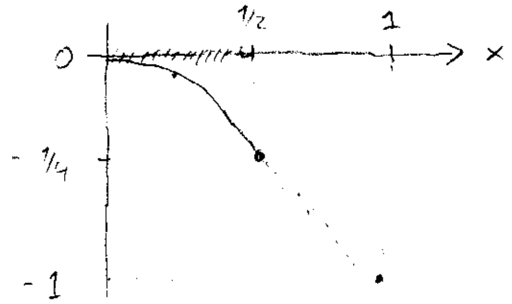
$$\Rightarrow\quad f(\bar{x}) \le f(x) + \sum_i \underbrace{\bar{\lambda}_i}_{\ge 0}\, \underbrace{g_i(x)}_{\le 0} \qquad \forall x \in S$$

$$\Rightarrow\quad f(\bar{x}) \le f(x) \qquad \forall x \in S \qquad \square$$

Beachte: Dieser Satz gilt ohne Voraussetzungen an f und die g_i, also auch für ganzzahlige Optimierung

ABER: Es kann Minima ohne Sattelpunkt geben.

Bsp: $\min f(x) = -x^2$ unter $2x - 1 \le 0$, $0 \le x \le 1$



Wähle $S = [0, 1]$

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

↑

konkav in x für festen λ

$\Rightarrow \min\limits_x L(x,\lambda)$ wird bei $x = 0$ oder $x = 1$ angenommen

aber: $x^0 = \frac{1}{2}$ ist Minimum von $f(x)$

kein $\bar{\lambda}$ liefert Minimum von $L(x,\bar{\lambda})$ bei $\frac{1}{2}$

$\Rightarrow L(x,\lambda)$ hat keinen Sattelpunkt

2.3 Notwendige und hinreichende Bedingungen für die Existenz von Sattelpunkten: Stör-Funktionen

Idee: (NLP) einbetten in eine Familie von Problemen mit gestörter rechter Seite

$$(NLP_y): \quad \min f(x) \quad \text{unter} \quad \left.\begin{array}{l} g_i(x) \le y_i \quad i \in I \\ x \in S \end{array}\right\} X(y)$$

$$\Phi(y) := \min_{x \in X(y)} f(x)$$

"Störfunktion"

$$\Rightarrow \Phi(0) = \min_{x \in X} f(x)$$

Störfunktion Φ ist antiton, d.h. $y' \le y \Rightarrow \Phi(y') \ge \Phi(y)$

(klar, da $X(y') \subseteq X(y)$)

2.7 SATZ: (NLP) habe ein globales Minimum $\bar{x}$ mit endlichem Wert $f(\bar{x})$. Dann gilt:

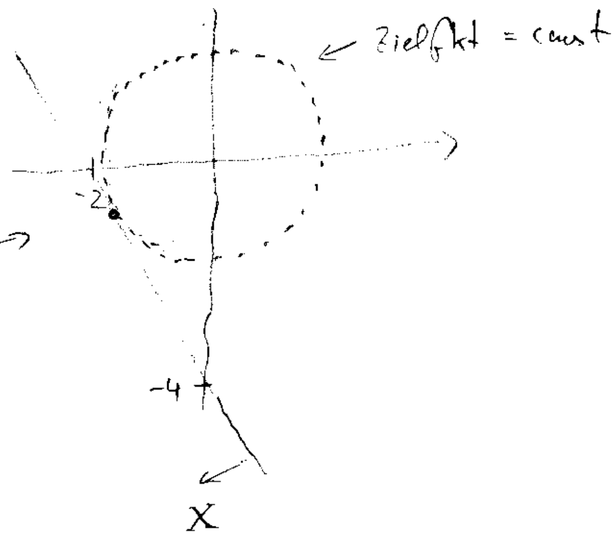
$(\bar{x}, \bar{\lambda})$ ist Sattelpunkt

$\Leftrightarrow$ die Hyperebene $z = \Phi(0) - \bar{\lambda}^T y$ ist Stützhyperebene des Graphen der Störfunktion Φ in $y = 0$

$\Leftrightarrow$ $\forall\, y \in \mathbb{R}^m$ gilt $\Phi(y) \ge \Phi(0) - \bar{\lambda}^T y$

Beispiel: $\min x_1^2 + x_2^2$

unter $2x_1 + x_2 \le -4$

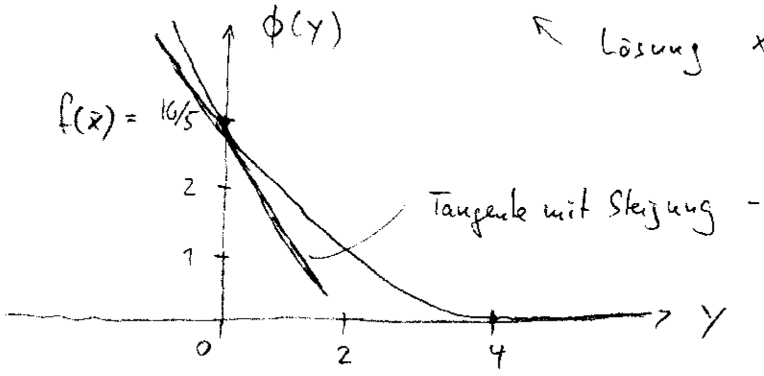


Minimum in $\bar{x} = \begin{pmatrix} -8/5 \\ -4/5 \end{pmatrix}$

Optimalwert ist $16/5$

Gestörtes Problem (Störfunktion) ist eindimensional

$$\Phi(y) = \min x_1^2 + x_2^2 \quad \text{unter} \quad 2x_1 + x_2 + 4 \le y$$



Lösung $x_1 = \frac{2y-8}{5}$ $x_2 = \frac{y-4}{5}$ für $y \le 4$

0 0 für $y \ge 4$

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

$\bar{\lambda} = -8/5$ definiert Sattelpunkt $(\bar{x}, \bar{\lambda})$

und $\Phi(y) \ge \Phi(0) - \bar{\lambda} y = 16/5 - 8/5\, y = $

Beweis: "$\Rightarrow$" $\qquad g(x) := \begin{pmatrix} g_1(x) \\ \vdots \\ g_m(x) \end{pmatrix}$

Sei $(\bar{x}, \bar{\lambda})$ ein Sattelpunkt $\Rightarrow \bar{\lambda} \ge 0$

$$f(x) + \bar{\lambda}^T g(x) \ge f(\bar{x}) + \bar{\lambda}^T g(\bar{x}) \quad \forall x \in S$$

$$\bar{\lambda} \cdot g(\bar{x}) = 0, \quad g(\bar{x}) \le 0$$

$\Rightarrow f(x) + \bar{\lambda}^T g(x) \ge f(\bar{x}) = \Phi(0)$ (da $\bar{x}$ min von f)

$$\Rightarrow f(x) \ge \underbrace{\Phi(0) - \bar{\lambda}^T g(x)}_{} \quad \forall x \in S$$

$\ge \Phi(0) - \bar{\lambda}^T y$ für jedes y, für das (NLO_y) eine Lösung hat

und für jedes $x \in X(y)$

(da $g_i(x) \le y_i$ nach Def von Φ und $\bar{\lambda} \ge 0$)

$$\Rightarrow \Phi(y) = \min_{x \in X(y)} f(x) \ge \Phi(0) - \lambda^T y$$

Falls (NLO_y) keine Lösung hat, so ist $\Phi(y)=\infty$ und

$\Phi(y) \geq \Phi(0) - \lambda^T y$ automatisch erfüllt

„$\Leftarrow$" Sei $\Phi(y) \geq \Phi(0) - \bar{\lambda}^T y \quad \forall y \in \mathbb{R}^m \qquad (*)$

(a) $\bar{\lambda} \geq 0$

Falls $\bar{\lambda}_i < 0$ so lasse $y_i \to \infty$ gehen und $y_j = 0$ sonst

$\Rightarrow \Phi(0) - \bar{\lambda}^T \cdot y \to \infty \overset{(*)}{\Rightarrow} \Phi(y) \to \infty$

$\Rightarrow$ Widerspruch zu $\Phi(y) \leq \Phi(0) \quad \forall y \geq 0$ (Antitonie)

(b) $L(\bar{x},\bar{\lambda}) = \min\limits_{x \in S} L(x,\bar{\lambda})$

$y = g(x) \Rightarrow f(x) \geq \Phi(y) = \Phi(g(x)) \qquad \forall x \in X(g(x)) \underset{\substack{\uparrow \\ g(x) \leq g(x) \text{ auf } \mathbb{R}^n}}{=} S$

$\Rightarrow f(x) \geq \Phi(g(x)) \overset{(*)}{\geq} \Phi(0) - \bar{\lambda}^T g(x) \qquad \forall x \in S \quad (**)$

Für $x = \bar{x}$ erhält man

$f(\bar{x}) + \bar{\lambda}^T g(\bar{x}) \geq \Phi(0) = f(\bar{x}) \qquad \Rightarrow \bar{\lambda}^T g(\bar{x}) \geq 0$

$\bar{\lambda} \geq 0, \; g(\bar{x}) \leq 0 \quad \Rightarrow \bar{\lambda}^T g(\bar{x}) = 0$

Nun gilt

$$\underbrace{f(x) + \bar{\lambda}^T g(x)}_{L(x,\bar{\lambda})} \overset{(**)}{\geq} f(\bar{x}) = \underbrace{f(\bar{x}) + \bar{\lambda}^T g(\bar{x})}_{L(\bar{x},\bar{\lambda})} \qquad \forall x \in S$$

$\Rightarrow L(\bar{x},\bar{\lambda}) = \min\limits_{x \in S} L(x,\bar{\lambda})$

(c) $(\bar{x},\bar{\lambda})$ ist Sattelpunkt

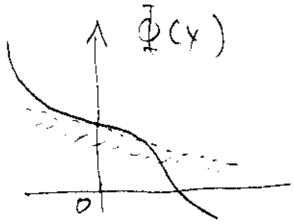
folgt aus (b), $\bar{\lambda}^T g(\bar{x}) = 0$ und Satz 2.5 □

Satz 2.7 kann auf Gleichheitsbedingungen $h_\ell(x) = 0$ verallgemeinert werden (zugehörige μ_ℓ sind dann nicht vorzeichenbeschränkt)

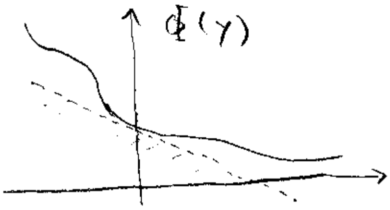
Obiges Beispiel war ein Fall mit Φ konvex

Oft ist (NLO) nicht konvex und Φ nicht konvex

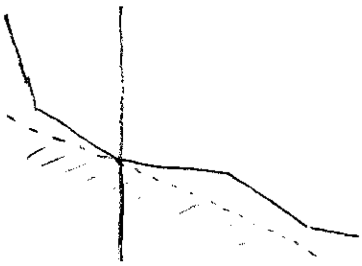
und $\nexists$ Stützhyperebene in $y=0$ $\Rightarrow$ kein Sattelpunkt



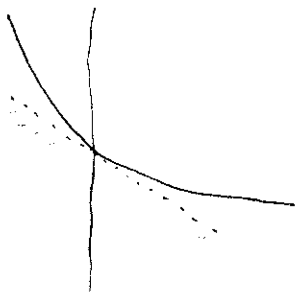
Φ kann nicht konvex sein, aber dennoch existiert eine Stützhyperebene in $y=0$ ($\Rightarrow$ Sattelpunkt $\Rightarrow$ Minimum)



Φ nicht diff.bar, nicht konvex, aber $\exists$ Sattelpunkt



Φ konvex $\Rightarrow$ $\exists$ Stützhyperebene $\Rightarrow$ $\exists$ Sattelpunkt



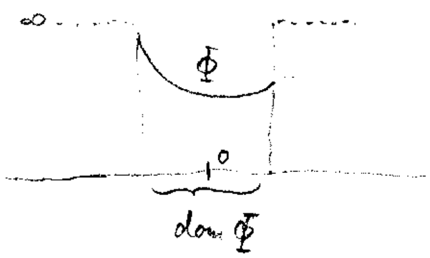
2.4 Sattelpunkte bei konvexen Optimierungsproblemen

2.8 SATZ f, g_i seien konvex, S sei konvex und es gebe $x \in S$ mit $g_i(x) < 0 \ \forall i$.
Hat dann (NLO) eine minimale Lösung $\bar{x}$, so existiert ein $\bar{\lambda} \geq 0$, so dass $(\bar{x}, \bar{\lambda})$ ein Sattelpunkt ist.
Also: Minima kommen von Sattelpunkten.

Beweis:

(1) $\Phi(y)$ ist konvex (2) $\mathrm{dom}(\Phi) := \{ y \mid \Phi(y) < \infty \} \neq \emptyset$

(3) $0 \in$ inneres$(\mathrm{dom}(\Phi))$



(1)-(3) $\Rightarrow$ $\Phi(y)$ hat einen Subgradienten γ in $y = 0$

$$\Rightarrow \Phi(y) \geq \Phi(0) + \gamma^T y \quad \forall y \in \mathbb{R}^m$$

[vgl. Def Subgradient aus ADM II für konkave Fkt
Subgradient in u = Vektor γ mit $f(v) - f(u) \leq \gamma^T (v-u) \ \forall v$
$\Rightarrow f(v) - f(u) \geq \gamma^T (v-u) \ \forall v$ für konvexe Fkt]

$\bar{\lambda} := -\gamma$ erfüllt dann die Bedingung für Sattelpunkt in Satz 2.7

denn $\Phi(y) \geq \Phi(0) - \bar{\lambda}^T y \quad \forall y \geq 0$

zu (1) - (3): $\boxed{\text{Aufgabe 4}}$

2.9 FOLGERUNG Im konvexen Falle gilt:
$(\bar{x}, \bar{\lambda})$ ist Sattelpunkt $\iff$ $-\bar{\lambda}$ ist Subgradient der Störfunktion Φ in $y = 0$

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

2.5 Kuhn Tucker Bedingungen für den konvexen differenzierbaren Fall

In diesem Fall erweisen sich die KT-Bedingungen als notwendig und hinreichend

> **2.10 SATZ** Betrachte
> $$\left.\begin{array}{l} \min f(x) \\ \text{unter } g_i(x) \leq 0 \quad i \in I \quad , \\ f, g_i \text{ konvex diffbar} \end{array}\right\} \text{NLP}$$
> mit $\exists\ x \in \mathbb{R}^n$ mit $g_i(x) < 0 \ \forall i$
>
> Dann gilt
> $\bar{x}$ ist globales Minimum von NLP
> $\Leftrightarrow$ die Kuhn Tucker Bedingungen gelten in $\bar{x}$,
> d.h. $\exists\ \bar{\lambda} \geq 0$ mit $\nabla_x L(\bar{x}, \bar{\lambda}) = 0$
> $$\bar{\lambda}_i g_i(\bar{x}) = 0 \quad \forall i$$

Beweis: Satz 2.7 sagt: $\bar{x}$ globales Minimum $\Leftrightarrow \exists\ \bar{\lambda} \geq 0$ und $(\bar{x}, \bar{\lambda})$ ist Sattelpunkt

$\Leftrightarrow$ (a) $\bar{x}$ minimiert $L(x, \bar{\lambda})$ auf S

(b) $g_i(\bar{x}) \leq 0 \quad \forall i$

(c) $\bar{\lambda}_i g_i(\bar{x}) = 0 \quad \forall i$

f, g_i konvex, diffbar $\Rightarrow L(x, \bar{\lambda})$ konvex, diffbar in x

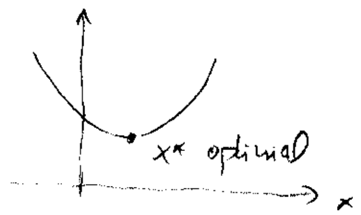
$\Rightarrow$ (a) ist äquivalent zu $\underbrace{\nabla_x L(\bar{x}, \bar{\lambda})}_{} = 0$

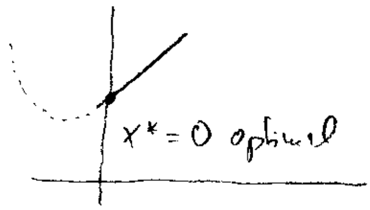
$$= \nabla f(\bar{x}) + \sum_i \bar{\lambda}_i \nabla g_i^T(\bar{x}) \qquad (*)$$

(*) und (c) entsprechen den KT Bedingungen $\square$

Berücksichtigung von Nichtnegativitätsbedingungen $x_j \geq 0$

Betrachte 1-dim Fall $\quad \min f(x) \quad x \geq 0$





$$\frac{df}{dx}(x^*) = 0 \qquad\qquad \frac{df}{dx}(x^*) \geq 0 \quad \text{und } x^* = 0$$

sowie $x^* \geq 0$

zusammengefasste notwendige Bedingungen:

$$x^* \cdot \frac{df}{dx}(x^*) = 0 \qquad \frac{df}{dx}(x^*) \geq 0 \qquad x^* \geq 0$$

Entsprechend im mehrdimensionalen Fall $f : \mathbb{R}^n \to \mathbb{R}^1$

$\min f(x)$ unter $x_j \geq 0$

wird entweder für $x^* > 0$ angenommen $\Rightarrow \nabla f(x^*) = 0$

oder am Rand mit $x_j^* = 0$ für ein oder mehrere j

$\Rightarrow$ notwendige Bedingungen sind

$$\boxed{x_j^* \frac{\partial f}{\partial x_j}(x^*) = 0 \qquad \frac{\partial f}{\partial x_j}(x^*) \geq 0 \qquad x_j^* \geq 0} \qquad (*)$$

exakte Herleitung von (*) aus der Kuhn-Tucker Bedingungen

mit $\quad g_j(x) = -x_j \leq 0$

$$(KT): \begin{cases} \nabla f(x^*) + \sum_j \lambda_j \nabla g_j(x^*) = 0 & (1) \\ \lambda_j \cdot g_j(x_j) = 0 \quad \forall j & (2) \\ \lambda_j \geq 0 \end{cases}$$

(1) $\Rightarrow \frac{\partial f}{\partial x_j}(x^*) + \lambda_j(-1) = 0 \quad \forall j \quad \Rightarrow \lambda_j = \frac{\partial f}{\partial x_j}(x^*)$

(1),(2) $\Rightarrow \quad x_j^* \cdot \frac{\partial f}{\partial x_j}(x^*) = 0 \quad \forall j$

$\lambda_j \geq 0 \quad \Rightarrow \frac{\partial f}{\partial x_j}(x^*) \geq 0 \quad \forall j$

Kuhn-Tucker anwendbar, da konvexe Nebenbedingungen mit inneren Punkt $\Rightarrow$ (CQ) gilt nach Lemma 2.2 $\square$

<u>Berücksichtigung von linearen Gleichheitsbedingungen und Nichtnegativität</u>

$$\min \; f(x) \quad \text{unter} \quad h_i(x) = -\sum_j h_{ij} x_j + b_i = 0$$
$$g_j(x) = -x_j \leq 0$$

(CQ) erfüllt, da Gleichheitsbed. linear und die $\geq$ Bedingungen konvex mit inneren Punkt

$\Rightarrow$ die Techniken aus Beweis von Lemma 2.2 können übertragen werden

$\Rightarrow$ (KT) Bedingungen sind notwendig:

(KT): $\begin{cases} \exists \cdot \text{Zahlen } \lambda_j \geq 0 \quad \mu_i \text{ bel. mit} \\ \nabla f(x^*) + \sum_j \lambda_j \nabla g_j(x^*) + \sum_i \mu_i \nabla h_i(x^*) = 0 & (1) \\ \lambda_j \cdot g_j(x^*) = 0 \quad \forall j & (2) \end{cases}$

(1) $\Rightarrow \frac{\partial f}{\partial x_j}(x^*) - \lambda_j + \sum_i \mu_i(-h_{ij}) = 0 \quad \forall j = 1, \dots, n$

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

λ_j in (2) eingesetzt ergibt

$$x_j \left(\frac{\partial f}{\partial x_j}(x^*) - \sum_i \mu_i h_{ij} \right) = 0 \qquad (2.1)$$

$\lambda_j \geq 0$ ergibt

$$\frac{\partial f}{\partial x_j}(x^*) - \sum_i \mu_i h_{ij} \geq 0 \qquad (2.2)$$

plus die Nebenbedingungen

$$\sum_j h_{ij} x_j^* = b_i \qquad (2.3)$$

$$x_j^* \geq 0 \qquad (2.4)$$

Also: notwendige Bedingungen für Optimalität von x^* ist die Existenz von Zahlen μ_i, $i = 1, \dots, m$ mit (2.1) - (2.4)

Bemerkung: in all diesen Fällen lässt sich über Sattelpunkte analog zu Satz 2.8 und (2.10) auch die Umkehrung zeigen (ohne Beweis)

2.6 Anwendung auf das Systemoptimum der Verkehrsplanung

$$\min\ C(x(f)) = \sum_{a\in A} x_a \cdot \tau_a(x_a) \qquad \text{mit} \quad x_a = \sum_{P\ni a} f_P$$

$$\text{unter} \quad \sum_{P\in \mathcal{P}_k} f_P = d_k \qquad k \in C$$

$$f_P \geq 0$$

(hier zur besseren Unterscheidung: f = Wegfluss

$x(f) = x$ = zugehöriger Kantenfluss

Kuhn-Tucker Bedingungen:

$$f_P \cdot \left(\frac{\partial C}{\partial f_P}(x(f)) - \mu_k \cdot 1 \right) = 0 \qquad \forall\, k \in C,\ P \in \mathcal{P}_k \qquad (1)$$

$$\frac{\partial C}{\partial f_P}\, x(f) - \mu_k \geq 0 \qquad \forall\, k \in C,\ P \in \mathcal{P}_k \qquad (2)$$

$$\sum_{P\in\mathcal{P}_k} f_P = d_k \qquad \forall\, k \qquad (3)$$

$$f_P \geq 0 \qquad \forall\, k,\ P \in \mathcal{P}_k \qquad (4)$$

Betrachte zunächst $\frac{\partial C}{\partial f_P}(x(f))$

$$= \frac{\partial}{\partial f_P} \sum_{a\in A} \underbrace{x_a(f) \cdot \tau_a(x_a(f))}_{=:\, c_a(x_a(f))}$$

$$= \sum_{a\in A} \frac{\partial c_a}{\partial x_a}(x_a) \frac{\partial x_a}{\partial f_P}(f) \qquad \text{Kettenregel}$$

$$= \sum_{a\in A} \frac{\partial c_a}{\partial x_a}(x_a) \cdot \delta_{a,P} \qquad \text{mit } \delta_{a,P} = \begin{cases} 1 & a \in P \\ 0 & \text{sonst} \end{cases}$$

$$\text{da} \quad x_a = \sum_{P' \ni a} f_{P'}$$

$$= \sum_{P \ni a} \frac{\partial C_a}{\partial x_a}(x_a) = \sum_{P \ni a} \underbrace{\left(x_a \cdot \frac{d}{dx_a} \tau_a(x_a) + \tau_a(x_a) \right)}_{= c'(x_a) \text{ aus § 1}}$$

"modifizierte Fahrzeitfkt",
"marginale Kosten"

$\Rightarrow$ (1) wird zu $f_P \cdot \left(\sum_{a \in P} c'(x_a) - \mu_k \right) = 0 \quad \forall k, P \in \mathcal{P}_k$

(2) wird zu $\sum_{a \in P} c'(x_a) \geq \mu_k \quad \forall k, P \in \mathcal{P}_k$

(1), (2) bedeuten

(1): Fluss führende Wege zu k haben dieselbe Länge μ_k bzgl. $c'(x_a)$

(2) jeder Weg zu k hat mindestens die Länge μ_k bzgl. $c'(x_a)$ als Fahrzeitfkt

$\Rightarrow$ SO = UE bzgl. $c'(x_a)$