

Grundkurs
"Mathematische Optimierung"
(für mathematische Studiengänge)

Prof. Dr. Beer

Literaturverzeichnis

- [1] Elster, Reinhardt, Schäuble, Donath: *Einführung in die nichtlineare Optimierung*, Teubner-Verlag, Leipzig 1977
- [2] D.G. Luenberger: *Linear and Nonlinear Programming*, Addison-Wesley Publ. Comp. 1989
- [3] J.F. Shapiro: *Mathematical Programming. Structures und Algorithms*, J. Wiley & Sons 1979
- [4] O.L. Mangasarian: *Nonlinear Programming*, SIAM-Publications 1994
- [5] J. Borwein, A.S. Lewis: *Convex Analysis and Nonlinear Optimization*, Springer-Verlag 2000
- [6] M.S. Bazaraa, H.D. Sherali, C.M. Shetty: *Nonlinear programming-theory and algorithms*, J. Wiley & Sons 1994

Inhaltsverzeichnis

1	Einige Klassen von Optimierungsaufgaben	4
1.1	Die allgemeine Aufgabenstellung	4
1.1.1	Das Minimierungsproblem	4
1.1.2	Einteilung von Optimierungsaufgaben in Klassen	5
1.1.3	Bemerkungen zur Geschichte der Optimierung	8
1.1.4	Der Aspekt: global - lokal	9
1.1.5	Zur Existenz optimaler Lösungen	11
1.2	Die Aufgabe des freien Optimums	11
1.3	Die klassische Aufgabe des restringierten Optimums	19
1.4	Die lineare Optimierungsaufgabe	25
1.4.1	Aufgabenstellung	25
1.4.2	Theoretische Grundlagen der Simplexmethode	30
1.4.3	Die primale modifizierte Simplexmethode	36
1.4.4	Die Phase I der SM	40
2	Grundlagen der konvexen Analysis (differenzierbarer Fall)	43
2.1	Konvexe Mengen	43
2.1.1	Der Begriff "konvexe Menge"	43
2.1.2	Algebraische Eigenschaften konvexer Mengen	45
2.1.3	Die lineare konvexe Hülle	47
2.1.4	Topologische Eigenschaften konvexer Mengen	51
2.2	Trennbarkeit konvexer Mengen	55
2.2.1	Die Projektion eines Punktes	56
2.2.2	Der Begriff der Trennung	57
2.2.3	Trennungssätze	59
2.2.4	Drei erste Anwendungen der Trennungssätze	63
2.3	Konvexe Funktionen	66
2.3.1	Der Begriff der konvexen Funktion	66
2.3.2	Erste Eigenschaften konvexer Funktionen	70
2.3.3	Differentielle Konvexitätskriterien	74
2.3.4	Ungleichungssysteme mit konvexen Funktionen	76

2.4	Konvexe Optimierungsaufgaben	77
3	Die allgemeine Aufgabe der Mathematischen Optimierung	81
3.1	Die Aufgabe (MOA)	81
3.2	Optimalitätskriterien für (MOA)	82
3.3	Regularität des zulässigen Bereichs	87
4	Dualitätstheorie	91
4.1	Die duale Aufgabe	91
4.2	Starke Dualität	96
4.3	Die Duale zu einer LOA	98
4.4	Optimalitäts- und Lösbarkeitskriterien für LOA	100

Kapitel 1

Einige Klassen von Optimierungsaufgaben

1.1 Die allgemeine Aufgabenstellung

1.1.1 Das Minimierungsproblem

Definition.

Wir sagen, daß eine **Optimierungsaufgabe** vorliegt, wenn 3 Dinge gegeben sind

$$\begin{cases} H - \text{Raum} \\ S \subseteq H - \text{Teilmenge des Raumes } H \\ f : S \rightarrow \mathbf{R} - \text{Funktion, die auf } S \text{ definiert ist} \end{cases}$$

und gesucht sind

$$(OA) \begin{cases} \text{eine Zahl } f^* \in \mathbf{R} \text{ sowie} \\ \text{ein Element } x^* \in S, \text{ so daß} \\ f^* = f(x^*) \leq f(x), \forall x \in S \end{cases}$$

Bemerkung.

1. Folgende Begriffe sind üblich:

f - Zielfunktion,

S - zulässiger Bereich, Restriktionsbereich,

f^* - Optimalwert,

x^* - optimale Lösung.

Ein $x \in S$ wird **zulässige Lösung** genannt.

2. Folgende Bezeichnungen sind für die obige Optimierungsaufgabe üblich:

$$f^* = \inf\{f(x) : x \in S\}$$

oder auch

$$(OA) \begin{cases} f(x) \rightarrow \min \\ x \in S \end{cases}$$

3. Wir vermerken, daß über die Natur von f und S nichts Näheres vorausgesetzt wird, wie etwa:

S - abgeschlossen, konvex, zusammenhängend

f - stetig, differenzierbar, konvex

4. Wir betrachten generell nur Minimierungsprobleme, da sich Maximierungsprobleme sofort auf Minimierungsprobleme zurückführen lassen und die Theorie der Minimierungsprobleme sich unmittelbar auf Maximierungsprobleme umschreiben läßt. Der Grund dafür ist in folgendem Fakt zu sehen: Sind ein $x^{**} \in S$ und ein $f^{**} \in \mathbf{R}$ gesucht, so daß

$$f^{**} = f(x^{**}) \geq f(x), \quad \forall x \in S, \quad (Max)$$

dann läßt sich über $F(x) := -f(x)$ das Minimierungsproblem

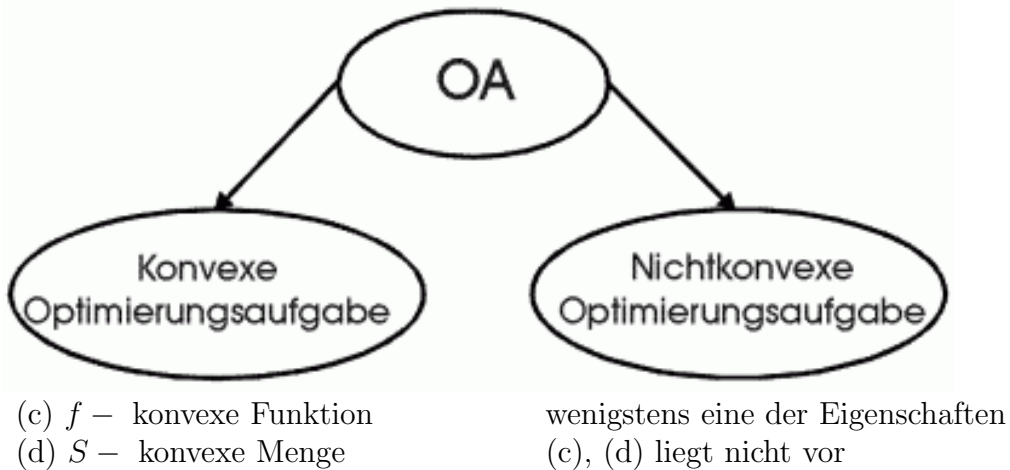
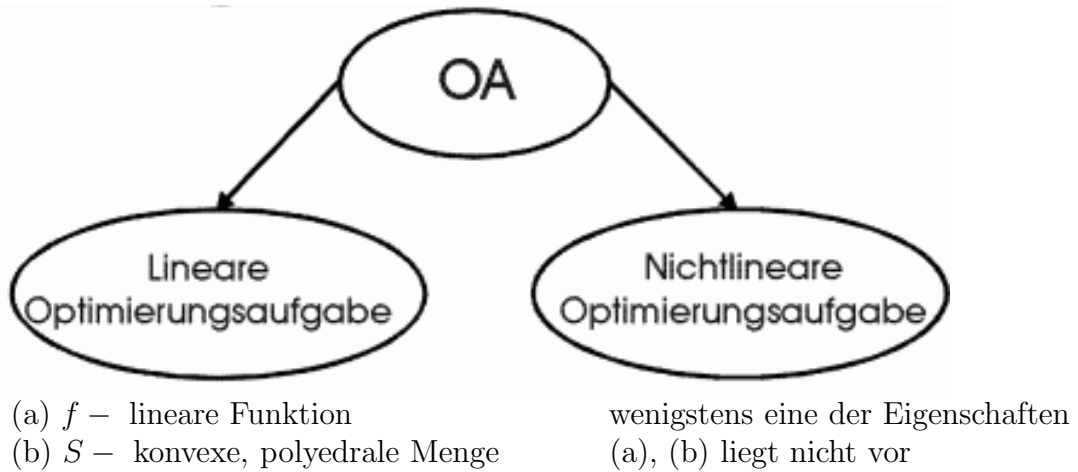
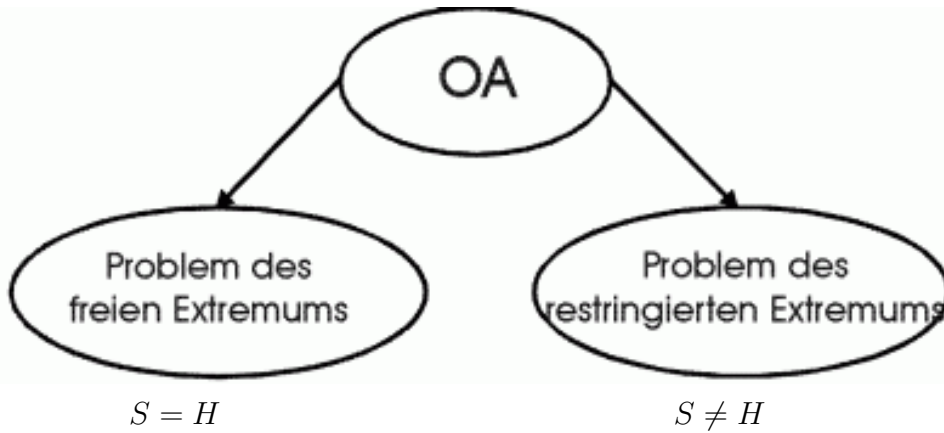
$$F^* = \inf\{F(x) : x \in S\} \quad (\widehat{OA})$$

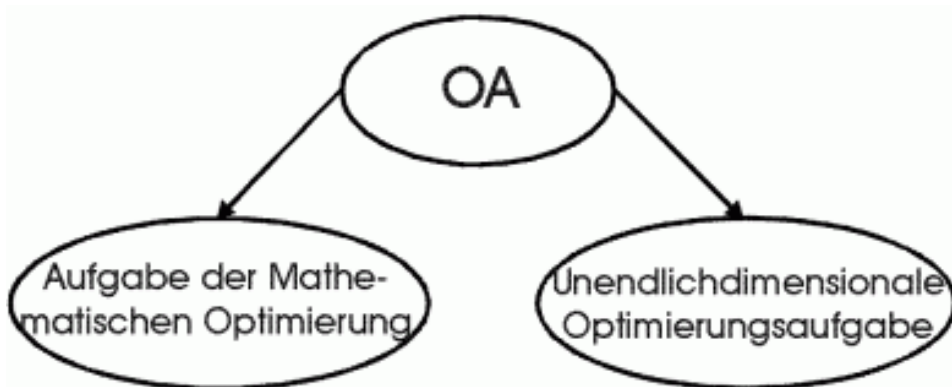
zuordnen und es gilt: $f^{**} = -F^*$ und x^{**} ist optimale Lösung von (\widehat{OA}) genau dann, wenn x^{**} das Problem (Max) löst.

5. Für die Menge $\{x \in S : f(x) = f^*\}$ der optimalen Lösungen von (OA) ist die Bezeichnung $Argmin\{f(x) : x \in S\}$ üblich. Für eine optimale Lösung von (OA) ist auch die Bezeichnung $argmin\{f(x) : x \in S\}$.

1.1.2 Einteilung von Optimierungsaufgaben in Klassen

Ausgehend von den Eigenschaften der 3 Bestandteile H, f, S einer Optimierungsaufgabe kann man diese klassifizieren.



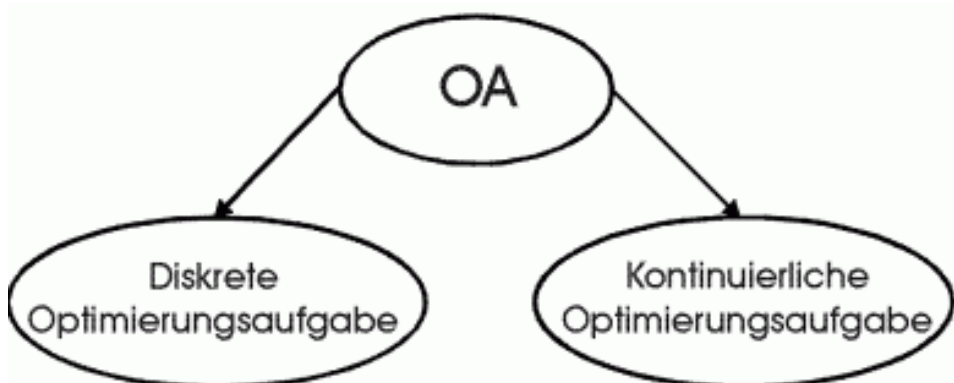


(e) $H = \mathbf{R}^n$

(f) S ist Lösungsmenge eines Systems von endlich vielen Gleichungen und Ungleichungen, d.h. $\exists g_i : \mathbf{R}^n \rightarrow \mathbf{R}$, $i = 1, 2, \dots, p + q$, so daß

$$S = \{x \in \mathbf{R}^n : g_i(x) \leq 0, i = 1, 2, \dots, p \\ \forall g_i(x) = 0, i = p + 1, \dots, p + q\}$$

wenigstens eine der Eigenschaften (e), (f) liegt nicht vor



(g) S oder eine Teilmenge von S besteht nur aus isolierten Punkten ($x^o \in S$ ist ein isolierter Punkt von S , wenn ein $\epsilon > 0$ existiert, so daß $S \cap U_\epsilon(x^o) = \{x^o\}$), in der Praxis ist die Diskretheit meistens dadurch erzeugt, daß gewisse der Koordinaten des Vektors $x \in \mathbf{R}^n$ nur ganzzahlige Werte oder die Werte 0 bzw. 1 annehmen dürfen.

explizit sind keine Diskretheitsforderungen erhoben

Wir behandeln in dieser Vorlesung weder diskrete noch unendlichdimensionale Optimierungsaufgaben. In der Regel werden wir die Differenzierbarkeit der vorkommenden Funktionen voraussetzen.

Abschnitt 1 der Vorlesung ist

- der Theorie der Aufgabe des freien Extremums (ohne numerische Verfahren),
- der Theorie der klassischen Extremwertaufgabe (ohne numerische Verfahren)
- den generellen Eigenschaften konvexer Optimierungsaufgaben (ohne Optimalitätskriterien)
- der Simplexmethode zur Lösung linearer Optimierungsaufgaben gewidmet.

Abschnitt 2 der Vorlesung beschäftigt sich mit den Bestandteilen einer Optimierungsaufgabe. Im wesentlichen geht es um Eigenschaften konvexer Mengen und konvexer Funktionen. Die hier angeschnittenen Fragen gehören in die sogenannte "Konvexe Analysis".

In Abschnitt 3 werden Optimalitätskriterien der Aufgabe der Mathematischen Optimierung behandelt.

Abschnitt 4 behandelt Beziehungen zwischen der Aufgabe der Mathematischen Optimierung und einer ihr zugeordneten sogenannten Lagrange-dualen Aufgabe. Diese Beziehungen werden abschließend näher für lineare Aufgaben analysiert.

1.1.3 Bemerkungen zur Geschichte der Optimierung

- 1947 - G.B. Dantzig beschreibt eine erste Variante der Simplexmethode für Zuordnungsprobleme
- 1948 - Fritz John publiziert ein Optimalitätskriterium für Optimierungsaufgaben, in denen Ungleichungen unter den Restriktionen erlaubt sind.
- 1950 - Kuhn-Tucker-Theorem
- 1950-1970: Ausarbeitung der linearen Optimierung und der Theorie der nichtlinearen Optimierung
- 1970-1980: Ausarbeitung der Theorie und Verfahren für nichtdifferenzierbare Optimierungsaufgaben und der Optimierung auf Graphen. Betrachtung von großen Aufgaben. Abrundung der Theorie der kontinuierlichen Aufgaben.

- 1980-1990: Komplexitätstheorie, insbesondere für diskrete Aufgaben. Wiederentdeckung der inneren-Punkt-Methoden.
- 1990-:....: Versuch, die Aufteilung in lineare und nichtlineare Probleme zu unterlaufen, indem semidefinite Aufgaben verstärkt untersucht werden.

Die Optimierungstheorie und ihre Verfahren entwickelten sich nicht zufällig erst mit dem Entstehen von elektronischen Rechnern. Es geht in ihr um $x \in \mathbf{R}^n$, wobei $n \geq 10$ bei nichtlinearen Aufgaben und $n \geq 10^6$ bei linearen Aufgaben typisch sind.

1.1.4 Der Aspekt: global - lokal

Oben hatten wir den Begriff der optimalen Lösung eingeführt. In der Regel wird diese Lösung gesucht. Häufig erweist es sich aber, daß das Aufsuchen dieser optimalen Lösung auf Schwierigkeiten stößt und man nur in der Lage ist, Punkte zu berechnen, unter denen die gesuchte optimale Lösung sein kann. Das führt zu dem Begriff der lokal-optimalen Lösung. Die oben bereits betrachtete optimale Lösung werden wir in diesem Zusammenhang als global-optimalen Lösung bezeichnen.

Definition.

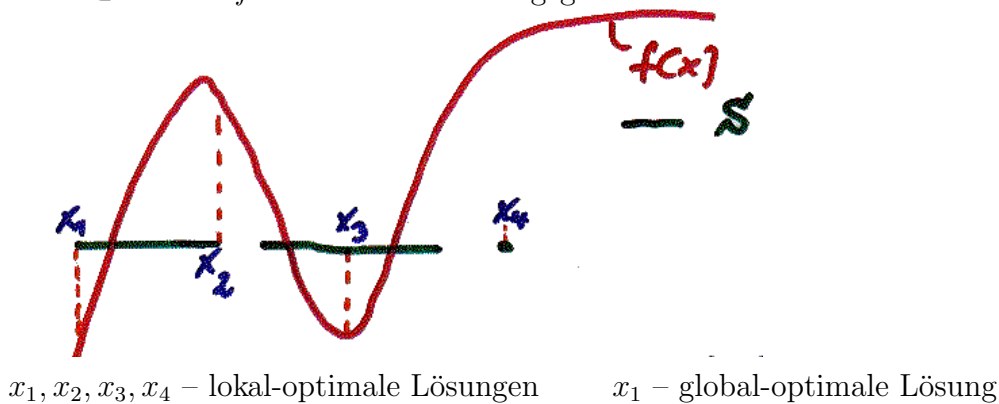
Gegeben sei die Aufgabe $f^* = \inf\{f(x) : x \in S\}$ (OA).

Der Punkt $x^* \in S$ heißt **global-optimale Lösung** von (OA), wenn $f(x^*) \leq f(x), \forall x \in S$.

Der Punkt $x^* \in S$ heißt **lokal-optimale Lösung** von (OA), wenn $\exists \epsilon > 0 : f(x^*) \leq f(x), \forall x \in S \cap U_\epsilon(x^*)$.

Beispiel. aus dem \mathbf{R}^1

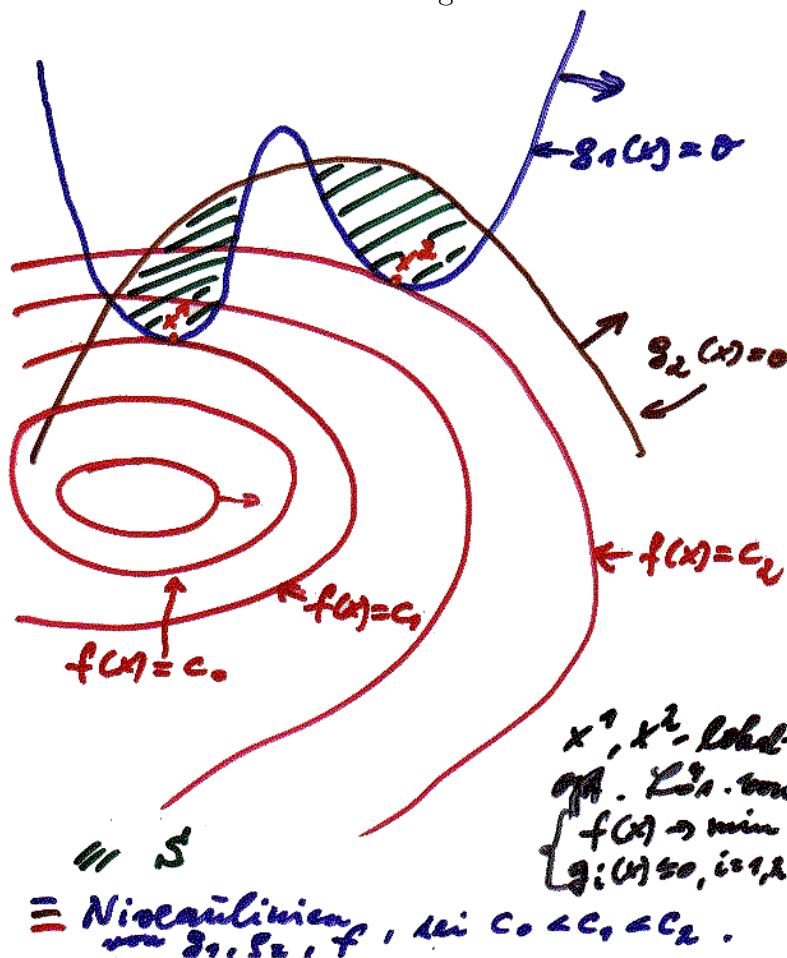
Sei $S \subset \mathbf{R}^1$ und $f : S \rightarrow \mathbf{R}$ wie unten gegeben.



Beispiel. aus dem \mathbf{R}^2

Sei $S = \{x \in \mathbf{R}^2 : g_1(x) \leq 0, g_2(x) \leq 0\}$, wobei $g_1, g_2 : \mathbf{R}^2 \rightarrow \mathbf{R}$. Sei $f : \mathbf{R}^2 \rightarrow \mathbf{R}$.

Wir tragen in die Ebene Niveaulinien von f , d.h. Punktmenge, die $f(x) = \text{const}$ erfüllen, sowie die Randkurven $g_1(x) = 0$ und $g_2(x) = 0$ von S (Niveaulinien von g_i zum Niveau Null) ein. Durch Pfeile deuten wir an, wo die Niveaulinien für höheres Niveau liegen.



Bemerkung.

1. Eine global-optimale Lösung ist stets auch lokal-optimale. Damit erscheint es sinnvoll, die global-optimale Lösung unter den lokal-optimale zu suchen. Wenn f^* aber in keinem Punkt aus S angenommen wird, so ist es möglich, daß die global-optimale Lösung nicht unter den lokal-optimale zu finden ist (vgl. hierzu Satz 1.5)

2. Das Dilemma der diskreten Optimierung wird u.a. durch folgende Aussage beschrieben: Wenn x^o ein isolierter Punkt von S ist, dann ist x^o offensichtlich lokal-optimal.

1.1.5 Zur Existenz optimaler Lösungen

Die Lösbarkeit einer Optimierungsaufgabe (d.h. die Existenz einer global-optimalen Lösung) nachzuweisen, werden wir stets mit im Auge haben. Das generell Mögliche wird bereits in der Analysis untersucht.

Satz (1.1). (K. Weierstraß)

Gegeben sei die OA

$$(OA) \quad f^* = \inf\{f(x) : x \in S\}$$

mit $S \subset \mathbf{R}^n$, $f : S \rightarrow \mathbf{R}$. Wenn

$$\begin{cases} S \neq \emptyset, \text{ beschränkt und abgeschlossen} \\ f \text{ stetig auf } S \end{cases}$$

dann besitzt (OA) eine global-optimale Lösung.

Satz (1.2).

Gegeben sei (OA) mit $S \subset \mathbf{R}^n$. Wenn

$$\begin{cases} S \text{ abgeschlossen} \\ f \text{ stetig auf } S \\ \exists x^o \in S, \text{ so daß } \{x \in S : f(x) \leq f(x^o)\} \text{ beschränkt ist} \end{cases}$$

dann besitzt (OA) eine global-optimale Lösung.

1.2 Die Aufgabe des freien Optimums

Definition.

Wenn in der Optimierungsaufgabe (OA) $S \equiv \mathbf{R}^n$, dann sagen wir, daß (OA) eine Aufgabe des **freien Optimums** ist.

Bemerkung.

1. Wir bezeichnen

$$(FO) \begin{cases} f(x) \rightarrow \min \\ x \in \mathbf{R}^n \end{cases}$$

Dabei wird vorausgesetzt, daß $f : \mathbf{R}^n \rightarrow \mathbf{R}$.

2. Viele andere OA lassen sich unter Umständen auf diesen Typ zurückführen:

- über einen Strafansatz (vgl. Beweis zu Satz 1.7)
- über Variablenelimination
- über die Lagrangefunktion
- wenn S eine offene Menge ist (vgl. Satz 1.6)

3. Wir wollen wiederholen, was wir unter Differenzierbarkeit verstehen wollen. Sei f in einer Umgebung von x^o definiert. Wir sagen, daß f in x^o **differenzierbar** ist (im Sinne von Frechet), wenn ein Vektor $\nabla f(x^o) \in \mathbf{R}^n$ existiert, so daß

$$\lim_{x \rightarrow x^o} \frac{f(x) - f(x^o) - \langle \nabla f(x^o), x - x^o \rangle}{\|x - x^o\|} = 0$$

$\nabla f(x^o)$ bezeichnen wir als **Gradienten** von f in x^o .

4. Es ist bekannt, wenn $\nabla f(x^o)$ existiert, dann besteht er aus den partiellen Ableitungen:

$$\nabla f(x^o) = \left[\frac{\partial f(x^o)}{\partial x_1}, \frac{\partial f(x^o)}{\partial x_2}, \dots, \frac{\partial f(x^o)}{\partial x_n} \right]$$

5. Wir sagen, daß f in x^o **zweimal differenzierbar** ist, wenn es einmal in x^o differenzierbar ist und eine Matrix $\nabla^2 f(x) \in \mathbf{R}^{n,n}$ existiert, so daß

$$\lim_{x \rightarrow x^o} \frac{f(x) - f(x^o) - \langle \nabla f(x^o), x - x^o \rangle - \frac{1}{2} \langle x - x^o, \nabla^2 f(x^o)(x - x^o) \rangle}{\|x - x^o\|^2} = 0$$

$\nabla^2 f(x^o)$ wird **Hessematrix** von f in x^o genannt.

6. Es ist bekannt, wenn $\nabla^2 f(x^o)$ existiert, dann besteht es aus den zweiten partiellen Ableitungen von f in x^o :

$$\nabla^2 f(x^o) = \begin{pmatrix} \frac{\partial^2 f(x^o)}{\partial x_1^2} & \frac{\partial^2 f(x^o)}{\partial x_2 \partial x_1} & \cdots & \frac{\partial^2 f(x^o)}{\partial x_n \partial x_1} \\ \frac{\partial^2 f(x^o)}{\partial x_1 \partial x_2} & \frac{\partial^2 f(x^o)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x^o)}{\partial x_n \partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x^o)}{\partial x_1 \partial x_n} & \frac{\partial^2 f(x^o)}{\partial x_2 \partial x_n} & \cdots & \frac{\partial^2 f(x^o)}{\partial x_n^2} \end{pmatrix}$$

Satz (1.3). (Notwendiges Optimalitätskriterium)

Sei $f : \mathbf{R}^n \rightarrow \mathbf{R}$. Sei $x^* \in \mathbf{R}^n$ eine lokal-optimale Lösung von (FO). Dann gilt:

1. Wenn f in x^* differenzierbar ist, dann ist $\nabla f(x^*) = 0$.
2. Wenn f in x^* zweimal differenzierbar ist, dann ist $\nabla^2 f(x^*)$ eine positiv semidefinite Matrix.

Bemerkung.

1. Worin besteht der Wert von Satz 1.3 ?
 - (a) Im differenzierbaren Fall müssen die lokal-optimalen Lösungen unter den Lösungen des Systems

$$\nabla f(x) = 0 \quad (+)$$

sein. Mit anderen Worten: Jede Lösung des Systems (+) von n Gleichungen mit n Unbekannten ist verdächtig, lokal-optimale Lösung von (FO) zu sein. Das eröffnet einen Weg, akademische Beispiele von (FO) zu lösen. Es muß gesagt werden, daß die numerische Praxis diesen Weg i.a. nicht geht. Anstelle das i.a. nicht-lineare Gleichungssystem (+) zu lösen, wendet man direkte Verfahren auf (FO) an.

- (b) Jeder Punkt $\hat{x} \in \mathbf{R}^n$, der wenigstens eine der Bedingungen $\nabla f(\hat{x}) = 0$ (im einmal differenzierbaren Fall) bzw. $\nabla^2 f(\hat{x})$ positiv semidefinit (im zweimal differenzierbaren Fall) nicht erfüllt, kann nicht lokal-optimal sein.
2. Wie können wir testen, ob eine quadratische symmetrische Matrix A positiv semidefinit ist? A ist positiv semidefinit \iff eine der Bedingungen (a), (b) ist erfüllt
 - (a) Alle Eigenwerte von A sind nichtnegativ.
 - (b) Alle Minoren entlang der Hauptdiagonalen von A sind nichtnegativ (d.h. alle einreihigen, alle zweireihigen, usw. entlang der Hauptdiagonalen, nicht nur die Hauptminoren).

Später (vgl. Satz 2.30) werden wir sehen, dass bei konvexen, 2-mal differenzierbaren Funktionen f stets $\nabla^2 f(x)$ positiv semidefinit ist.

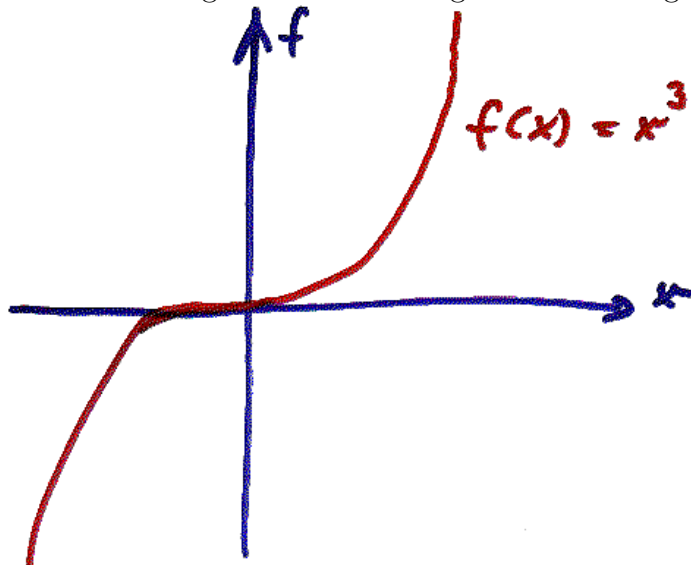
3. Oben hatten wir behauptet, daß in der Praxis wir kaum Möglichkeiten besitzen, die global-optimale Lösung zu berechnen, aber häufig in der Lage sind (d.h. Möglichkeiten besitzen, z.B. durch Auswerten notwendiger Optimalitätsbedingungen oder Anwenden von Minimierungsverfahren) lokal-optimale Lösungen zu berechnen. Exakterweise müßten wir anstelle von lokal-optimal in diesem Zusammenhang besser von stationär sprechen.

Definition.

Wenn $\nabla f(x^*) = 0$, dann sagen wir, daß x^* ein **stationärer Punkt** von f ist.

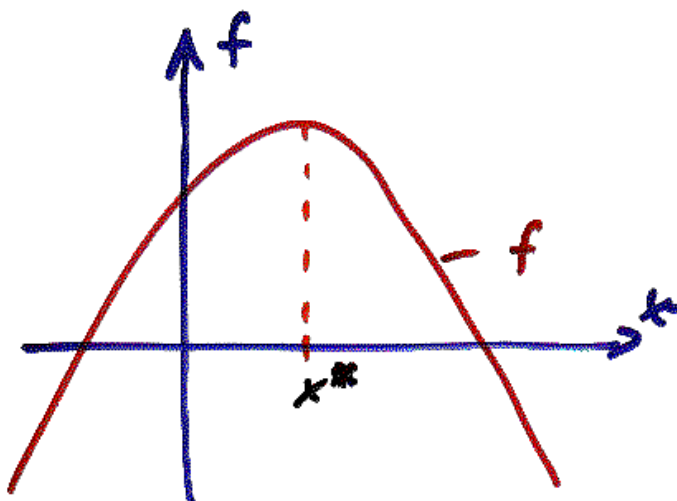
Satz 1.3 sagt also, daß im differenzierbaren Fall lokal-optimale Lösungen von (FO) stationäre Punkte von f sind.

Offensichtlich gilt die Umkehrung nicht unbedingt:



$x^* = 0$ ist stationärer Punkt, aber keine lokal-optimale Lösung von

$$\begin{cases} f(x) \rightarrow \min \\ x \in \mathbf{R}^1 \end{cases}$$



x^* ist stationärer Punkt, aber keine lokal-optimale Lösung von

$$\begin{cases} f(x) \rightarrow \min \\ x \in \mathbf{R}^1 \end{cases}$$

Wir wollen der Frage nachgehen, bei welchen Funktionen das Kriterium $\nabla f(x^*) = 0$ nicht nur notwendig, sondern auch hinreichend ist.

Definition.

Sei $S \subset \mathbf{R}^n$, $f : S \rightarrow \mathbf{R}$, $\bar{x} \in S$, f in \bar{x} differenzierbar.

Wir sagen, daß f in \bar{x} bzgl. S **pseudokonvex** ist, wenn aus

$$\begin{cases} \langle \nabla f(\bar{x}), x - \bar{x} \rangle \geq 0 \\ x \in S \end{cases}$$

stets folgt, daß $f(x) \geq f(\bar{x})$.

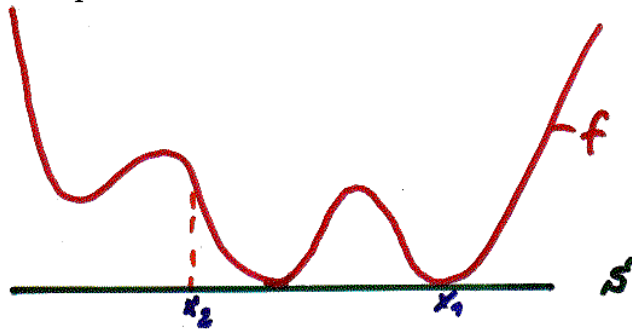
Wir sagen, daß f in \bar{x} bzgl. S **streng pseudokonvex** ist, wenn aus

$$\begin{cases} \langle \nabla f(\bar{x}), x - \bar{x} \rangle \geq 0 \\ x \in S, x \neq \bar{x} \end{cases}$$

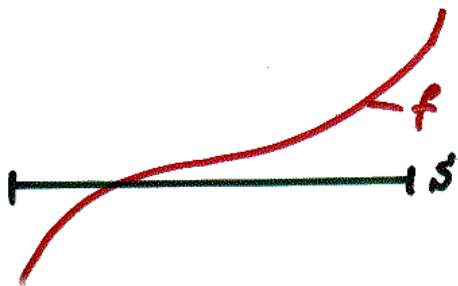
stets folgt, daß $f(x) > f(\bar{x})$.

Wir sagen, daß f auf S pseudokonvex (bzw. streng pseudokonvex) ist, wenn f in jedem Punkt $\bar{x} \in S$ bzgl. S pseudokonvex (bzw. streng pseudokonvex) ist.

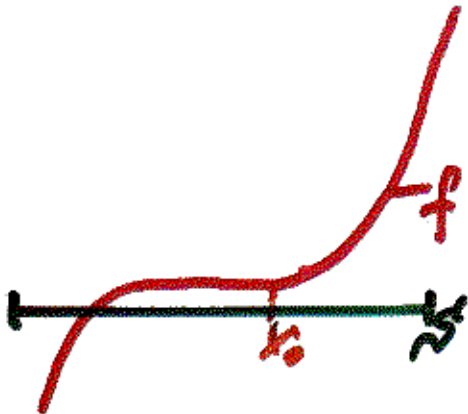
Beispiel.



- f in x_1 pseudokonvex bzgl. S , aber nicht streng pseudokonvex
- f in x_2 streng pseudokonvex bzgl. S
- f auf S nicht pseudokonvex

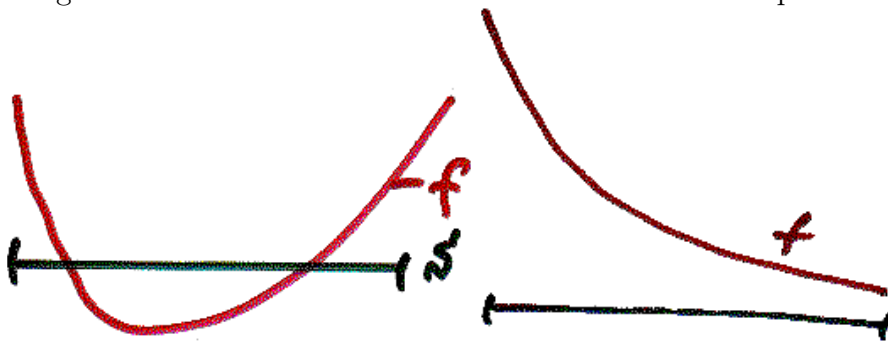


- f auf S pseudokonvex



- f auf S nicht pseudokonvex (x_0 ist horizontaler Wendepunkt von f)

Vorgriff: Differenzierbare konvexe Funktionen sind stets pseudokonvex.



Satz (1.4.1). (Notw. und hinreich. Opt.-kriterium erster Ordnung)
 Sei $f : \mathbf{R}^n \rightarrow \mathbf{R}$ pseudokonvex in x^* bzgl. \mathbf{R}^n . Dann ist x^* global-optimale Lösung von (FO) genau dann, wenn $\nabla f(x^*) = 0$.

Satz (1.4.2). (Hinreich. Opt.-kriterium zweiter Ordnung)

Sei $f : \mathbf{R}^n \rightarrow \mathbf{R}$. In $x^* \in \mathbf{R}^n$ sei f zweimal differenzierbar. Dann gilt: Wenn

$$\begin{cases} \nabla f(x^*) = 0 \\ \nabla^2 f(x^*) \text{ positiv definit,} \end{cases}$$

dann ist x^* eine lokal-optimale Lösung von (FO) .

Bemerkung.

1. Aus dem Beweis folgt, daß unter den Voraussetzungen des Satzes x^* sogar streng lokal-optimale Lösung von (FO) ist (d.h. $\exists \epsilon > 0$, so daß $f(x) > f(x^*), \forall x \in U_\epsilon(x^*) \setminus \{x^*\}$).
2. Wie können wir eine quadratische Matrix A auf positive Definitheit überprüfen?
 - alle Eigenwerte von A sind positiv
 - alle Hauptminoren von A sind positiv
 - in der Cholesky-Zerlegung $A = L^\top L$ (L - obere Dreiecksmatrix) sind alle Diagonalelemente von L positiv
3. Später werden wir sehen, daß die Forderung $\nabla^2 f(x^*)$ positiv definit bedeutet, daß f in einer Umgebung von x^* streng konvex ist.

Wir wollen noch einen Satz betrachten, der etwas über die Existenz einer global-optimale Lösung in (FO) aussagt.

Definition.

Sei

$$\begin{cases} f : \mathbf{R}^n \rightarrow \mathbf{R} \\ f \text{ stetig in } \mathbf{R}^n. \end{cases}$$

Wir sagen, daß f **koerzibel** (=bezwingbar) ist, wenn zu jeder Zahl $M \in \mathbf{R}$ eine Zahl $r_M \in \mathbf{R}$ existiert, so daß aus $\|x\| \geq r_M$ stets folgt, daß $f(x) \geq M$.

Bemerkung.

Bei koerziblen Funktionen ist also

$$\lim_{\|x\| \rightarrow \infty} f(x) = +\infty.$$

Satz (1.5).

Wir betrachten die Aufgabe (FO) mit koerziblem f . Dann gilt:

1. (FO) besitzt eine global-optimale Lösung.

2. Wenn f zusätzlich in \mathbf{R}^n differenzierbar ist, dann ist die global-optimale Lösung von (FO) unter den stationären Punkten von f zu finden.

Folgerung.

Bei differenzierbaren, koerziblen f kann man eine global-optimale Lösung von (FO) finden, indem man alle Lösungen $x^1, x^2, x^3, \dots, x^k$ von $\nabla f(x) = 0$ berechnet und dann einen Punkt $x^{i_o}, i_o \in \{1, 2, \dots, k\}$ aussucht, für den $\min_{1 \leq i \leq k} f(x^i)$ realisiert wird (gilt auch, wenn $k = \infty$ und das Minimum existiert).

Abschließend zu (FO) wollen wir die Frage untersuchen, wann die allgemeine Aufgabe der Optimierung als Aufgabe des freien Optimums behandelt werden kann.

Satz (1.6).

Wir betrachten die beiden Optimierungsaufgaben

$$(OA) \begin{cases} f(x) \rightarrow \min \\ x \in S \end{cases} \quad \text{und} \quad (FO) \begin{cases} f(x) \rightarrow \min \\ x \in \mathbf{R}^n \end{cases}$$

wobei $S \subset \mathbf{R}^n, f : \mathbf{R}^n \rightarrow \mathbf{R}$. Es sei $x^* \in \text{int}(S)$ (d.h. $\exists \epsilon > 0 : U_\epsilon(x^*) \subset S$).

Dann gilt:

x^* ist lokal-optimale Lösung von $(OA) \iff x^*$ ist lokal-optimale Lösung von (FO) .

Bemerkung.

1. Unter den Voraussetzungen des Satzes (speziell $x^* \in \text{int}(S)$) sind die notwendigen und hinreichenden Optimalitätskriterien für x^* in (OA) die gleichen wie in (FO) . (OA) ist deshalb erst eine wesentlich andere Aufgabe als (FO) , wenn die lokal-optimalen Lösungen von (OA) auf dem Rand von S liegen (d.h., $\forall \epsilon > 0 : U_\epsilon(x^*) \not\subset S$).
2. Die Aussage des Satzes läßt sich verallgemeinern: Es seien $S, T \subset \mathbf{R}^n, x^* \in S, x^* \in \text{int}(T)$: Wir betrachten 2 Aufgaben

$$(OA_1) \begin{cases} f(x) \rightarrow \min \\ x \in S \cap T \end{cases} \quad \text{anstelle von (OA) und}$$

$$(OA_2) \begin{cases} f(x) \rightarrow \min \\ x \in S \end{cases} \quad \text{anstelle von (FO).}$$

Dann gilt: x^* ist lokal-optimal in $(OA_1) \iff x^*$ ist lokal-optimal in (OA_2) .

1.3 Die klassische Aufgabe des restringierten Optimums

In der Analysis wird unsere allgemeine Optimierungsaufgabe (OA) für den Fall untersucht, daß S durch ein endliches System von Gleichungen definiert ist.

Es seien $g_i : \mathbf{R}^n \rightarrow \mathbf{R}, i = 1, \dots, m$ gegeben. Wir setzen $S = \{x \in \mathbf{R}^n : g_i(x) = 0, i = 1, \dots, m\}$. (o)

Im weiteren soll nur der Fall betrachtet werden, daß die g_i differenzierbar sind. Wenn (OA) mit einem nach (o) definierten S vorliegt, wollen wir von der Aufgabe ($OAGR$) sprechen.

$$(OAGR) \begin{cases} f(x) \rightarrow \min \\ g_1(x) = 0 \\ g_2(x) = 0 \\ \vdots \\ g_m(x) = 0 \end{cases}$$

Bevor wir das notwendige Optimalitätskriterium formulieren, wollen wir einige Bezeichnungen festlegen.

Definition.

Für $\lambda_o \in \mathbf{R}, \lambda = [\lambda_1, \lambda_2, \dots, \lambda_m] \in \mathbf{R}^m$ bezeichnen wir mit $L(x, \lambda_o, \lambda)$ die sogenannte **Lagrangefunktion** zu ($OAGR$): $L(x, \lambda_o, \lambda) = \lambda_o f(x) + \sum_{i=1}^m \lambda_i g_i(x)$

Mit $\nabla_x L(x, \lambda_o, \lambda) = \lambda_o \nabla_x f(x) + \sum_{i=1}^m \lambda_i \nabla_x g_i(x)$

und $\nabla_{xx}^2 L(x, \lambda_o, \lambda) = \lambda_o \nabla_{xx}^2 f(x) + \sum_{i=1}^m \lambda_i \nabla_{xx}^2 g_i(x)$

bezeichnen wir den x -**Gradienten** bzw. die x -**Hessematrix** der Lagrange-funktion.

Definition.

Sei x^* eine zulässige Lösung von ($OAGR$), die g_i seien in x^* differenzierbar. Dann wird der Unterraum

$$T(x^*) = \{z \in \mathbf{R}^n : \langle \nabla g_i(x^*), z \rangle = 0, i = 1, 2, \dots, m\}$$

des \mathbf{R}^n als **Tangentialraum** zu x^* bezeichnet.

Bemerkung.

$T(x^*)$ ist die Lösungsmenge eines homogenen linearen Gleichungssystems und hat die Dimension $n - \text{rank}\{\nabla g_i(x^*)\}_{i=1}^m$.

Satz (1.7). (Notwendiges Optimalitätskriterium)

Es sei x^* lokal-optimale Lösung von (OAGR). Dann gilt:

1. Wenn $f, g_i, i = 1, 2, \dots, m$ in einer Umgebung von x^* stetig differenzierbar sind, dann existieren $\lambda_o^* \geq 0$ und $\lambda_1^*, \lambda_2^*, \dots, \lambda_m^* \in \mathbf{R}$, so daß

$$(+)\begin{cases} \sum_{i=0}^m (\lambda_i^*)^2 > 0 \\ \nabla_x L(x^*, \lambda_o^*, \lambda^*) = 0 \end{cases}$$

2. Wenn zusätzlich zur Voraussetzung von (1) noch gilt, daß $f, g_i, i = 1, 2, \dots, m$ in x^* zweimal differenzierbar sind, dann gilt für beliebige $\lambda_o^* \lambda_1^*, \dots, \lambda_m^*$, die (+) erfüllen, daß auch

$$\langle z, \nabla_{xx}^2 L(x^*, \lambda_o^*, \lambda^*) z \rangle \geq 0, \forall z \in T(x^*) \quad (++)$$

sein muß (d.h. der x -Hessematrix der Lagrangefunktion muss auf dem zu x^* gehörenden Tangentialraum positiv semidefinit sein).

Bemerkung.

1. Geometrische Interpretation: Wenn $\lambda_o^* \neq 0$, dann folgt aus (+) $\nabla f(x^*) = -\sum_{i=1}^m \lambda_i^* / \lambda_o^* \nabla g_i(x^*)$, d.h. lokal-optimale Lösungen von (OAGR) zeichnen sich dadurch aus, daß in ihnen der Gradient der Zielfunktion in dem durch die Gradienten der Nebenbedingungen aufgespannten Unterraum liegt.

2. Später werden wir uns besonders dafür interessieren, unter welchen Bedingungen garantiert werden kann, daß ein $(m+1)$ -Tupel $\lambda_o^*, \dots, \lambda_m^*$ mit $\lambda_o^* \neq 0$ existiert, das (+) erfüllt (sogenannte Regularität des zulässigen Bereichs).

Im Analysis-Kurs wird gewöhnlich in Verbindung mit Satz 1.7 noch die Forderung $\{\nabla g_i(x^*)\}_{i=1}^m$ ist eine linear unabhängige Menge (sog. (LICQ)) gestellt. Unter dieser Zusatzforderung folgt aus (+), daß $\lambda_o^* \neq 0$, da bei linearer Unabhängigkeit der g_i -Gradienten und $\lambda_o^* = 0$ die zweite Zeile von (+) nur bei $\lambda_1^* = \dots = \lambda_m^* = 0$ möglich ist, was der ersten Zeile von (+) widerspricht.

3. Die Zahlen $\lambda_o^*, \lambda_1^*, \dots, \lambda_m^*$, die (+) erfüllen, heißen Lagrangesche Multiplikatoren zu lokal-optimale Lösung x^* .

4. Da die λ in die Lagrangefunktion homogen eingehen, braucht man eigentlich nur 2 Möglichkeiten für λ_o^* betrachten: $\lambda_o^* = 0$ und $\lambda_o^* = 1$. (bei $\lambda_o^* \neq 0$ führt $\lambda_i^* := \lambda_i^*/\lambda_o^*, i = 0, 1, 2, \dots, m$, in (+) zu $\lambda_o^* = 1$).
5. Wenn (OAGR) die Aufgabe (FO) ist (d.h. $m = 0$), dann ergeben die Aussagen von Satz 1.7 genau die Aussagen von Satz 1.3 (da $T(x^*) = \mathbf{R}^n, \lambda_o^* \neq 0$).
6. Der Beweis gibt Hinweise, wie man (OAGR) lösen kann (zugehöriges (FO) lösen) und wie man die Lagrangemultiplikatoren dabei berechnen kann.

Definition.

Der Punkt $x^* \in \mathbf{R}^n$ heißt **stationärer Punkt** von Aufgabe (OAGR), wenn Zahlen $\lambda_o^* \geq 0, \lambda_1^*, \dots, \lambda_m^*$ mit $\sum_{i=0}^m (\lambda_i^*)^2 > 0$ existieren, so daß gilt

$$(\circ) \begin{cases} \nabla_x L(x^*, \lambda_o^*, \lambda^*) = 0 \\ g_i(x^*) = 0, i = 1, 2, \dots, m \end{cases}$$

Bemerkung.

1. Satz 1.7 besagt u.a.: Im stetig-differenzierbaren Fall sind die lokal-optimalen Punkte unter den stationären Punkten zu finden.
2. (o) ist ein System aus $(m + n)$ (i.a. nichtlinearen) Gleichungen mit $(m + n + 1)$ Unbekannten $x_1^*, \dots, x_n^*, \lambda_o^*, \dots, \lambda_m^*$.
3. Bei der Suche nach Punkten, die verdächtig sind, lokal-optimale Lösungen von (OAGR) zu sein, muß man (aus akademischer Sicht) folgende 2 Gleichungssysteme aus je $(m + n)$ Gleichungen mit je $(m + n)$ Unbekannten analysieren:

$$(\text{Fall } \lambda_o = 0) \quad (\Delta) \begin{cases} \sum_{i=1}^m \lambda_i \nabla g_i(x) = 0 \\ g_i(x) = 0, \quad i = 1, 2, \dots, m \\ \sum_{i=1}^m (\lambda_i)^2 > 0 \end{cases}$$

$$(\text{Fall } \lambda_o = 1) \quad (\Delta\Delta) \begin{cases} \nabla f(x) + \sum_{i=1}^m \lambda_i \nabla g_i(x) = 0 \\ g_i(x) = 0, \quad i = 1, 2, \dots, m \end{cases}$$

4. Die Numerik der Optimierung geht nicht den Weg des Lösens von (Δ) und $(\Delta\Delta)$, sondern liefert Iterationsverfahren, die $(OAGR)$ direkt angehen. Satz 1.7 wird in der Theorie zu diesen Verfahren benutzt.

Satz (1.8.1). (Hinreichendes Optimalitätskriterium zweiter Ordnung für ein lokales Optimum)

Sei x^* stationärer Punkt von $(OAGR)$. Seien f, g_1, g_2, \dots, g_m in x^* zweimal differenzierbar. Dann gilt:

Wenn $\nabla_{xx}^2 L(x^*, \lambda_o^*, \lambda^*)$ auf dem Tangentialraum positiv definit ist, dann ist x^* lokal-optimale Lösung von $(OAGR)$.

Genauer: Wenn $[x^*, \lambda_o^*, \lambda^*]$ Lösung von (o) ist und

$$\langle z, \nabla_{xx}^2 L(x^*, \lambda_o^*, \lambda^*) z \rangle > 0, \quad \forall z \in T(x^*) \setminus \{0\}, \quad (\square)$$

dann existiert ein $\epsilon > 0$, so daß

$$\begin{aligned} f(x^*) < f(x), \quad \forall x \in U_\epsilon(x^*) \cap (S \setminus \{x^*\}) \\ (S = \{x \in \mathbf{R}^n : g_i(x) = 0, i = 1, 2, \dots, m\}). \end{aligned}$$

Bemerkung.

1. In einigen (akademischen) Beispielen kann man die allgemeine Lösung des $T(x^*)$ definierenden Gleichungssystems gewinnen und mit ihrer Hilfe (\square) überprüfen.
2. Interessant sind 2 Grenzfälle:
 - (a) $T(x^*) = \{0\}$ (d.h. $m = n, \text{rank}\{\nabla g_i(x^*)\}_{i=1}^m = m$) Dann ist (\square) erfüllt.
 - (b) $\nabla_{xx}^2 L(x^*, \lambda_o^*, \lambda^*)$ ist eine positiv definite Matrix. Dann ist (\square) ebenfalls erfüllt.
3. Im Falle $\lambda_o^* \neq 0$ gilt (\square) genau dann, wenn von den $(n + m)$ Hauptminoren der Matrix

$$\begin{pmatrix} 0 & \nabla g(x^*) \\ (\nabla g(x^*))^\top & \nabla_{xx}^2 L(x^*, \lambda_o^*, \lambda^*) \end{pmatrix}$$

die letzten $(n - m)$ (d.h vom $(2m + 1)$ -ten bis zum $(m + n)$ -ten) alle das Vorzeichen $(-1)^m$ haben. ($\nabla g(x^*)$ sei hier die Matrix, deren Zeilen die $\nabla g_i(x^*)$ bilden.)

4. Satz 1.8 ist ein Spezialfall von Theorem 3.6 (das wir in der Vorlesung beweisen).

Zum Abschluß wollen wir noch ein hinreichendes Kriterium für ein globales Optimum aufstellen.

Definition.

Sei $S \subset \mathbf{R}^n, f : S \rightarrow \mathbf{R}, \bar{x} \in S$. Wir sagen, daß f in \bar{x} bzgl. S **quasikonvex** ist, wenn aus

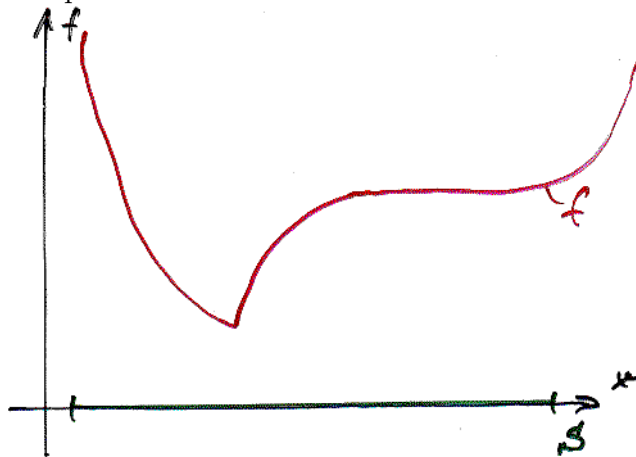
$$\begin{cases} x \in S \\ f(x) \leq f(\bar{x}) \\ 0 \leq \lambda \leq 1 \\ (\lambda x + (1 - \lambda)\bar{x}) \in S \end{cases} \quad \text{immer folgt, daß } f(\lambda x + (1 - \lambda)\bar{x}) \leq f(\bar{x})$$

Wir sagen, daß f auf S quasikonvex ist, wenn es in jedem Punkt von S quasikonvex ist. Wir sprechen von **Quasikonkavität**, wenn $(-f)$ quasikonvex ist.

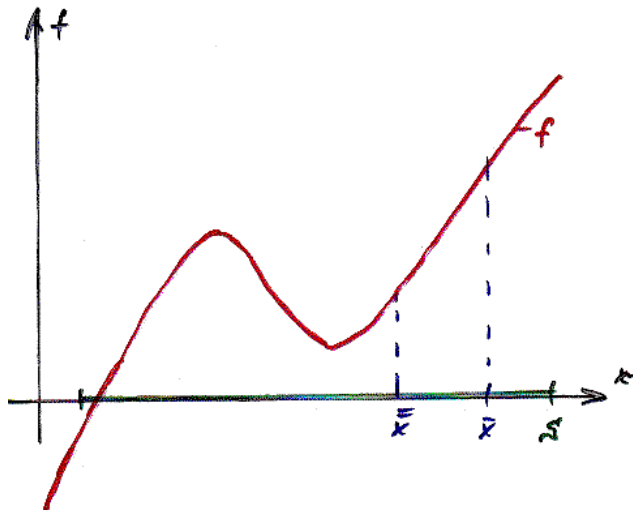
Bemerkung.

1. In den Anwendungen wird häufig S eine sog. "konvexe" Menge sein. Dann ist die Forderung $(\lambda x + (1 - \lambda)\bar{x}) \in S$ automatisch erfüllt, wenn $x, \bar{x} \in S$.

2. Beispiel



f ist quasikonvex auf S



f weder quasikonvex noch quasikonkav auf S , aber f ist quasikonvex in \bar{x} , in \bar{x} nicht quasikonvex bzgl. S

3. Im besonders interessanten Fall, daß S eine sog. konvexe Menge ist, d.h. mit $x^1, x^2 \in S$ und $\lambda \in [0, 1]$ gilt stets $\lambda x^1 + (1 - \lambda)x^2 \in S$, haben wir: f ist auf S genau dann quasikonvex, wenn $\forall x^1, x^2 \in S, \lambda \in [0, 1]$ gilt

$$f(\lambda x^1 + (1 - \lambda)x^2) \leq \max\{f(x^1), f(x^2)\} \quad (+)$$

(Funktionswert in einem Punkt des Intervalls $[x^1, x^2]$ ist nie größer als der größere Funktionswert an den Intervallenden).

Satz (1.8.2). (Hinreichendes Optimalitätskriterium für globales Optimum)
Wir betrachten Aufgabe (OAGR). Seien $f, g_i, i = 1, \dots, m$ in \bar{x} differenzierbar. Sei \bar{x} stationärer Punkt von (OAGR), d.h. es existieren $\lambda_o^* \geq 0, \lambda_1^*, \dots, \lambda_m^* \in \mathbb{R}$, so daß

$$(\circ) \begin{cases} \lambda_o^* \nabla f(\bar{x}) + \sum_{i=1}^m \lambda_i^* \nabla g_i(\bar{x}) = 0 \\ g_i(\bar{x}) = 0, \quad i = 1, 2, \dots, m \\ \sum_{i=0}^m (\lambda_i^*)^2 > 0 \end{cases}$$

Wenn gilt

$$\begin{cases} \lambda_o^* \neq 0 \\ g_i \text{ ist bei } \lambda_i^* > 0 \text{ quasikonvex in } \bar{x} \text{ bzgl. } \mathbb{R}^n \\ g_i \text{ ist bei } \lambda_i^* < 0 \text{ quasikonkav in } \bar{x} \text{ bzgl. } \mathbb{R}^n \\ f \text{ ist in } \bar{x} \text{ pseudokonvex bzgl. } S \end{cases}$$

dann ist \bar{x} global-optimale Lösung von (OAGR). (S - zulässiger Bereich von (OAGR))

Bemerkung.

Im Beweis haben wir die Differenzierbarkeit der g_i bei $i : \lambda_i^* = 0$ in $x = \bar{x}$ nicht benötigt. Da wir die erste Zeile von (o) auch so formulieren können:

$$\lambda_o^* \nabla f(\bar{x}) + \sum_{i:\lambda_i^* \neq 0} \lambda_i^* \nabla g_i(\bar{x}) = 0$$

so ist die Differenzierbarkeit der g_i für die $\lambda_i^* = 0$ im Satz 1.8.2 auch nicht erforderlich.

1.4 Die lineare Optimierungsaufgabe

1.4.1 Aufgabenstellung

Definition.

Wir sagen, daß die Funktion $f : \mathbf{R}^n \rightarrow \mathbf{R}$ eine **affin-lineare Funktion** ist, wenn ein Vektor $a \in \mathbf{R}^n$ und eine Zahl $\alpha \in \mathbf{R}$ existieren, so daß $\forall x \in \mathbf{R}^n$ gilt

$$f(x) = \langle a, x \rangle + \alpha = \sum_{j=1}^m a_j x_j + \alpha.$$

Bemerkung.

1. Für affin-lineare Funktionen f ist also $\nabla f(x) = a = \text{const.}$
2. Bei $\alpha = 0$ sprechen wir von linearen Funktionen.

Definition.

Sei $S \subset \mathbf{R}^n$. Wir sagen, daß S eine **polyedrale Menge** ist, wenn S als Lösungsmenge eines endlichen Systems von Gleichungen und Ungleichungen mit affin-linearen Funktionen darstellbar ist.

Bemerkung.

Wenn $S \subset \mathbf{R}^n$ polyedral ist, dann existieren also $m_1 \geq 0$ affin-lineare Funktionen

$$f_i(x) = \langle a^i, x \rangle + \alpha_i, \quad i = 1, 2, \dots, m_1$$

und $m_2 \geq 0$ affin-lineare Funktionen

$$h_j(x) = \langle b^j, x \rangle + \beta_j, \quad j = 1, 2, \dots, m_2$$

so, daß

$$S = \{x \in \mathbf{R}^n : f_i(x) \leq 0, i = 1, 2, \dots, m_1 \text{ und } h_j(x) = 0, j = 1, 2, \dots, m_2\}$$

In der Regel werden wir S in der Matrix-Vektor-Schreibweise darstellen:

$$S = \{x \in \mathbf{R}^n : Ax \leq a \text{ und } Bx = b\}$$

wobei

$$\begin{cases} A - \text{Matrix vom Typ } [m_1, n] \\ B - \text{Matrix vom Typ } [m_2, n] \\ a - \text{Vektor des } \mathbf{R}^{m_1} \\ b - \text{Vektor des } \mathbf{R}^{m_2}, \end{cases}$$

die Zeilen von A sind die a^i , die Zeilen von B sind die b^j , die Komponenten von a sind die $(-\alpha_i)$, die Komponenten von b sind die $(-\beta_j)$.

Definition.

Wir sagen, daß

$$(OA) \begin{cases} f(x) \rightarrow \min \\ x \in S \end{cases}$$

eine **lineare Optimierungsaufgabe (LOA)** ist, wenn f eine lineare Funktion des \mathbf{R}^n und $S \subset \mathbf{R}^n$ eine polyedrale Menge ist.

Bemerkung.

1. Die LOA sind die verbreitetsten Optimierungsaufgaben in der Anwendung. Für ihre Lösung gibt es ein reichhaltiges (kommerzielles) Angebot sicherer Software.
2. Im allgemeinen hat eine LOA also die Gestalt

$$(LOA) \begin{cases} \langle c, x \rangle \rightarrow \min \\ Ax = b \\ Bx \leq d \end{cases} \quad \text{mit} \quad \begin{cases} c \in \mathbf{R}^n, b \in \mathbf{R}^{m_1}, d \in \mathbf{R}^{m_2} \\ A \in \mathbf{R}^{m_1 \times n}, B \in \mathbf{R}^{m_2 \times n}. \end{cases}$$

3. Es gibt viele spezielle Typen von LOA (Transportprobleme, LOA auf Graphen, ...) für die auch spezielle Lösungsverfahren entwickelt wurden. Wir betrachten in Abschnitt 1.4 nur einen allgemeinen Typ (Fall $B := -I, d = 0$), auf den alle LOA zurückgeführt werden können (vgl. Übung).

In diesem Abschnitt geht es auch nur um das klassische Lösungsverfahren für diesen Aufgabentyp, die sog. Simplexmethode. Gegen Ende des Semesters (vgl. Abschnitte 4.3, 4.4) beschäftigen wir uns mit der Theorie der LOA näher.

Definition.

Wir sagen, daß eine LOA in der **(K)-Form** (=kanonische Form) gegeben ist, wenn die Ungleichungsnebenbedingungen in der Form $x \geq 0$ erscheinen und ansonsten nur Nebenbedingungen in Gleichungsform vorkommen.

Bemerkung.

In der (K)-Form hat eine LOA also die Gestalt

$$(K) \begin{cases} \langle c, x \rangle \rightarrow \min \\ Ax = b \\ x \geq 0 \end{cases}$$

Für das Weitere vereinbaren wir $c \in \mathbf{R}^n, b \in \mathbf{R}^m$, A - Matrix vom Typ $[m, n]$

Definition.

Es sei eine LOA in (K)-Form gegeben. Es bezeichne $A_j \in \mathbf{R}^m$ die j -te Spalte von Matrix A . Für ein $x \in \mathbf{R}^n$ setzen wir $J(x) = \{j : x_j \neq 0\}$. Wir sagen, daß $x \in \mathbf{R}^n$ eine Basislösung von (K) ist, wenn gilt

$$\begin{cases} x \text{ ist zulässige Lösung von (K) (d.h. } Ax = b, x \geq 0) \\ \{A_j : j \in J(x)\} \text{ ist eine linear unabhängige Menge.} \end{cases}$$

Bemerkung.

- Wir führen hier ein Standardbeispiel ein, das wir im Weiteren zu Illustrationszwecken heranziehen wollen:

$$(\square) \begin{cases} x_1 + 6x_2 - 7x_3 + x_4 + 5x_5 + 10x_6 \rightarrow \min \\ 5x_1 - 4x_2 + 13x_3 - 2x_4 + x_5 + 5x_6 = 20 \\ x_1 - x_2 + 5x_3 - x_4 + x_5 + 2x_6 = 8 \\ x_j \geq 0, \quad j = 1, 2, \dots, 6 \end{cases}$$

$$c = [1, 6, -7, 1, 5, 10], b = \begin{pmatrix} 20 \\ 8 \end{pmatrix} \quad A = \begin{pmatrix} 5 & -4 & 13 & -2 & 1 & 5 \\ 1 & -1 & 5 & -1 & 1 & 2 \end{pmatrix}$$

$$\bar{x} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 4 \end{pmatrix}, \quad \hat{x} = \begin{pmatrix} 3 \\ 0 \\ 0 \\ 0 \\ 5 \\ 0 \end{pmatrix}, \quad \tilde{x} = \begin{pmatrix} \frac{3}{2} \\ 0 \\ 0 \\ 0 \\ \frac{5}{2} \\ 0 \end{pmatrix} = 1/2\bar{x} + 1/2\hat{x}, \quad x^* = \begin{pmatrix} 0 \\ \frac{4}{7} \\ \frac{12}{7} \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

sind zulässige Lösungen von (\square) . Die zugehörigen Zielfunktionswerte sind $\langle c, \bar{x} \rangle = 40, \langle c, \hat{x} \rangle = 28, \langle c, \tilde{x} \rangle = 34, \langle c, x^* \rangle = -60/7$.

2. $J(\bar{x}) = \{6\} \Rightarrow \bar{x}$ ist eine Basislösung,
 $J(\hat{x}) = \{1, 5\} \Rightarrow \hat{x}$ ist eine Basislösung,
 $J(\tilde{x}) = \{1, 5, 6\} \Rightarrow \tilde{x}$ ist keine Basislösung,
 $J(x^*) = \{2, 3\} \Rightarrow x^*$ ist eine Basislösung.
3. Wir vereinbaren, daß bei $b = 0$ die zulässige Lösung $x = 0$ als Basislösung gilt ($J(x) = \emptyset$).
4. Es sei x eine Basislösung von (K). Da $Ax = b$ in der Form $\sum_{j=1}^n A_j x_j = b$ aufgeschrieben werden kann und $x_j = 0$ bei $j \notin J(x)$, so ist x Lösung des Systems (+)

$$(+)\begin{cases} \sum_{j \in J(x)} A_j z_j = b \\ z_j = 0, & j \notin J(x) \end{cases}$$

Wir setzen $B := (A_j)_{j \in J(x)}$ und $I_x = (e^j)_{j \notin J(x)}$, wobei

$$e^j = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix}$$

B ist eine Matrix vom Typ $[m, |J(x)|]$.
 I_x ist eine Matrix vom Typ $[n, n - |J(x)|]$ ($|J(x)|$ steht für $\text{card}(J(x))$).
Das System (+) hat dann die Systemmatrix

$$L = \begin{pmatrix} B & 0 \\ 0 & I_x \end{pmatrix}$$

vom Typ $[m + n, |J(x)| + n - |J(x)|] = [m + n, n]$ und da

$$\begin{cases} \text{rank}(B) = |J(x)| & (\text{da } x \text{ Basislösung}), \\ \text{rank}(I_x) = n - |J(x)| \end{cases}$$

so ist $\text{rank}(L) = n$. D.h. (+) ist ein lineares Gleichungssystem mit n Unbekannten mit vollem Rang der Systemmatrix.

Deshalb gibt es für (+) nur 2 Möglichkeiten. Entweder ist (+) nicht lösbar oder es ist eindeutig lösbar. Da aber (siehe oben) x Lösung von (+) ist, so ist es die einzige Lösung von (+).

Damit haben wir ein Charakteristikum einer Basislösung von (K) gefunden: Eine Basislösung x von (K) zeichnet sich dadurch aus, daß sie einzige Lösung des mit x verbundenen Systems (+) ist.

5. Wir vermerken, daß (K) höchstens endlich viele (d.h. keine oder endlich viele) Basislösungen besitzen kann. Denn es gibt nur endlich viele unterschiedliche Teilmengen $J \subset \{1, 2, \dots, n\}$ und damit nur endlich viele Systeme (+)
 (dabei fallen diejenigen J sogar noch weg, für die $\{A_j\}_{j \in J}$ linear abhängig und auch die J , für die $\{A_j\}_{j \in J}$ zwar linear unabhängig ist, aber die zugehörige (eindeutige) Lösung $z = [z_1, \dots, z_n]$ von (+) negative Komponenten enthält).

Definition.

Es sei $x \in \mathbf{R}^n$ eine Basislösung von (K). Jede Indexmenge $S \subset \{1, 2, \dots, n\}$ mit den Eigenschaften

$$\begin{cases} J(x) \subset S \\ \text{card } S = m \\ B = \{A_j : j \in S\} \text{ ist eine linear unabhängige Menge} \end{cases}$$

heißt **Basisindexmenge** zur Basislösung x .

Die zugehörige Matrix $B = \{A_j : j \in S\}$ heißt **Basismatrix** zur Basisindexmenge S .

Bemerkung.

1. Wir betrachten unser Standardbeispiel.

$$\bar{x}^\top = [0, 0, 0, 0, 0, 4], \quad J(\bar{x}) = \{6\}.$$

$$\text{Mögliche Wahl von } S \text{ und } B \text{ zu } \bar{x} : \quad S = \{6, 1\},$$

$$B = \begin{pmatrix} 1 & 5 \\ 5 & 2 \end{pmatrix}$$

oder $S = \{6, 2\}$,

$$B = \begin{pmatrix} -4 & 5 \\ -2 & 2 \end{pmatrix}$$

oder ...

$$\hat{x}^\top = [3, 0, 0, 0, 5, 0], \quad J(\hat{x}) = \{1, 5\}$$

$$\text{Einzig mögliche Wahl von } S \text{ und } B : \quad S = J(\hat{x}),$$

$$B = \begin{pmatrix} 5 & 1 \\ 1 & 1 \end{pmatrix}$$

2. Man sieht leicht, wenn $\text{rank}(A) < m$, dann kann es zu einer beliebigen Basislösung (BL) von (K) keine Basismatrix (BM) und auch keine Basisindexmenge (BiM) geben. Wir werden deshalb im Weiteren stets $\text{rank}(A) = m$ voraussetzen.

Das ist keine Einschränkung der Allgemeinheit, denn bei $\text{rank}(A) < m$ sind im lösbaeren System $Ax = b$ genau $k = m - \text{rank}(A)$ der linearen Gleichungen überflüssig. Genauer: Es existieren k Zeilen in A und b so, daß für die nach Streichen dieser Zeilen verbliebenen Restmatrix \hat{A} und den Restvektor \hat{b} gilt

$$\{x : Ax = b\} = \{x : \hat{A}x = \hat{b}\}.$$

Die Restmatrix \hat{A} hat dann genau $\text{rank}(A)$ Zeilen und $\text{rank}(\hat{A}) = \text{rank}(A)$, d.h. in ihr stimmen Zeilenanzahl und Rang überein.

3. Wenn $\text{rank}(A) = m$, dann existiert zu jeder BL x von (K) wenigstens eine BM B und eine BiM S , denn dann ist $\{A_j : j \in J(x)\}$ eine linear unabhängige Menge von Vektoren des \mathbf{R}^m .

Da $\text{rank}(A) = m$, so enthält die Menge der Spalten von A wenigstens eine Basis \mathcal{L}_o des \mathbf{R}^m . Nach einem Satz der linearen Algebra läßt sich aber jede linear unabhängige Menge zu einer Basis \mathcal{L}_1 des Raumes ergänzen, wobei die ergänzenden Vektoren aus der Basis \mathcal{L}_o entnommen werden können.

Also, bei $\text{rank}(A) = m$ läßt sich $\{A_j : j \in J(x)\}$ unter Zuhilfenahme weiterer Spalten von A zu einer Basis des \mathbf{R}^m ergänzen, d.h. es existieren m linear unabhängige Spalten von A , die die Spalten A_j mit $j \in J(x)$ enthalten, d.h. S und B existieren (nicht unbedingt eindeutig).

4. Eine Basismatrix ist eine quadratische Matrix vom Typ $[m, m]$ aus m linear unabhängigen Spalten, d.h. $\det(B) \neq 0$ und es existiert B^{-1} .

1.4.2 Theoretische Grundlagen der Simplexmethode

Satz (1.9). (Hinreichendes Optimalitätskriterium)

Gegeben sei eine LOA in (K)-Form. Es sei x^* eine BL von (K) und S eine mit x^* verbundene BiM. Es sei y^* Lösung des Systems

$$c_j + \langle y^*, a_j \rangle = 0 \quad , \forall j \in S. \quad (+)$$

Wir setzen $\Delta_j := c_j + \langle y^*, A_j \rangle, j \notin S$. Wenn

$$\Delta_j \geq 0 \quad , \forall j \notin S, \quad (++)$$

dann ist x^* optimale (d.h. global-optimale) Lösung von (K).

Bemerkung.

1. Wie aus Satz 1.11 und der Definition des später einzuführenden Begriffes "nichtentartete BL" folgen wird, ist das Kriterium (++) für nichtentartete BL auch notwendig für die Optimalität. Bei solchen BL ist die Basisindexmenge S eindeutig festgelegt. Im allgemeinen Fall gilt für BL folgendes Optimalitätskriterium: Die BL x^* ist optimal genau dann, wenn zu x^* eine Basisindexmenge S existiert, in der für den nach (+) berechneten Vektor y^* die Eigenschaft (++) vorliegt.
2. Aus dem später zu betrachtenden Kuhn-Tucker-Theorem (Theorem 3.11) wird folgendes allgemeine Optimalitätskriterium folgen: Die zulässige Lösung x^* (nicht unbedingt BL) von Aufgabe (K) ist optimal genau dann, wenn ein $y^* \in \mathbf{R}^m$ existiert, für das

$$\begin{cases} c + y^*A \geq 0 \\ \langle c + y^*A, x^* \rangle = 0 \end{cases}$$

erfüllt ist (vgl. auch Satz 4.5).

3. Die Δ_j werden Optimalitätsindikatoren genannt. Genauer: Δ_j ist der Optimalitätsindikator zur Variablen x_j bei der BiM S .
4. Wir betrachten unser Standardbeispiel.

(a) $\hat{x} = [3, 0, 0, 0, 5, 0]^\top$, $S = \{1, 5\}$

$$B = \begin{pmatrix} 5 & 1 \\ 1 & 1 \end{pmatrix}$$

(+) ist das System

$$\begin{cases} 5y_1 + y_2 + 1 = 0 \\ y_1 + y_2 + 5 = 0 \end{cases}$$

Es hat die Lösung $y_1^* = 1, y_2^* = -6$. Daraus erhalten wir nach Definition der Δ_j : $\Delta_1 = c_1 + \langle y^*, A_1 \rangle = 0$

$$\Delta_2 = 8, \quad \Delta_3 = -24, \quad \Delta_4 = 5, \quad \Delta_5 = 0, \quad \Delta_6 = 3.$$

Weil $\Delta_3 < 0$, so ist unser notwendiges Optimalitätskriterium nicht erfüllt. \implies Vermutlich ist \hat{x} keine optimale Lösung.

(b) $x^* = [0, 4/7, 12/7, 0, 0, 0]^\top$. Hier ist $S = \{2, 3\}$,

$$B = \begin{pmatrix} -4 & 13 \\ -1 & 5 \end{pmatrix}$$

und damit ist (+) das System

$$\begin{cases} -4y_1 - y_2 + 6 = 0 \\ 13y_1 + 5y_2 - 7 = 0 \end{cases}$$

Es hat die Lösung $y_1^* = 23/7, y_2^* = -50/7$. Daraus erhalten wir $\Delta_1 = c_1 + \langle y^*, A_1 \rangle = 72/7$

$$\Delta_2 = 0, \quad \Delta_3 = 0, \quad \Delta_4 = 11/7, \quad \Delta_5 = 8/7, \quad \Delta_6 = 85/7.$$

Alle $\Delta_j \geq 0 \Rightarrow x^*$ ist optimale Lösung des Standardbeispiels.

Da $\langle c, x^* \rangle = -60/7$, aber $\langle c, \hat{x} \rangle = 28$, so ist \hat{x} sicher keine optimale Lösung (in Bemerkung 4(a) hatten wir festgestellt, daß \hat{x} vermutlich nicht optimal ist).

Für die Unlösbarkeit von

$$f^* = \inf\{\langle c, x \rangle : Ax = b, x \geq 0\} \quad (K)$$

sind 3 Gründe denkbar

- $S = \{x \in \mathbf{R}^n : Ax = b, x \geq 0\} = \emptyset$
- $S \neq \emptyset$, aber $f^* = -\infty$
- $S \neq \emptyset, f^* > -\infty$, aber $\nexists x^* \in S$, so daß $f^* = \langle c, x^* \rangle$

Wie wir das Vorliegen des ersten Falls austesten können, wird Satz 1.14 zeigen. In einem späteren Abschnitt (Theorem 2.22) zeigen wir, daß der dritte Fall bei LOA unmöglich ist. Jetzt zeigen wir, wie wir das Vorliegen des zweiten Falles austesten können.

Satz (1.10). (Optimum im Unendlichen)

Wir betrachten Aufgabe (K). Es sei x eine BL von (K) und S sowie B seien eine zu x gehörende BiM bzw. BM. Schließlich sei y die Lösung des Systems

$$c_j + \langle y, A_j \rangle = 0, j \in S \quad (+)$$

Wenn ein Index $k \in \{1, 2, \dots, n\} \setminus S$ existiert, für den gilt

$$\begin{cases} \Delta_k := c_k + \langle y, A_k \rangle < 0 \\ x_{\cdot k} := B^{-1}A_k \leq 0 \end{cases}$$

dann ist $f^* = \inf\{\langle c, z \rangle : Az = b, z \geq 0\} = -\infty$, d.h. Aufgabe (K) ist dann unlösbar.

Bemerkung.

1. Wie im Beweis des Satzes sei angenommen, daß $S = \{s_1, s_2, \dots, s_m\} \subset \{1, 2, \dots, n\}$ die mit x verbundene BiM ist. Wie der Beweis zeigt, ist unter den Voraussetzungen des Satzes $x(\Theta) = x + \Theta d$, $\Theta \geq 0$ mit $d^T = [d_1, d_2, \dots, d_n]$,

$$d_j = \begin{cases} -x_{ik} & , j = s_i \\ 1 & , j = k \\ 0 & , sonst \end{cases}$$

ein unendlicher Strahl in

$$S = \{x \in \mathbf{R}^n : Ax = b, x \geq 0\}$$

und auf ihm gilt $\lim_{\Theta \rightarrow \infty} \langle c, x(\Theta) \rangle = -\infty$.

2. Anfänglich hatten wir die Optimalitätsindikatoren über den Vektor y aus System (+) definiert:

$$\Delta_j := c_j + \langle y, A_j \rangle .$$

Wie die Rechnungen im Beweis zeigen, läßt sich Δ_j über x_k und $c_B = [c_{s_1}, c_{s_2}, \dots, c_{s_m}]$ auch wie folgt berechnen:

$$\Delta_j = c_j - \langle c_B, x_k \rangle .$$

Satz (1.11). (Möglichkeit des Übergangs zu einer benachbarten Basislösung)
Wir betrachten Aufgabe (K). Es sei x eine BL von (K) und $S = \{s_1, s_2, \dots, s_m\}$ sowie B seien eine zu x gehörende BiM bzw. BM. Es sei y die Lösung des Systems

$$c_j + \langle y, A_j \rangle = 0, \quad \forall j \in S \quad (+)$$

Schließlich sei für ein $k \in \{1, 2, \dots, n\} \setminus S$ $\Delta_k := c_k + \langle y, A_k \rangle < 0$

$$x_k = \begin{pmatrix} x_{1k} \\ \vdots \\ x_{mk} \end{pmatrix} = B^{-1} A_k \not\leq 0.$$

Wir setzen

$$(\Delta) \begin{cases} \Theta_o = \min_i \left\{ \frac{x_{s_i}}{x_{ik}} : i, \text{ so da\ss } x_{ik} > 0 \right\} = \frac{x_{s_r}}{x_{rk}} \\ S' = (S \setminus \{s_r\}) \cup \{k\} \\ x' = [x'_1, \dots, x'_n]^\top \text{ mit} \\ x'_j = \begin{cases} x_j - \Theta_o x_{ik}, & j = s_i \\ \Theta_o, & j = k \\ 0, & \text{sonst} \end{cases} \quad \text{f\"ur } j = 1, 2, \dots, n \\ B' = \{A_j\}_{j \in S'} \end{cases}$$

Dann gilt:

$$\begin{cases} x' \text{ ist Basisl\"osung von (K)} \\ S' \text{ ist eine zu } x' \text{ geh\"orende BiM} \\ B' \text{ ist eine zu } x' \text{ geh\"orende BM} \\ \langle c, x' \rangle = \langle c, x \rangle + \Theta_o \Delta_k \leq \langle c, x \rangle . \end{cases}$$

Bemerkung.

1. Satz 1.11 besagt: Wenn auf x weder das Optimalit\"atskriterium von Satz 1.9 noch das Unl\"osbarkeitskriterium von Satz 1.10 erf\"ullt sind, dann ist nach den Formeln (Δ) der \"Ubergang von der BL x zu einer neuen BL x' mit zugeh\"origer BM und BiM m\"oglich. Wenn dabei $\Theta_o > 0$ (i.a. ist $\Theta_o \geq 0$), gilt, dann ist $\langle c, x' \rangle < \langle c, x \rangle$.
2. Wir demonstrieren die Aussage des Satzes am Vorrechenbeispiel:
Auf der zul. BL $\hat{x} = [3, 0, 0, 0, 5, 0]$ mit $S = J(\hat{x}) = \{1, 5\}$,

$$B = \begin{pmatrix} 5 & 1 \\ 1 & 1 \end{pmatrix}$$

war $\Delta_k = \Delta_3 < 0$ (vgl. Bemerkung 4(a) zu Satz 1.9)

Wir haben $S = \{s_1, s_2\} = \{1, 5\}$, $x_{.3} = B^{-1}A_3 \implies Bx_{.3} = A_3 \implies$

$$\begin{pmatrix} 5 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x_{13} \\ x_{23} \end{pmatrix} = \begin{pmatrix} 13 \\ 5 \end{pmatrix}$$

\implies

$$x_{.3} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$$

Dann ist $I = \{i : x_{ik} > 0\} = \{1, 2\}$

$$\Theta = \min \left\{ \frac{x_{s_1}}{x_{13}}, \frac{x_{s_2}}{x_{23}} \right\} = \min \{3/2, 5/3\} = 3/2$$

$$\implies s_r = s_1 = 1.$$

$$\begin{aligned}
&\text{Neue BL } x' = \{x'_j\}_{j=1}^5, \quad \text{wobei} \\
&x'_3 = \Theta = 3/2 \\
&x'_j = 0, \text{ für } j = 2, 4, 6, \text{ da diese } j \notin S \text{ und } j \neq k(= 3) \\
&x'_1 = 3 - 3/2 \cdot 2 = 0 \\
&x'_5 = 5 - 3/2 \cdot 3 = 1/2 \\
&\implies x' = [0, 0, 3/2, 0, 1/2, 0], S' = \{3, 5\} \text{ und} \\
&\langle c, x' \rangle = \langle c, x \rangle + \Theta \cdot \Delta_3 = 28 - 3/2 \cdot 24 = -8
\end{aligned}$$

Die Bemerkung suggeriert den Gedanken, auf diese Weise zu einer optimalen Lösung zu gelangen: starten mit BL, testen auf Optimalität, wenn Optimalität der vorliegenden BL nicht feststellbar, gehen wir zu einer neuen BL mit nichtschlechteren ZF-Wert über. Nach Satz 1.11 ist das möglich, wenn auf der vorliegenden BL die Unlösbarkeit von (K) nicht feststellbar ist.

Unten (vgl. Satz 1.13) werden wir für den sogenannten Nichtentartungsfall zeigen, daß man auf diese Weise zum Optimum gelangen kann.

Aus der Vorgehensweise wird folgen, daß die so erhaltene Lösung eine BL ist. Mit unserem letzten Satz zur Theorie wollen wir zeigen, daß die Vorgehensweise auch dann sinnvoll sein kann, wenn der sogenannte Nichtentartungsfall nicht vorliegt.

Satz (1.12). (Existenz einer optimalen BL)

Aufgabe (K) sei lösbar (d.h. es existiert eine global-optimale Lösung). Dann existiert unter den optimalen Lösungen eine, die Basislösung ist.

Bemerkung.

1. Man könnte meinen, die Aussage des Satzes ist trivial, da alle optimalen Lösungen von (K) Basislösungen sind. Das ist i.a. nicht so, wie folgende Überlegung nahelegt:

Wenn x^* und x^{**} zwei verschiedene optimale Lösungen sind, die BL sind, dann ist $x(\lambda) = \lambda x^* + (1 - \lambda)x^{**}$ bei $\lambda \in [0, 1]$ auch optimale Lösung, aber bei $\lambda \in (0, 1)$ i.a. keine BL (da $\{A_j\}_{j \in J(x(\lambda))}$ nicht linear unabhängig sein wird).

2. Auf analoge Weise wie Satz 1.12 von uns bewiesen wurde, kann man auch folgende Aussage zeigen: Wenn Aufgabe (K) eine zulässige Lösung hat, dann hat sie auch eine zulässige Lösung, die BL ist. Diese Aussage erhalten wir u.a. in Satz 1.14 auf andere Weise.

3. Basislösungen von (K) sind sogenannte Extrempunkte der Menge $S = \{x \in \mathbf{R}^n : Ax = b, x \geq 0\}$.

S ist eine sogenannte konvexe Menge (vgl. Abschnitt 2.1), $f(x) = \langle$

$c, x >$ ist eine sogenannte konkave Funktion (vgl. Abschnitt 2.3). Für Aufgaben

$$(OA) \begin{cases} f(x) \rightarrow \min \\ x \in S \end{cases} \quad \text{mit } f \text{ konkav, } S \text{ konvex}$$

zeigen wir eine Verallgemeinerung der Aussage von Satz 1.12: Wenn S wenigstens einen Extrempunkt besitzt und (OA) lösbar ist, dann ist unter den optimalen Lösungen von (OA) wenigstens ein Extrempunkt von S (vgl. Satz 2.35).

1.4.3 Die primale modifizierte Simplexmethode

Die Sätze aus Abschnitt 1.4.2 legen folgenden Prinzipialgorithmus nahe (Wir betrachten vorläufig nur die Situation, daß eine zul. BL als Startlösung bekannt ist \implies impliziert, daß $\{x : Ax = b, x \geq 0\} \neq \emptyset$).

- **0.Schritt:**

Es sei x^o eine (zulässige) BL von (K) und $S_o = \{s_1^o, \dots, s_m^o\}$ sowie $B_o = \{A_j\}_{j \in S_o}$ eine zugehörige BiM bzw. BM. Es sei $c_{B_o} = [c_{s_1^o}, \dots, c_{s_m^o}]^T$. Wir setzen $t := 0$ und berechnen für alle $j \in \{1, 2, \dots, n\}$ die Vektoren

$$x_{\cdot j}^t = B_t^{-1} A_j = [x_{1j}^t, \dots, x_{mj}^t]^T$$

(= Koordinaten der Vektoren A_j in der aktuellen Basis) Wenn es unmöglich ist, eine BL x^o zu berechnen, dann Ende A.

- **1. Schritt:**

Für $j=1,2,\dots,n$ berechnen wir $\Delta_j^t = c_j - \langle c_{B_t}, x_{\cdot j}^t \rangle$ und setzen

$$z_t = \langle c_B, x_{B_t}^t \rangle, \text{ wobei } x_{B_t}^t = [x_{s_1}^t, \dots, x_{s_m}^t]^T.$$

- **2.Schritt:**

Wir berechnen $\Delta_k^t = \min_{0 \leq j \leq n} \Delta_j^t$. Wenn $\Delta_k^t \geq 0$, dann Ende B. (x^t ist optimale Lösung von (K), vgl. Satz 1.9)

- **3.Schritt:**

Wenn $x_{\cdot k}^t \leq 0$, dann Ende C. (Aufgabe (K) ist unlösbar, vgl. Satz 1.10) Andernfalls berechnen wir einen Index $s_r^t \in I_t$ und die Zahl Θ_t , so daß

$$\Theta_t := \frac{x_{s_r}^t}{x_{rk}^t} := \min \left\{ \frac{x_{s_i}^t}{x_{ik}^t} : i \in I_t \right\}.$$

Hierbei bezeichnet $I_t = \{i : x_{ik}^t > 0\}$.

• **4.Schritt:**

Wir setzen $x^{t+1} := [x_1^{t+1}, \dots, x_n^{t+1}]^\top$ und $S_{t+1} := (S_t \setminus \{s_r^t\}) \cup \{k\}$ wobei

$$x_j^{t+1} := \begin{cases} x_{s_i^t}^t - \Theta_t x_{ik}^t, & \text{wenn } j = s_i^t \\ \Theta_t & , \text{wenn } j = k \\ 0 & , \text{sonst} \end{cases} \quad \text{für } j = 1, 2, \dots, n$$

(Nach Satz 1.11 ist x^{t+1} BL und S_{t+1} eine zugehörige BiM.)

• **5.Schritt:**

Wir berechnen die mit x^{t+1} und S_{t+1} verbundene Information

$$x_j^{t+1}, \forall j \quad | \quad \Delta_j^{t+1} \quad | \quad z_{t+1} \quad | \quad x_{B_{t+1}}^{t+1} \quad | \quad c_{B_{t+1}}$$

und gehen mit $t := t + 1$ zu Schritt 2)

Bemerkung.

1. Es bleibt noch zu zeigen, daß der Algorithmus gegen eine optimale Lösung konvergiert (siehe unten).
2. Die Wahl des Indexes k nach $\Delta_k^t = \min_{j \notin S_t} \Delta_j^t$ ist nicht zwingend für die SM erforderlich. Wie Satz 1.11 zeigt, ist in der primalen SM nur notwendig, k so zu wählen, daß $\Delta_k^t < 0$.
3. Für die praktische Rechnung ist es sinnvoll, die mit der aktuellen BL verbundene Information in einem Schema anzuordnen. Üblich ist die sogenannte modifizierte Simplextabelle. (wir lassen den Iterationsindex t weg).

S	c_B	x_B	1	2	...	n	Θ
s_1	c_{s_1}	x_{s_1}	x_{11}	x_{12}	...	x_{1n}	
s_2	c_{s_2}	x_{s_2}	x_{21}	x_{22}	...	x_{2n}	
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	
s_m	c_{s_m}	x_{s_m}	x_{m1}	x_{m2}	...	x_{mn}	
		$-z$	Δ_1	Δ_2	...	Δ_n	

In die Spalte Θ tragen wir später (nur in den Zeilen, die in I_t eingehen) die Quotienten $\frac{x_{s_i}}{x_{ik}}$ ein.

4. Es sei vermerkt, daß sich das Ausfüllen der ersten Tabelle besonders einfach gestaltet, wenn $B_o = I$.

Es seien hier nochmals die Formeln zusammengestellt (vgl. Sätze 1.9 - 1.11), nach denen sich der Inhalt der Spalten $x_B, 1, 2, \dots, n$ berechnet:

$$(+)\begin{cases} x_{.j} = [x_{1j}, x_{2j}, \dots, x_{mj}]^T = B^{-1}A_j \\ x_B = B^{-1}b = [x_{s_1}, x_{s_2}, \dots, x_{s_m}]^T \\ z = \langle c_B, x_B \rangle \\ \Delta_j = c_j - \langle c_B, x_{.j} \rangle \quad , j = 1, \dots, n \end{cases}$$

5. Ab der ersten Iteration ($t = 1$) läßt sich die nächste Simplextabelle rekursiv aus der vorhergehenden Tabelle aufdatieren (d.h. im 5. Schritt braucht nicht wieder auf die Formeln (+) aus der vierten Bemerkung zurückgegriffen zu werden).

Wir verzichten hier auf die formale Begründung der Aufdatierungsformeln und führen nur an, wie mit der Tabelle zu verfahren ist (Hintergrund ist ein Übergang zu einer neuen Basis, die sich von der vorhergehenden nur durch einen Vektor unterscheidet).

Um die Aufdatierungsformeln anzugeben nehmen wir an, der Index k (Index der sogenannten Leitspalte) sei in Schritt 2 ausgewählt worden ($\Delta_k^t < 0$). Ebenso sei der Index r (Index der sogenannten Leitzeile) im Schritt 3 festgelegt worden. Das Element x_{rk}^t an der Überschneidung von Leitzeile und Leitspalte (es ist positiv) wird als Leitelement bezeichnet.

Die Spalten $x_B, 1, 2, \dots, n$ der Simplextabelle sehen dann so aus (wieder ohne Iterationsindex t):

x_{s_1}	x_{11}	x_{12}	\dots	x_{1k}	\dots	x_{1n}
x_{s_2}	x_{21}	x_{22}	\dots	x_{2k}	\dots	x_{2n}
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
x_{s_r}	x_{r1}	x_{r2}	\dots	x_{rk}	\dots	x_{rn}
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
x_{s_m}	x_{m1}	x_{m2}	\dots	x_{mk}	\dots	x_{mn}
$-z$	Δ_1	Δ_2	\dots	Δ_k	\dots	Δ_n

mit **x_{rk}** ...Leitelement, $-$...Leitzeile und $||$...Leitspalte

Der neue Inhalt dieser Tabelle berechnet sich dann wie folgt:

$$\left\{ \begin{array}{l} \text{Zuerst wird die Leitzeile aufdatiert nach:} \\ \quad \text{neuer Inhalt} := \text{alter Inhalt, dividiert durch Leitelement} \\ \text{Anschließend wird der Rest der Tabelle aufdatiert nach:} \\ \quad \text{neuer Inhalt} := \text{alter Inhalt minus Produkt aus Element} \\ \quad \text{aus neuer Leitzeile (selbe Spalte wie aktuell aufdatierende} \\ \quad \text{Zelle) und Element aus alter Leitspalte (selbe Zeile wie} \\ \quad \text{aktuell aufdatierendes Element).} \end{array} \right.$$

Wir trainieren das in der Übung.

Definition.

Es sei x eine Basislösung von (K). S eine zugehörige Basisindexmenge. Wir sagen, daß x **primal nichtentartet** ist, wenn $S = J(x)$ (d.h. $x_j > 0, \forall j \in S$ bzw. $\text{card}J(x) = m$).

Satz (1.13). (Endlichkeit der SM)

Gegeben sei eine Aufgabe (K) in der alle Basislösungen primal nichtentarten. Es gilt: Wenn die Simplexmethode mit einer BL x^o gestartet wird, dann tritt nach endlich vielen Iterationen entweder Ende B oder Ende C ein.

Bemerkung.

1. Wie der Beweis zeigt, benötigen wir nur, daß alle BL, welche die SM generiert, primal nichtentarten.
2. Ohne Nichtentartungsvoraussetzung ist in der oben beschriebenen SM eine sog. Verzyklung (Erzeugen einer Kette sich ständig wiederholender Basisindexmengen ($S_t \rightarrow S_{t+1} \rightarrow \dots \rightarrow S_{t+v} \rightarrow S_{t+v+1} = S_t \rightarrow S_{t+1} \dots$) möglich.
Das erste Beispiel, das diese Eigenschaft hat, stammt von Beale.
Natürlich ist solch eine Verzyklung nur möglich, wenn primal entartete BL auftreten. Eine genauere Analyse zeigt, daß die Verzyklung mit der Nichteindeutigkeit der Wahl des Indexes s_r in Schritt 3 (Θ_t wird auf mehreren Indizes realisiert) einhergeht.
3. Unser Satz 1.12 legt nahe, daß man aber auch im Entartungsfall über BL zum Optimum gelangen kann (der Satz 1.12 benötigte keine Nichtentartungsvoraussetzung).
Das abzusichern ist aber in der obigen SM nur mit einer Verfeinerung

der Auswahl des Indexes s_r möglich.

Wir wollen uns nicht näher mit solch einer Regel beschäftigen. Ziel einer jeden Regel ist, die Gültigkeit der Aussage von Satz 1.13 zu sichern (natürlich, ohne Nichtentartungsannahme).

4. R.G.Bland gab eine einfache Regel an:

Immer dann, wenn es für die Wahl des Indexes k (Schritt 2 aber nur $\Delta_k < 0$, nicht $\Delta_k = \min_j \Delta_j$, $k = \text{Index der in die Basis aufzunehmenden Variablen}$) oder die Wahl des Indexes r (Schritt 3, $s_r = \text{Index der die Basis verlassenden Variablen}$) eine Wahlmöglichkeit gibt, wähle den Index mit dem kleinsten Wert.

Diese Regel beißt sich aber mit (von uns nicht betrachteten) Überlegungen zur numerischen Stabilität der SM. Diese Überlegungen verlangen, daß $|\Delta_k|$ und $|x_{rk}|$ nicht zu nahe bei Null liegen dürfen.

1.4.4 Die Phase I der SM

Die oben gegebene Beschreibung der Simplexmethode ging im 0.Schritt davon aus, daß wir eine Startbasislösung x^o und die zugehörige Simplextabelle kennen. Nicht immer ist diese Ausgangsinformation leicht beschaffbar:

Im allgemeinen ist die Gewinnung einer Basislösung eine Aufgabe, welche einen Aufwand erfordert, der mit dem der Simplexmethode vergleichbar ist. In der einschlägigen Software für die Simplexmethode ist es üblich, die Methode selbst einzusetzen, um eine Startbasislösung zu berechnen. Dazu teilt man den Rechenprozeß in zwei Phasen auf:

- Phase I: Lösen einer Hilfs-LOA, für die das Problem der Start-Simplextabelle nicht besteht. Anhand der optimalen Lösung dieser Hilfs-LOA kann man entscheiden, ob die eigentlich zu lösende LOA eine BL besitzt und wenn ja, wie eine solche aussieht.
- Phase II: Lösen der eigentlichen LOA mit der SM, beginnend mit der in Phase I erhaltenen Startbasislösung.

Wir wollen Phase I und den Übergang zu Phase II nun näher beschreiben. Gegeben sei eine Aufgabe

$$(K) \begin{cases} \langle c, x \rangle \rightarrow \min \\ Ax = b \\ x \geq 0 \end{cases}$$

Die bisher übliche Forderung $\text{rank}(A) = m$ kann fallengelassen werden. Für das Weitere sei o.E.d.A. angenommen, daß $b \geq 0$. Wir führen die Hilfsaufgabe (H) ein

$$(H) \begin{cases} \sum_{i=1}^m u_i \rightarrow \min \\ Ax + Iu = b \\ x \geq 0, u \geq 0 \end{cases}$$

Hier bezeichne I die Einheitsmatrix der Ordnung m . (H) ist eine LOA in (K)-Form, wie die folgenden Bezeichnungen zeigen:

$$\begin{aligned} \hat{c} &= [0, e] \in \mathbf{R}^{m+n} \text{ mit } e = [1, 1, \dots, 1] \in \mathbf{R}^m \\ z &= [x, u]^T \in \mathbf{R}^{m+n}, \hat{A} = [A, I] \text{ vom Typ } [m, n+m]. \end{aligned}$$

In diesen Bezeichnungen ist (H) die Aufgabe

$$(H) \begin{cases} \langle \hat{c}, z \rangle \rightarrow \min \\ \hat{A}z = b \\ z \geq 0 \end{cases}$$

Wir erkennen leicht folgende Eigenschaften dieser Aufgabe:

1. Der zulässige Bereich von (H) ist nicht leer. So ist, z.B. $z^o = [x^o, u^o]$ mit $x^o := 0, u^o := b$ eine zulässige Lösung (da $b \geq 0$). z^o ist sogar eine BL, denn $S_o := \{n+1, n+2, \dots, n+m\}$ = Indizes der u_i -Variablen in z^o und $B_o := I$ ist eine BiM bzw. BM zu z^o .
2. $\hat{f} = \inf \{ \langle \hat{c}, z \rangle : \hat{A}z = b, z \geq 0 \} \geq 0 \implies \underline{\text{Ende C}}$ (Optimum bei $-\infty$) scheidet für (H) aus.
3. Aus (1) und (2) wird nach einem später auch formal bewiesenen Theorem 2.22 folgen, daß (H) eine optimale Lösung besitzt. Nach Satz 1.11 hat sie also auch eine optimale Basislösung. Nach Satz 1.12 besteht eine Chance, solch eine optimale BL von (H) nach endlich vielen Iterationen, startend mit z^o , zu erhalten.
4. $m = \text{rank}(\hat{A}) \geq \text{rank}(A)$

Satz (1.14).

Es sei $z^* = [x^*, u^*]^T$ eine optimale Basislösung von (H). Es sei $\hat{f} = \langle \hat{c}, z^* \rangle = \langle e, u^* \rangle$. Dann gilt:

1. Wenn $\hat{f} = 0$, dann ist x^* eine Basislösung (d.h. auch zulässige) von Aufgabe (K)

2. Wenn $\hat{f} > 0$, dann ist der zulässige Bereich von Aufgabe (K) leer (Ausgang A der SM).

Bemerkung.

1. In der Zwei-Phasen-Methode zur Lösung von (K) mit der SM wenden wir zuerst die SM auf (H) an und wenden im Falle $\hat{f} = 0$ anschließend die SM auf (K) an (d.h. streichen in (H) die mit den u_i verbundenen Spalten und ersetzen die bisherige Zielfunktion durch $\langle c, x \rangle$.)
2. Wie steht es um BiM und BM für die Startlösung von Phase II ?
 Es sei $z^* = [x^*, u^*]$ optimale BL von (H) mit $u^* = 0$.
 Wenn sich keine der u_i - Variablen mehr in der Basis zu z^* befindet, d.h. die Basis von z^* nur aus Variablen aus dem x -Teil von z besteht, dann ist die BM und BiM von z auch BM bzw. BiM in Aufgabe (K).
 \implies Simplextabellen zu z^* in (H) können im wesentlichen als Startsimplextabellen zu x^* in (K) weiterverwendet werden: Spalte c_B , Zielfunktionswert z und Optimalitätsindikatoren Δ_j müssen entsprechend der Zielfunktion aus Phase II ersetzt werden.
3. Im Fall, daß $u^* = 0$ und sich aber noch wenigstens ein u_i in der Basis von z^* befindet, sollte man versuchen die Voraussetzung der 2. Bemerkung zu erfüllen, d.h. eine andere Basis zu z^* zu finden, in der kein u_i mehr BV ist.
 Dazu führe man am besten weitere Simplexiterationen nach der sogenannten dualen SM (die von uns behandelte SM haben wir primale SM genannt) aus. In dieser SM ist zuerst die aus der Basis zu eliminierende Variable (in unserem Fall u_i) zu fixieren (in der primalen SM wurde zuerst die in die Basis aufzunehmende Variable x_k fixiert).
4. Wenn $\text{rank}(A) < m$, dann wird die in der 3. Bemerkung vorgeschlagene Vorgehensweise (duale SM zur Elimination der u_i aus der Basis benutzen) nicht zum totalen Erfolg führen können.

Kapitel 2

Grundlagen der konvexen Analysis (differenzierbarer Fall)

Die konvexe Analysis liefert Grundlagen nicht nur für die Optimierungstheorie, sondern auch für eine Reihe verwandter Disziplinen, wie der Spieltheorie, des Operation Research, der Mathematischen Ökonomie, der Kontrolltheorie u.a..

2.1 Konvexe Mengen

2.1.1 Der Begriff "konvexe Menge"

Es seien x^1 und x^2 zwei verschiedene Punkte des \mathbf{R}^n . Dann beschreibt die Gleichung

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

für $\lambda \in (-\infty, \infty)$ eine Gerade g des \mathbf{R}^n .

Die Punkte $\phi = x^2 + \lambda(x^1 - x^2)$ für $0 \leq \lambda \leq 1$ sind die Punkte des Abschnittes $[x^1, x^2]$ auf g .

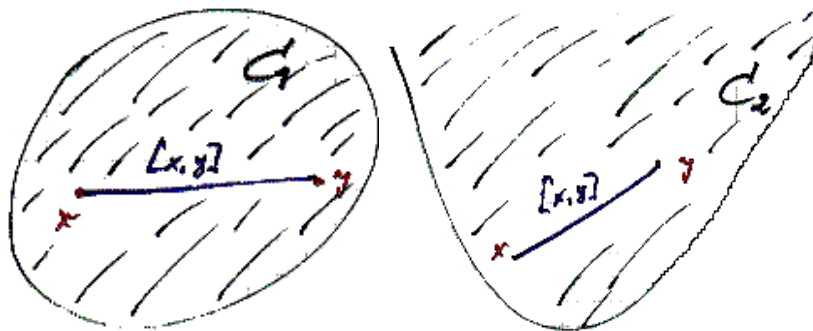
Wir wählen im Weiteren für diese Punkte meist die Darstellung $\phi = \lambda x^1 + (1 - \lambda)x^2$, $0 \leq \lambda \leq 1$.

Definition.

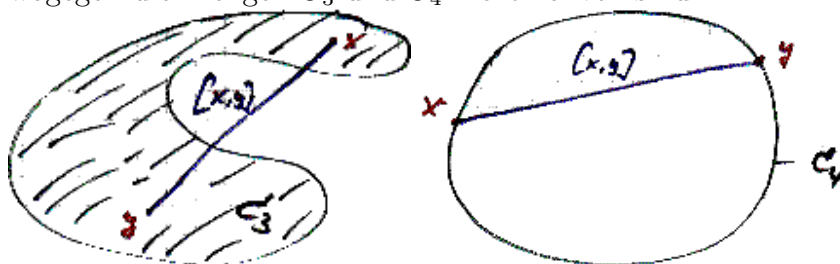
Sei $C \subset \mathbf{R}^n$. Wir sagen, daß C **konvex** ist, wenn bei $x, y \in C$ und $\lambda \in [0, 1]$ stets auch $(\lambda x + (1 - \lambda)y) \in C$.

Bemerkung.

1. C ist also konvex genau dann, wenn mit beliebigen Punkten x, y aus C auch der gesamte Abschnitt $[x, y]$ zu C gehört. Deshalb sind die Mengen C_1 und C_2 konvex,



wogegen die Mengen C_3 und C_4 nicht konvex sind.



2. Welche Punktmenge auf der Geraden $g = \{x \in \mathbf{R}^n : x = x^o + \alpha a, \alpha \in \mathbf{R}\}$, wobei $a \neq 0$, sind konvex ?

Durch Anwenden der Definition überzeugt man sich, daß

- (a) Ein Punkt $x^o + \alpha_0 a$ ($\alpha_0 = \text{fixiert}$) ist eine konvexe Menge.
- (b) Ein beliebiges Intervall

$$\{x = x^o + \alpha a : \alpha_1 < \alpha < \alpha_2\},$$

$$\{x = x^o + \alpha a : \alpha_1 \leq \alpha \leq \alpha_2\},$$

$$\{x = x^o + \alpha a : \alpha_1 \leq \alpha < \alpha_2\},$$

$$\{x = x^o + \alpha a : \alpha_1 < \alpha \leq \alpha_2\},$$

ist eine konvexe Teilmenge von g .

- (c) Der Strahl $\{x = x^o + \alpha a : \alpha \geq \alpha_1\}$ ist eine konvexe Menge.
- (d) Die gesamte Gerade $g = \{x = x^o + \alpha a : -\infty < \alpha < +\infty\}$ selbst ist eine konvexe Menge.
- (e) Aber z.B. zwei disjunkte Intervalle $\{x = x^o + \alpha a : \alpha_1 \leq \alpha \leq \alpha_2 \vee \alpha_3 \leq \alpha \leq \alpha_4\}$ mit $\alpha_1 \leq \alpha_2 < \alpha_3 \leq \alpha_4$ ist keine konvexe Menge.

3. Typische konvexe Mengen des \mathbf{R}^n sind:

- (a) Der \mathbf{R}^n selbst ist konvex.

- (b) Jeder Unterraum $L \subset \mathbf{R}^n$ ist konvex.
- (c) Jede affine Mannigfaltigkeit $N = \{x^o\} + L$ (wobei $x^o \in \mathbf{R}^n$, L ein Unterraum des \mathbf{R}^n) ist eine konvexe Menge.
- (d) Ebenen $\epsilon = \{x = x^o + \alpha a + \beta b : \alpha, \beta \in \mathbf{R}\}$ (wobei $x^o, a, b \in \mathbf{R}^n$, a und b fixiert sowie linear unabhängig) sind konvexe Mengen.
- (e) Hyperebenen $\pi = \{x \in \mathbf{R}^n : \langle p, x \rangle = \alpha\}$ (wobei $p \neq 0, \alpha \in \mathbf{R}$, α und p fixiert) sind konvex.
- (f) Die durch die Hyperebene $\pi = \{x \in \mathbf{R}^n : \langle p, x \rangle = \alpha\}$ erzeugten Halbräume

$$\begin{aligned} & \{x \in \mathbf{R}^n : \langle p, x \rangle \geq \alpha\}, \\ & \{x \in \mathbf{R}^n : \langle p, x \rangle > \alpha\}, \\ & \{x \in \mathbf{R}^n : \langle p, x \rangle \leq \alpha\}, \\ & \{x \in \mathbf{R}^n : \langle p, x \rangle < \alpha\}, \end{aligned}$$

sind konvex.

- 4. Per Definition sei bei uns die leere Menge konvex.
- 5. Die in der Linearen Optimierung eingeführten polyedralen Mengen (vgl. Abschnitt 1.4.1) sind konvex (polyedrale Menge = Durchschnitt von endlich vielen abgeschlossenen Halbräumen und Hyperebenen).

2.1.2 Algebraische Eigenschaften konvexer Mengen

Satz (2.1).

Der Durchschnitt von konvexen Mengen ist wieder eine konvexe Menge. Genauer: Es sei I eine Indexmenge (endlich oder abzählbar oder auch überabzählbar). Für beliebiges $i \in I$ sei $C_i \in \mathbf{R}^n$ eine konvexe Menge. Dann ist auch

$$C := \bigcap_{i \in I} C_i$$

konvex.

Satz (2.2).

Die algebraische Summe von konvexen Mengen ist wieder eine konvexe Menge. Genauer: Es sei $m \in \mathbf{N}$. Für beliebiges $i = 1, 2, \dots, m$ sei $C_i \subset \mathbf{R}^n$ eine

konvexe Menge und $\alpha_i \in \mathbf{R}$ eine gegebene Zahl. Dann ist auch

$$C := \sum_{i=1}^m \alpha_i C_i$$

$$=_{Def} \{x \in \mathbf{R}^n : \text{es existieren } x^i \in C_i, i = 1, 2, \dots, m \text{ so, daß } x = \sum_{i=1}^m \alpha_i x^i\}$$

eine konvexe Menge.

Folgerung (2.2.1).

Wenn $C_1, C_2 \subset \mathbf{R}^n$ konvexe Mengen sind, dann sind auch $C_1 + C_2, C_1 - C_2, C_1 - C_1$ konvexe Mengen. Beachten Sie dabei, daß bei $C_1 \neq \emptyset$ und C_1 keine einelementige Menge gilt, $C_1 - C_1 \neq \emptyset, C_1 - C_1 \neq \{0\}$.

Bemerkung.

Im allgemeinen liefert die Vereinigung konvexer Mengen keine konvexe Menge.

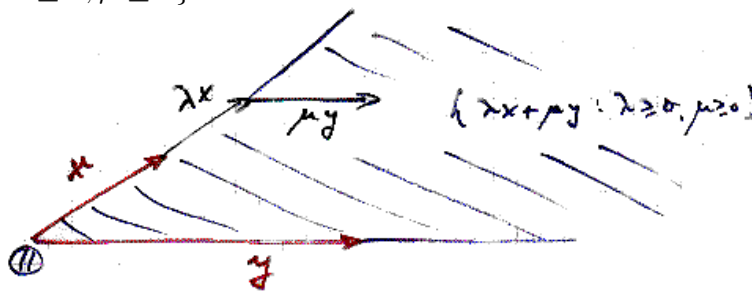
Eine typische unbeschränkte konvexe Menge ist der konvexe Kegel.

Definition.

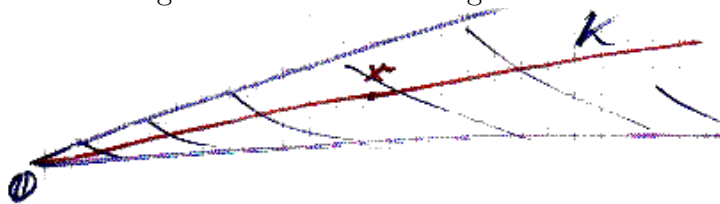
Die Menge $K \subset \mathbf{R}^n$ heißt **konvexer Kegel**, wenn mit $x, y \in K, \lambda \geq 0, \mu \geq 0$ stets auch $(\lambda x + \mu y) \in K$.

Bemerkung.

1. Mit 2 Elementen x, y muß ein konvexer Kegel auch die Menge $\{\lambda x + \mu y : \lambda \geq 0, \mu \geq 0\}$ enthalten.



2. Da mit $x \in K$ auch $0 = 0 \cdot x \in K$, so ist nach der 1. Bemerkung auch der in 0 beginnende und durch x gehende Strahl in K enthalten.



3. Der

$$\mathbf{R}_+^n = \{x \in \mathbf{R}^n : x \geq O\} \text{ und}$$

$$K = \{x \in \mathbf{R}^n : Ax \leq 0\}, \quad A - \text{Matrix vom Typ } [m, n]$$

sind typische Kegel (sogenannte polyedrale Kegel)

2.1.3 Die lineare konvexe Hülle

Definition.

Es seien $x^1, x^2, \dots, x^m \in \mathbf{R}^n, \lambda_1, \lambda_2, \dots, \lambda_m \in \mathbf{R}$. Wir betrachten das Element $x = \sum_{i=1}^m \lambda_i x^i$.

Wenn die $\lambda_i \geq 0, i = 1, 2, \dots, m$, dann sagen wir, daß x eine **nichtnegative Linearkombination** der $\{x^i\}$ ist.

Wenn $\sum_{i=1}^m \lambda_i = 1$, dann sagen wir, daß x eine **affine Linearkombination** der $\{x^i\}$ ist.

Wenn die Linearkombination sowohl nichtnegativ als auch affin ist, dann sprechen wir von **konvexer Linearkombination** (d.h. $\lambda_i \geq 0, \forall i, \sum_{i=1}^m \lambda_i = 1$).

Die Begriffe "konvexe Menge" und "konvexer Kegel" hatten wir über die Linearkombination von 2 Elementen eingeführt. Wie der folgende Satz zeigt, hätten wir auch von beliebig vielen (aber endlich vielen) ausgehen können.

Hilfssatz (2.3).

Eine Menge ist genau dann konvex, wenn sie jede endliche konvexe Linearkombination ihrer Elemente enthält.

Eine Menge ist genau dann ein konvexer Kegel, wenn sie jede endliche nichtnegative Linearkombination ihrer Elemente enthält.

Eine Menge ist genau dann eine affine Mannigfaltigkeit, wenn sie jede affine Linearkombination ihrer Elemente enthält.

Definition.

Es sei $A \subset \mathbf{R}^n$.

Die Menge $\text{conv}(A) =_{Def} \bigcap_{\substack{C \supset A, \\ C \text{ konvex}}} C$ heißt **konvexe lineare Hülle** von A .

Die Menge $\text{cone}(A) =_{Def} \bigcap_{\substack{K \supset A, \\ K \text{ konvexer Kegel}}} K$ heißt **konvexe konische Hülle**

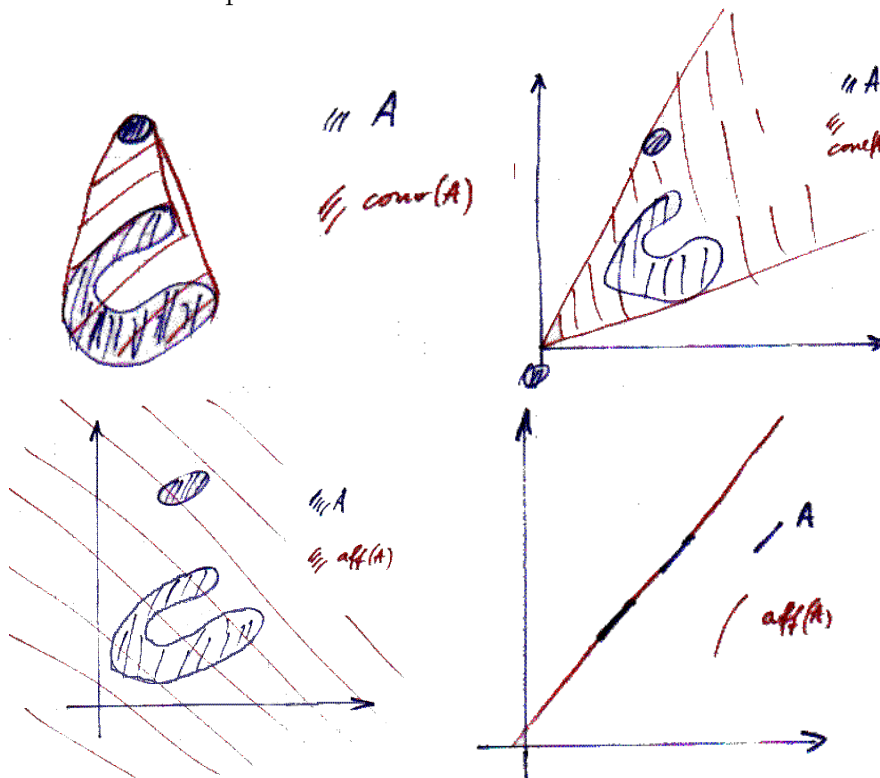
von A .

Die Menge $\text{aff}(A) =_{\text{Def}} \bigcap_{\substack{N \supset A, \\ N \text{ affine Mannigf.}}} N$ heißt **affine Hülle** von A .

Die Menge $\text{span}(A) =_{\text{Def}} \bigcap_{\substack{L \supset A, \\ L \text{ Unterraum}}} L$ heißt **lineare Hülle** von A .

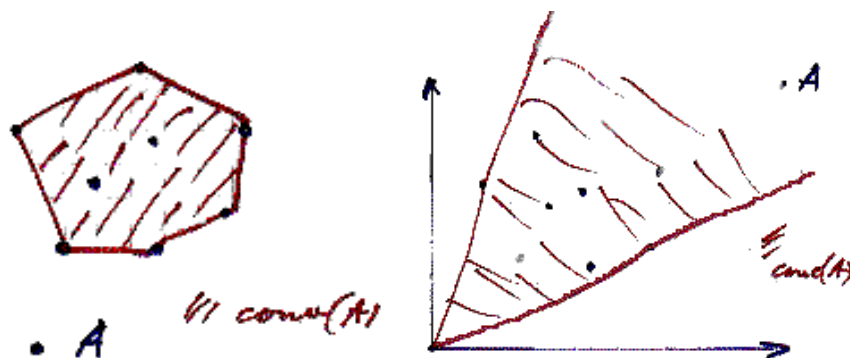
Bemerkung.

1. Illustrationsbeispiele im \mathbb{R}^2 :



2. Unabhängig davon, ob A konvex ist, $\text{conv}(A)$, $\text{cone}(A)$, $\text{aff}(A)$, $\text{span}(A)$ sind stets konvex. Außerdem ist $\text{cone}(A)$ ein konvexer Kegel, $\text{aff}(A)$ eine affine Mannigfaltigkeit, $\text{span}(A)$ ein Unterraum.

3. Wenn A eine endliche Menge ist, dann wird $\text{conv}(A)$ als **konvexes Polyeder** und $\text{cone}(A)$ als **polyedrischer konvexer Kegel** bezeichnet.



4. Zeigen Sie, daß

$$\text{conv}(A) = A \iff A \text{ ist konvex}$$

$$\text{cone}(A) = A \iff A \text{ ist ein konvexer Kegel}$$

$$\text{aff}(A) = A \iff A \text{ ist eine affine Mannigfaltigkeit}$$

$$\text{span}(A) = A \iff A \text{ ist ein Unterraum}$$

5. Wenn N eine lineare Mannigfaltigkeit ist und $x^o \in N$, dann ist $L =_{Def} N - \{x^o\}$ der zu N parallele Unterraum.

Der zu $\text{aff}(A)$ gehörende parallele Unterraum wird mit $\text{Lin}(A)$ bezeichnet ($\text{Lin}(A) = \text{aff}(A) - \{x^o\}$, wobei $x^o \in A$ möglich).

Wir benötigen später folgende Eigenschaften:

- Wenn $0 \in A$, dann ist $\text{Lin}(A) = \text{aff}(A) = \text{span}(A)$.
- Wenn $x \in A, y, z \in \text{Lin}(A), \alpha, \beta \in \mathbf{R}$, dann ist $(x + \alpha y + \beta z) \in \text{aff}(A)$.
- Wenn $x, y \in A$, dann ist $(x - y) \in \text{Lin}(A)$.

Hilfssatz (2.4).

Es sei N eine affine Mannigfaltigkeit des \mathbf{R}^n .

Dann existieren ein $m \in N$ und eine Matrix A vom Typ $[m, n]$ sowie ein Vektor $b \in \mathbf{R}^n$, so daß

$$N = \{x \in \mathbf{R}^n : Ax = b\}.$$

Folgerung.

Wenn A im Hilfssatz nicht die Nullmatrix ist (d.h. $N \neq \mathbf{R}^n$, dann ist $\dim\{x : Ax = 0\} = n - \text{rank}(A) < n$, d.h. $\dim(\text{Lin}(N)) < n$ und damit ist $\text{int}(N) = \emptyset$.

Eine affine Mannigfaltigkeit des \mathbf{R}^n besitzt folglich nur dann innere Punkte, wenn sie mit dem gesamten Raum \mathbf{R}^n identisch ist.

Satz (2.5).

Es sei $A \subset \mathbf{R}^n$. Dann gilt:

1. $\text{conv}(A)$ ist die Menge aller endlichen konvexen Linearkombinationen von Elementen aus A .
2. $\text{cone}(A)$ ist die Menge aller endlichen nichtnegativen Linearkombinationen von Elementen aus A .
3. $\text{aff}(A)$ ist die Menge aller endlichen affinen Linearkombinationen von Elementen aus A .
4. $\text{span}(A)$ ist die Menge aller endlichen Linearkombinationen von Elementen aus A .

Nach dem Satz ist bei $A \subset \mathbf{R}^n$ jeder Punkt aus $\text{conv}(A)$ als endliche konvexe Linearkombination von Punkten aus A darstellbar. Eines der fundamentalen Ergebnisse der konvexen Analysis ist, daß man sich hierbei auf Linearkombinationen von höchstens $n + 1$ Elementen aus A beschränken kann.

Hilfssatz (2.6).

Es seien m Zahlen $\lambda_i \geq 0, i = 1, 2, \dots, m$ mit $\sum_{i=1}^m \lambda_i = 1$ und m Punkte $x^i \in \mathbf{R}^n, i = 1, 2, \dots, m$ gegeben. Es sei $x = \sum_{i=1}^m \lambda_i x^i$ und $m > n + 1$.

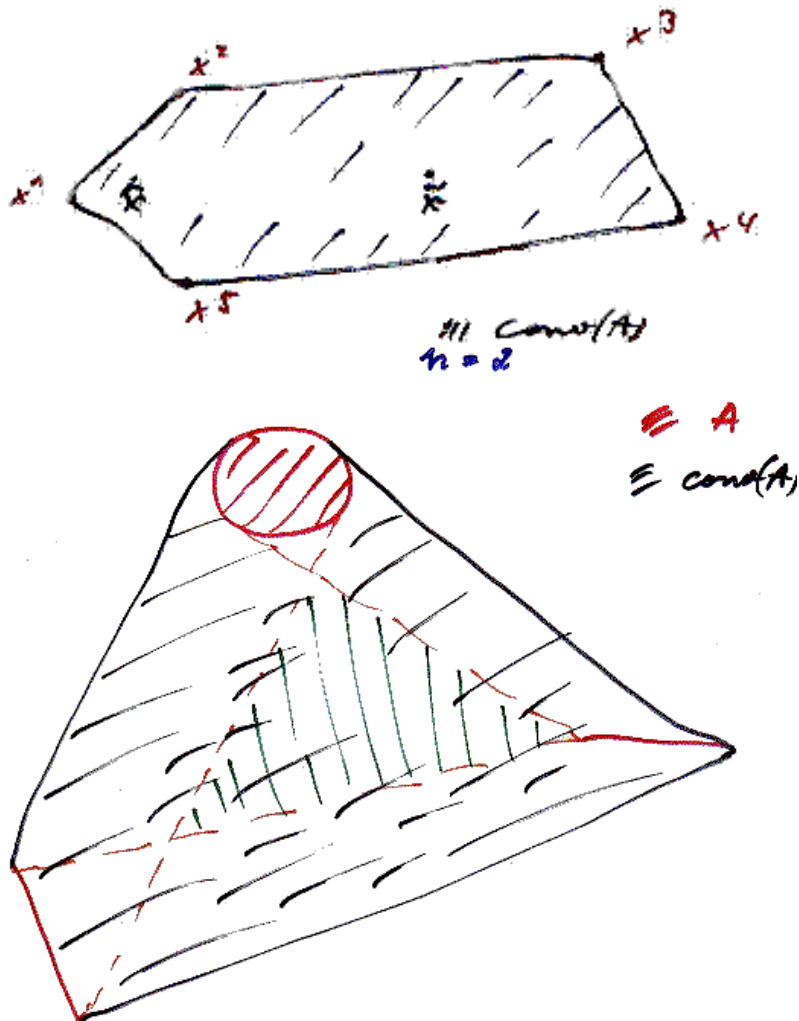
Dann existieren Zahlen $\alpha_i \geq 0, i = 1, 2, \dots, m$ mit $\sum_{i=1}^m \alpha_i = 1$, wobei wenigstens eines der α_i verschwindet, so daß $x = \sum_{i=1}^m \alpha_i x^i$ (d.h. bei $m > n + 1$ ist x auch als konvexe Linearkombination von weniger als der gegebenen m Punkte x^1, x^2, \dots, x^m darstellbar).

Theorem (2.7). (C. Caratheodory, 1907)

Es sei $A \subset \mathbf{R}^n$ und $x \in \text{conv}(A)$. Dann ist x eine konvexe Linearkombination von höchstens $(n + 1)$ Punkten aus A .

Bemerkung.

1. Warnung: Der Satz sagt nicht, daß man zur Darstellung verschiedener Punkte $x \in \text{conv}(A)$ immer auf dieselben $n + 1$ Punkte aus A zurückgreifen kann.
2. Beispiele: $A = \{x^1, x^2, x^3, x^4, x^5\}, n = 2$
 \hat{x} ist konvexe LK von x^1, x^2, x^5 .
 \tilde{x} ist konvexe LK von x^3, x^4, x^5 .



Punkte in doppelt schraffiertem Bereich sind nur über 3 Elemente aus A darstellbar, die anderen Punkte aus $\text{conv}(A)$ sind über 2 Punkte darstellbar.

Folgerung (2.8).

Es sei $A \subset \mathbf{R}^n$. Wenn A abgeschlossen und beschränkt ist, dann ist es auch $\text{conv}(A)$.

2.1.4 Topologische Eigenschaften konvexer Mengen

Im weiteren sei $U_\epsilon(x^o) = \{x \in \mathbf{R}^n : \|x - x^o\| \leq \epsilon\}$. Für $A \subset \mathbf{R}^n$ sei $\text{cl}(A) = \{x \in \mathbf{R}^n : \forall \epsilon > 0 \text{ ist } U_\epsilon(x) \cap A \neq \emptyset\}$, und $\text{int}(A) = \{x \in A : \exists \epsilon > 0, \text{ so daß } U_\epsilon(x) \subset A\}$.

$\text{cl}(A)$ werde **Abschließung** von A genannt, $\text{int}(A)$ soll **Inneres** von A heißen.

Satz (2.9).

Sei $C \subset \mathbf{R}^n$, C -konvex. Dann gilt:

1. $cl(C)$ ist konvex,
2. $int(C)$ ist konvex,
3. $aff(C) = aff cl(A)$.

Leider ist für viele in der Optimierung betrachteten konvexen Mengen C das Innere leer. Aus diesem Grund ist das sogenannte relative Innere von Interesse.

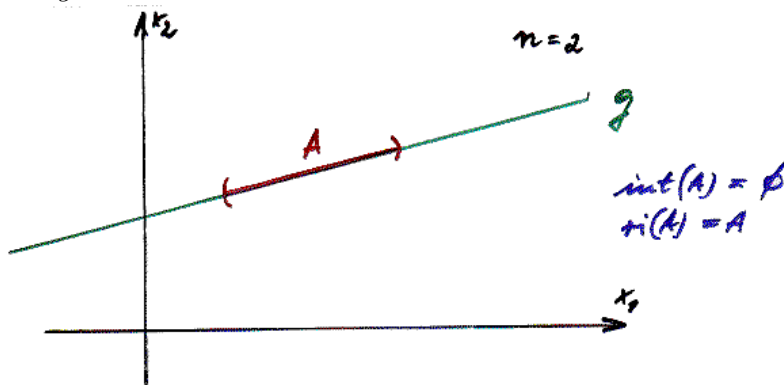
Definition.

Sei $x \in A \subset \mathbf{R}^n$. Wir sagen, daß x ein **relativ innerer Punkt** von A ist, wenn ein $\epsilon > 0$ existiert, so daß $(U_\epsilon(x) \cap aff(A)) \subset A$.

Die Menge aller relativ inneren Punkte von A wird mit $ri(A)$ bezeichnet und **relativ Inneres** von A genannt.

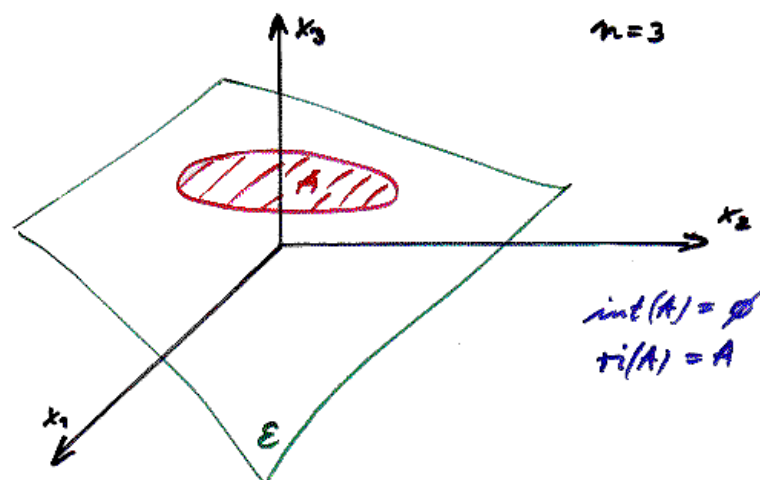
Bemerkung.

1. Relativ innere Punkte sind also innere Punkte in Bezug auf $aff(A)$, aber nicht unbedingt in Bezug auf den zugrunde gelegten Raum \mathbf{R}^n .
2. Sei g eine Gerade des \mathbf{R}^2 und A ein Intervall auf der Geraden.



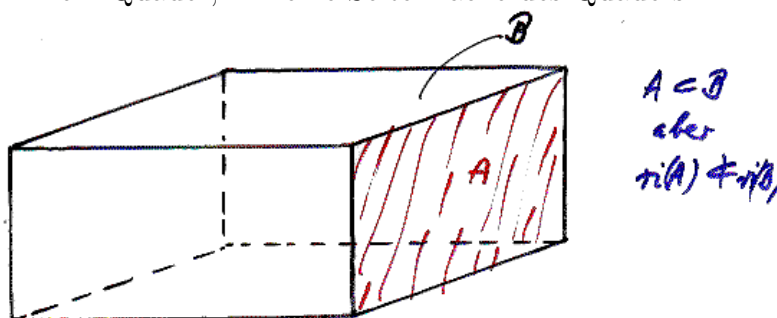
Dann ist $int(A) = \emptyset$, $aff(A) = g$ und damit $ri(A) = A$ ohne die Endpunkte des Intervalls.

3. Es sei ϵ eine Ebene des \mathbf{R}^3 und A eine in ϵ liegende Kreisscheibe.



Dann ist $\text{int}(A) = \emptyset$, $\text{aff}(A) = \epsilon$. Somit ergibt sich, daß $\text{ri}(A) = A$ ohne die Peripherie der Kreisscheibe.

4. Wenn $A \subset B$, dann ist $\text{cl}(A) \subset \text{cl}(B)$, $\text{int}(A) \subset \text{int}(B)$.
 Wie das folgende Beispiel zeigt, muß aber $\text{ri}(A) \subset \text{ri}(B)$ i.a. nicht gelten:
 B - ein Quader, A - eine Seitenfläche des Quaders.



Dann ist $\text{ri}(B) = \text{Quader ohne Seitenfläche}$, wogegen $\text{ri}(A) = \text{Seitenfläche ohne Kanten}$.

Damit ist kein einziger Punkt aus $\text{ri}(A)$ in $\text{ri}(B)$ enthalten, d.h. $\text{ri}(A) \not\subset \text{ri}(B)$.

5. $x^o \in \text{ri}(A)$ gilt genau dann, wenn aus $\{x^i\}_{i=1}^\infty \subset \text{aff}(A)$ mit $\lim_{i \rightarrow \infty} x^i = x^o$ stets folgt, daß es ein i_o gibt, so daß bei $i \geq i_o$ gilt $x^i \in A$.
6. Wir sagen, daß die Menge A **relativ offen** ist, wenn $A = \text{ri}(A)$.
7. Die Menge $\text{bd}(A) =_{\text{Def}} \text{cl}(A) \setminus \text{int}(A)$ wird **Rand** von A genannt. Analog nennt man die Menge $\text{rb}(A) =_{\text{Def}} \text{cl}(A) \setminus \text{ri}(A)$ **relativen Rand** von A .

8. Wenn $\text{int}(A) \neq \emptyset$, dann ist $\text{int}(A) = \text{ri}(A)$.

Satz (2.10).

Sei $C \subset \mathbf{R}^n, C$ - konvex, $C \neq \emptyset$. Dann ist $\text{ri}(C) \neq \emptyset$.

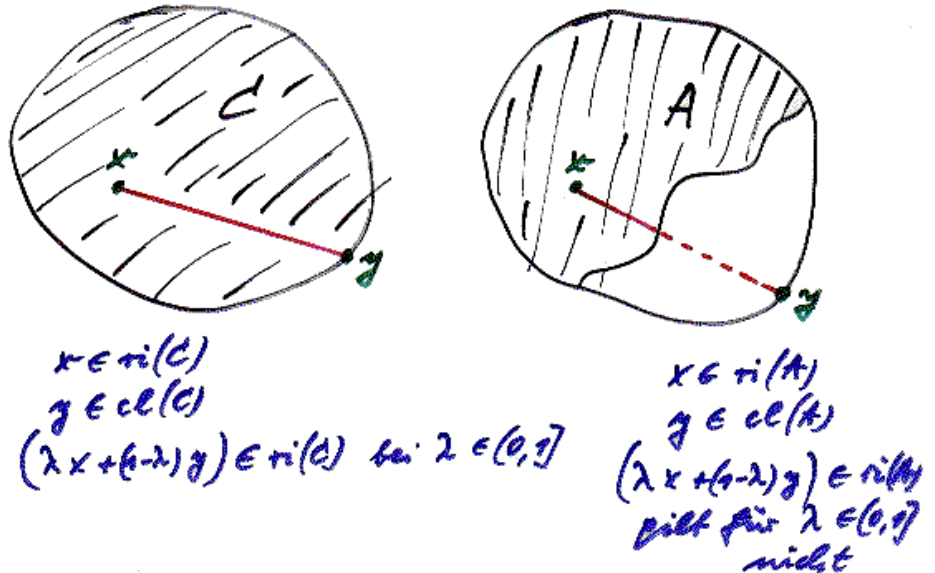
Lemma (2.11). (Zugänglichkeitslemma)

Sei $C \subset \mathbf{R}^n, C \neq \emptyset, C$ - konvex.

Wenn $x \in \text{ri}(C), y \in \text{cl}(C), \lambda \in (0, 1]$, dann ist $z =_{\text{Def}} (\lambda x + (1-\lambda)y) \in \text{ri}(C)$.

Bemerkung.

1. Unter der Bedingung des Satzes haben wir $[x, y) \subset \text{ri}(C)$.



2. Illustration:

Ohne Konvexität von C muß die Behauptung nicht gelten: $(\lambda x + (1-\lambda)y) \in \text{ri}(A), \forall \lambda \in (0, 1]$ ist nicht wahr, obwohl $x \in \text{ri}(A), y \in \text{cl}(A)$.

Folgerung (2.12).

Sei $C \subset \mathbf{R}^n, C$ -konvex. Dann ist auch $\text{ri}(C)$ konvex.

Hilfssatz (2.13).

Sei $C \subset \mathbf{R}^n, C \neq \emptyset, C$ - konvex. Dann gilt

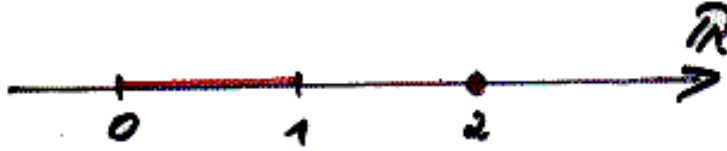
$$\begin{cases} \text{cl}(C) = \text{clri}(A) \\ \text{ri}(C) = \text{ricl}(C). \end{cases}$$

Bemerkung.

Ohne Konvexität müssen die Behauptungen nicht gelten:

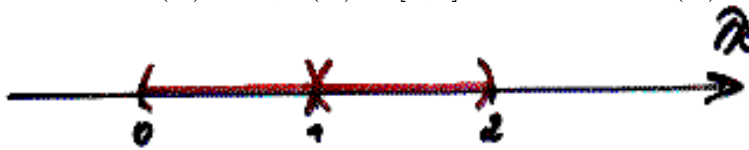
1. Sei $A = [0, 1] \cup \{2\} \subset \mathbf{R}$.

Dann ist $A = cl(A)$ und $clri(A) = [0, 1] \neq cl(A)$.



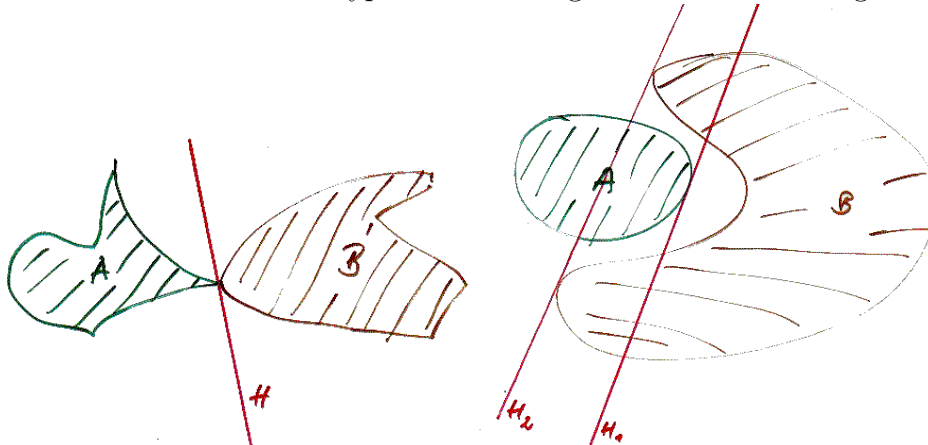
2. Sei $A = (0, 1) \cup (1, 2)$.

Dann ist $ri(A) = A$, $cl(A) = [0, 2]$ und damit $ricl(A) = (0, 2) \neq ri(A)$.



2.2 Trennbarkeit konvexer Mengen

In diesem Abschnitt geht es um die Existenz einer sogenannten trennenden Hyperebene. Aus geometrischer Sicht betrachtet ist es dabei das Ziel, eine Hyperebene so im Raum zu legen, daß zwei gegebene Mengen A und B in verschiedene durch die Hyperebene erzeugte Halbräume zu liegen kommen.



A und B trennbar

A und B nicht trennbar.

Es ist anschaulich klar, daß elementefremde konvexe Mengen trennbar sind. Dieser tiefe Fakt ist aber schwer zu beweisen und hat weitreichende Folgerungen. Viele Existenzaussagen lassen sich über Trennungssätze beweisen. Wir wählen einen Beweisweg, der mit den topologischen Eigenschaften der Norm im \mathbf{R}^n zusammenhängt.

Analoge Aussagen der Funktionalanalysis werden meist als Hahn-Banach-Theorem bezeichnet.

2.2.1 Die Projektion eines Punktes

Definition.

Sei $A \subset \mathbf{R}^n, a \in \mathbf{R}^n$. Unter der **Projektion** des Punktes a auf die Menge A verstehen wir einen Punkt $P_A(a)$ aus A mit der Eigenschaft $\|P_A(a) - a\| \leq \|x - a\|, \forall x \in A$.

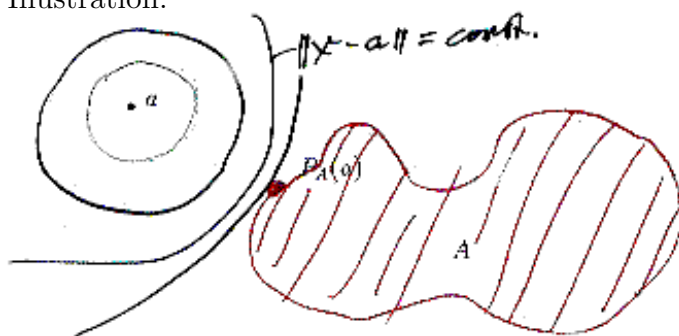
Bemerkung.

1. $P_A(a)$ ist damit optimale Lösung der Optimierungsaufgabe

$$\begin{cases} \|x - a\| \rightarrow \min \\ x \in A \end{cases} \quad \text{und damit auch der} \quad (\text{OA}) \begin{cases} \|x - a\|^2 \rightarrow \min \\ x \in A \end{cases}$$

2. Wenn $a \in A$, dann ist $P_A(a) = a$.

3. Illustration:

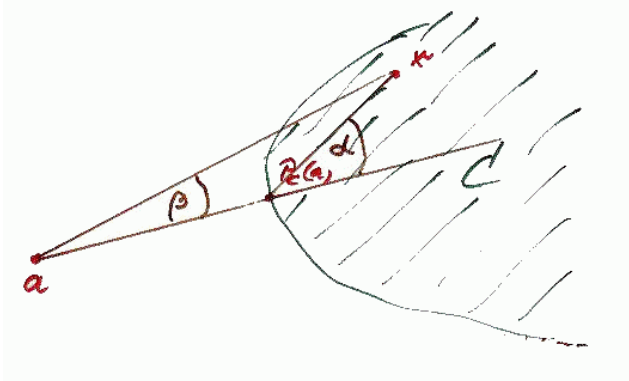


Hilfssatz (2.14).

Es sei $C \subset \mathbf{R}^n, C \neq \emptyset, C$ -konvex und abgeschlossen. Es sei $a \in \mathbf{R}^n, a \notin C$. Dann gilt:

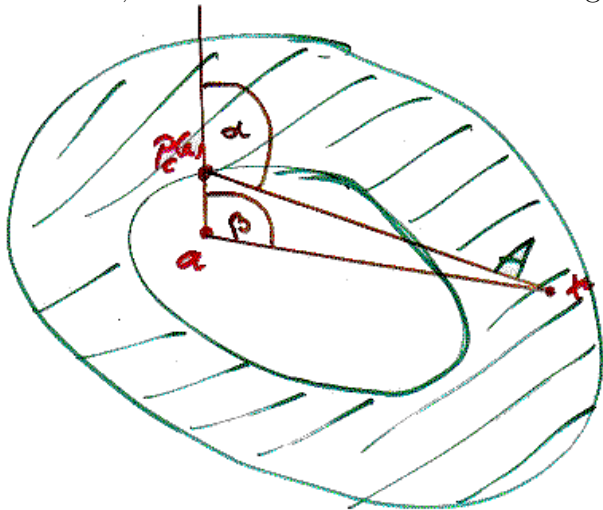
1. $P_C(a)$ existiert und ist eindeutig, $P_C(a) \neq a$,
2. $\langle P_C(a) - a, x - P_C(a) \rangle \geq 0, \forall x \in C$,
3. $\langle P_C(a) - a, x - a \rangle \geq \|P_C(a) - a\|^2 > 0, \forall x \in C$.

Illustration von (2) und (3):



Nach 2. ist α nichtstumpf. Nach 3. ist β spitz.

Es ist klar, daß bei nichtkonvexem C diese Eigenschaften nicht gelten müssen:



2.2.2 Der Begriff der Trennung

Definition.

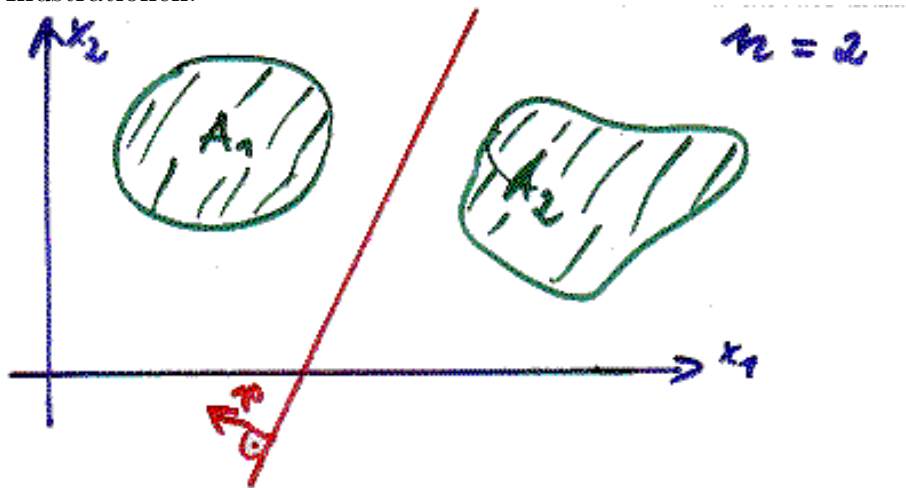
Seien $A_1, A_2 \subset \mathbf{R}^n$. Wir sagen, daß A_1 und A_2 **trennbar sind**, wenn ein $p \in \mathbf{R}^n, p \neq 0$ und ein $\beta \in \mathbf{R}$ existieren, so daß $\langle p, x \rangle \geq \beta \geq \langle p, y \rangle, \forall x \in A_1, \forall y \in A_2$.

Wir sagen, daß A_1 und A_2 **eigentlich trennbar** sind, wenn sie trennbar sind und dabei p so gewählt werden kann, daß ein $\bar{x} \in A_1$ und ein $\bar{y} \in A_2$ existieren, für die $\langle p, \bar{x} \rangle > \langle p, \bar{y} \rangle$.

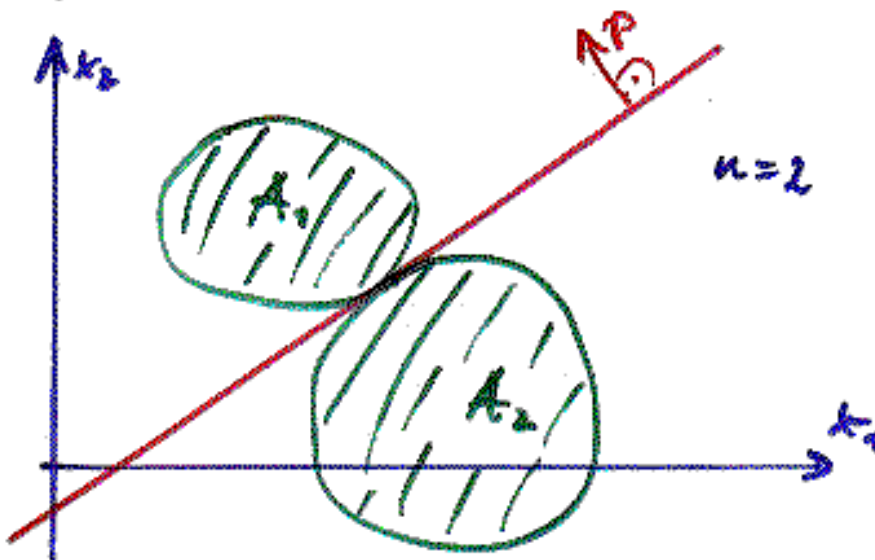
Schließlich sagen wir, daß A_1 und A_2 **stark trennbar** sind, wenn p und β so gewählt werden können, daß $\inf\{\langle p, x \rangle : x \in A_1\} > \beta > \sup\{\langle p, y \rangle : y \in A_2\}$.

Bemerkung.

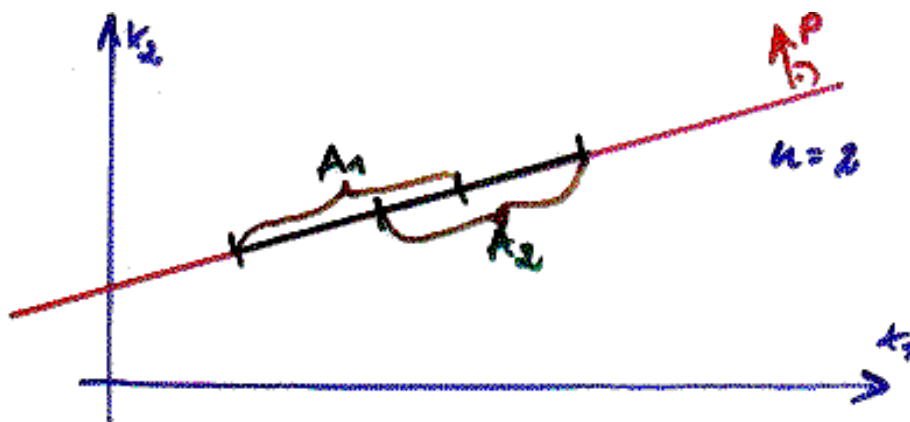
1. Die von uns betrachteten drei Arten der Trennbarkeit sind mit der Existenz einer trennenden bzw. eigentlich trennenden bzw. stark trennenden Hyperebene $H_p^\beta = \{x \in \mathbf{R}^n : \langle p, x \rangle = \beta\}$ verbunden. Geometrisch gesehen bedeutet die Trennbarkeit, daß eine Hyperebene existiert, so daß die Mengen A_1 und A_2 in verschiedene durch die Hyperebene erzeugte Halbräume zu liegen kommen.
2. Klar ist, daß: stark trennbar \implies eigentlich trennbar \implies trennbar.
3. Illustrationen:



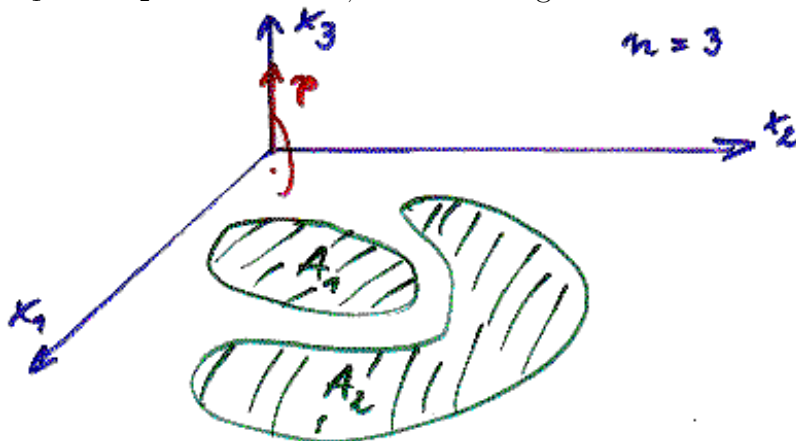
stark trennbar



eigentlich trennbar, aber nicht stark trennbar



A_1 und A_2 sind trennbar, aber nicht eigentlich trennbar



A_1 und A_2 liegen in der $x_1 - x_2$ -Ebene.

Als Mengen des zweidimensionalen $x_1 - x_2$ -Raumes sind A_1 und A_2 nicht trennbar. Als Mengen des \mathbf{R}^3 sind A_1 und A_2 trennbar, aber nicht eigentlich.

2.2.3 Trennungssätze

Theorem (2.15). (Notwendiges und hinreichendes Kriterium für starke Trennbarkeit)

Seien $C_1, C_2 \subset \mathbf{R}^n, C_1, C_2 \neq \emptyset, C_1, C_2$ -konvex.

C_1 und C_2 sind stark trennbar genau dann, wenn der Abstand $\rho(C_1, C_2)$ der beiden Mengen positiv ist.

Folgerung (2.16).

Seien $C_1, C_2 \subset \mathbf{R}^n, C_1, C_2 \neq \emptyset, C_1, C_2$ -konvex und abgeschlossen, wobei wenigstens eine der beiden Mengen beschränkt sei.

Dann gilt: Wenn $C_1 \cap C_2 = \emptyset$, dann sind die Mengen stark trennbar.

Folgerung (2.17).

Sei $C \subset \mathbf{R}^n, C \neq \emptyset, C$ -konvex und abgeschlossen. Sei $a \in \mathbf{R}^n, a \notin C$.

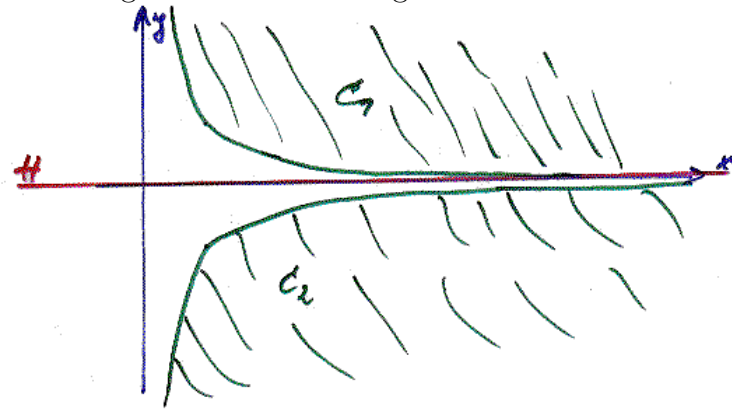
Dann existieren $p \in \mathbf{R}^n, \beta \in \mathbf{R}$, so daß $\inf\{\langle p, x \rangle : x \in C\} > \beta > \langle p, a \rangle$.

Bemerkung.

1. Wie das Beispiel

$$C_1 = \{[x_1, x_2] \in \mathbf{R}^2 : x_2 \geq \frac{1}{x_1}, x_1 > 0\},$$
$$C_2 = \{[x_1, x_2] \in \mathbf{R}^2 : x_2 \leq \frac{-1}{x_1}, x_1 > 0\}$$

zeigt, können wir in Folgerung 2.16 die Forderung nicht fallen lassen, daß wenigstens eine der Mengen beschränkt ist.

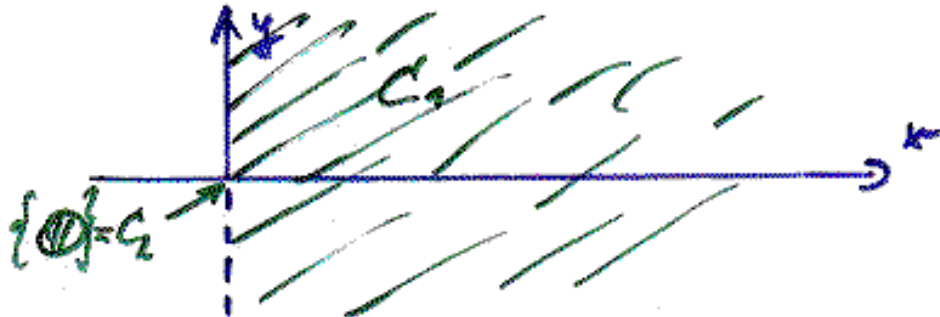


eigentlich trennbar, aber nicht stark trennbar, einzige trennende Hyper-ebene ist $\langle p, x \rangle = 0$ mit $p = [0, 1]$. $\inf_{x \in C_1} \langle p, x \rangle = \sup_{y \in C_2} \langle p, y \rangle = 0$.

2. Wie das Beispiel

$$C_1 = \{[x_1, x_2] \in \mathbf{R}^2 : x_1, x_2 \geq 0, x_1^2 + x_2^2 > 0\} \cup \{[x_1, x_2] \in \mathbf{R}^2 : x_1 > 0, x_2 < 0\},$$
$$C_2 = \{[0, 0]\}$$

zeigt, können wir in Folgerung 2.16 die Forderung nicht fallen lassen, daß die Mengen abgeschlossen sind.



C_1 und C_2 sind eigentlich trennbar, aber nicht stark trennbar. $H = \{[x_1, x_2] \in \mathbf{R}^2 : x_1 = 0\}$ ist die eigentlich trennende Hyperebene.

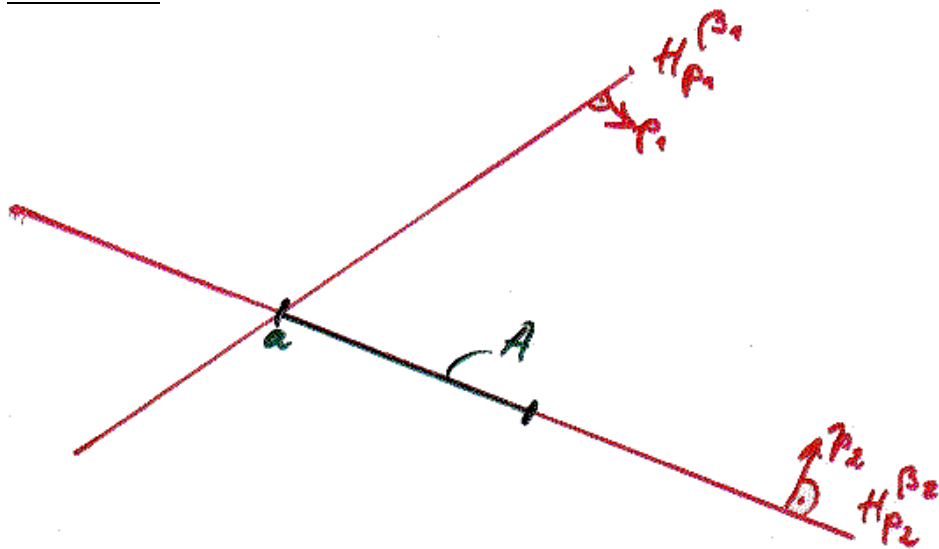
Definition.

Sei $p \neq 0, p \in \mathbf{R}^n$. Die Hyperebene $H_p^\beta = \{x \in \mathbf{R}^n : \langle p, x \rangle = \beta\}$ heißt **Stützhyperebene** an die Menge $A \subset \mathbf{R}^n$ im Punkte $a \in cl(A)$, wenn

$$(\circ) \begin{cases} \langle p, x \rangle \geq \beta & , \forall x \in A \\ \langle p, a \rangle = \beta \end{cases}$$

Wir sagen, daß H_p^β eine **eigentliche Stützhyperebene** ist, wenn zusätzlich zu (o) noch die Existenz eines $\bar{x} \in A$ gesichert ist, für das $\langle p, \bar{x} \rangle > \beta$ gilt.

Illustration: Sei $n = 2$.



$H_{p_1}^{\beta_1}$ ist eine eigentliche Stützhyperebene an A im Punkt a ,
 $H_{p_2}^{\beta_2}$ ist eine Stützhyperebene an A in a , aber keine eigentliche.

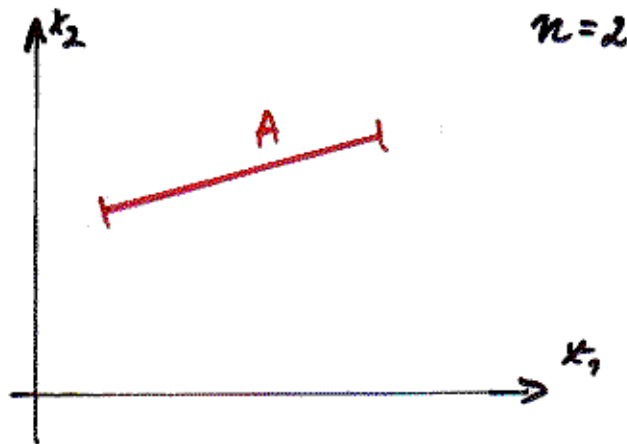
Im Abschnitt 2.1.4 hatten wir die Begriffe des Randes und des relativen Randes einer Menge eingeführt:

$$\begin{aligned} cl(A) \setminus int(A) &= \text{Rand von } A, \\ cl(A) \setminus ri(A) &= \text{relativer Rand von } A. \end{aligned}$$

Da $int(A) \subset ri(A)$ stets gilt, so ist der relative Rand stets Teil des Randes. Wenn $int(A) \neq \emptyset$, dann gilt $int(A) = ri(A)$ und damit ist der Rand identisch mit dem relativen Rand.

Wenn aber $int(A) = \emptyset$, dann können sich Rand und relativer Rand von A stark unterscheiden.

Beispiel: $n = 2$



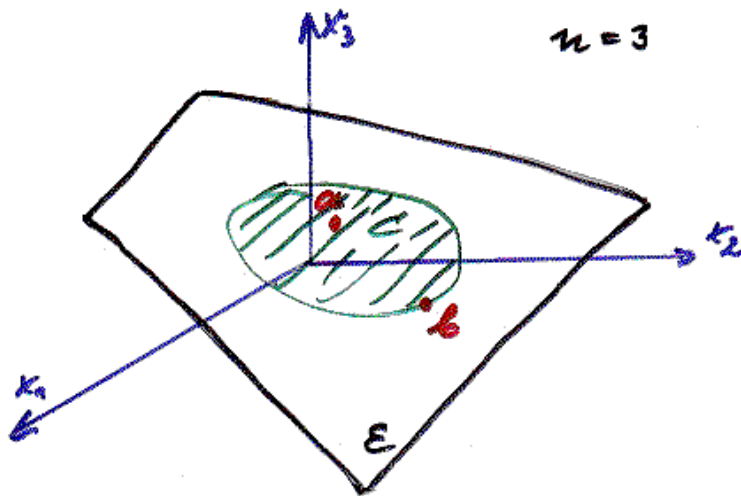
Der relative Rand von A besteht aus den Endpunkten des Abschnittes A . Der Rand von A ist identisch mit A .

Satz (2.18). (Existenz einer Stützhyperebene)

Sei $C \subset \mathbf{R}^n$, C -konvex, $C \neq \emptyset$. Dann gilt:

1. In einem Randpunkt von C existiert eine Stützhyperebene an C .
2. In einem relativen Randpunkt von C existiert eine eigentliche Stützhyperebene an C .

Illustration: Sei $n = 3$



Sei C eine konvexe Menge der Ebene ϵ . Wir haben $a \in (cl(C) \setminus int(C))$, da $int(C) = \emptyset$. Einzige Stützhyperebene in a an C ist die Ebene ϵ , die aber nicht eigentlich stützt.

Weiter haben wir $b \in (cl(C) \setminus ri(C))$. Mit jeder von ϵ verschiedenen, durch b verlaufenden Ebene des \mathbf{R}^3 , bei der C auf einer Seite liegt, ist eine eigentliche Stützhyperebene an C verbunden.

Theorem (2.19). (Notwendiges und hinreichendes Kriterium der eigentlichen Trennbarkeit)

Seien $C_1, C_2 \subset \mathbf{R}^n$, C_1, C_2 -konvex, $C_1, C_2 \neq \emptyset$.

Dann gilt: C_1 und C_2 sind genau dann eigentlich trennbar, wenn $ri(C_1) \cap ri(C_2) = \emptyset$.

2.2.4 Drei erste Anwendungen der Trennungssätze

Theorem (2.20). (Farkas, 1902)

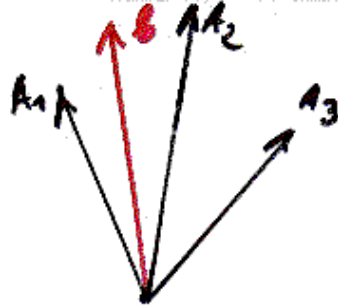
Es sei A eine Matrix vom Typ $[m, n]$, $b \in \mathbf{R}^m$. Dann gilt: Eines und genau eines der beiden Systeme (+) und (++) besitzt eine Lösung:

$$(+) \begin{cases} Ax = b \\ x \geq 0 \end{cases} \quad (++) \begin{cases} pA \geq 0 \\ \langle p, b \rangle < 0 \end{cases}$$

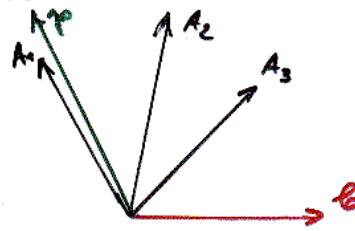
Bemerkung.

- Geometrisch können wir Theorem 2.20 wie folgt interpretieren:
Den Vektor $b \in \mathbf{R}^m$ und die Spalten der Matrix A können wir als Vektoren des \mathbf{R}^m interpretieren. Dann gilt genau eine der Alternativen:
 - Alternative I: b ist eine nichtnegative Linearkombination der Spalten von A ((+)) ist lösbar).

- Alternative II: Es gibt einen Vektor $p \in \mathbf{R}^m$, der mit allen Spalten von A einen nichtstumpfen Winkel bildet, mit b aber einen stumpfen Winkel bildet ((++) ist lösbar).



(+) lösbar
(++) nicht lösbar



(+) nicht lösbar
(++) lösbar

2. Ein Analogon des Satzes läßt sich für alle linearen Systeme zeigen (d.h., sie ist nicht an die spezielle Form von (+) gebunden). So gilt, z.B. auch die Aussage, daß von den zwei Systemen

$$Ax \leq b, \quad \begin{cases} pA = 0 \\ \langle p, b \rangle < 0 \\ p \geq 0 \end{cases}$$

stets genau eines lösbar ist.

Satz (2.21). (Gordan, 1973)

Es sei A eine Matrix vom Typ $[m, n]$ und B eine Matrix vom Typ $[l, n]$. Von den beiden Systemen

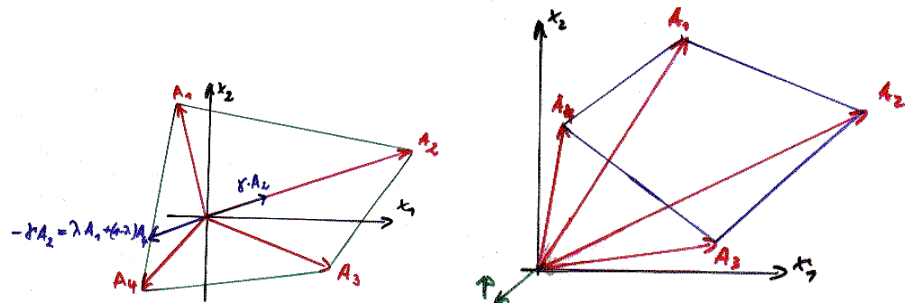
$$(\circ) \begin{cases} yA + zB = 0 \\ y \geq 0, y \neq 0 \end{cases} \quad (\circ\circ) \begin{cases} Ad < 0 \\ Bd = 0 \end{cases}$$

hat stets genau eine Lösung.

Bemerkung.

1. In $(\circ\circ)$ soll in jeder Komponente der Vektorungleichung $Ad < 0$ strenge Ungleichheit gelten.
2. Geometrisch kann man den Satz von Gordon für den Fall $B = 0$ so interpretieren: Es gilt stets genau eine von 2 Alternativen

- Alternative I: Der Ursprung kann als nichttriviale, nichtnegative Linearkombination der Zeilen A dargestellt werden ((\circ) lösbar).
- Alternative II: Es existiert ein Vektor p , der mit allen Zeilen von A echt stumpfe Winkel bildet ($(\circ\circ)$ lösbar).



(\circ) lösbar,

z.B.: $0 = \gamma A_2 + \lambda A_1 + (1 - \lambda)A_4 + oA_3$

$(\circ\circ)$ lösbar

Wir betrachten vorübergehend wieder mal die allgemeine Optimierungsaufgabe:

Für $S \subset \mathbf{R}^n$ und $f : S \rightarrow \mathbf{R}$ sei

$$(OA) \quad f^* = \inf\{f(x) : x \in S\}.$$

Bezüglich der Lösbarkeit können wir vier Möglichkeiten unterscheiden:

- 1. Möglichkeit: $S = \emptyset$, d.h. $f^* = +\infty$ und $\nexists x^* \in S$, so daß $f(x^*) = f^*$.
- 2. Möglichkeit: $S \neq \emptyset$, $f^* = -\infty$ und damit $\nexists x^* \in S$, so daß $f(x^*) = f^*$.
- 3. Möglichkeit: $S \neq \emptyset$, $f^* > -\infty$ und $\exists x^* \in S$, so daß $f(x^*) = f^*$.
- 4. Möglichkeit: $S \neq \emptyset$, $f^* > -\infty$ und $\nexists x^* \in S$, so daß $f(x^*) = f^*$.

Theorem 2.22 zeigt, daß bei LOA die vierte Möglichkeit ausgeschlossen ist.

Theorem (2.22). (Lösbarkeit für LOA)

Wir betrachten die LOA

$$(K) \quad f^* = \inf\{\langle c, x \rangle : Ax = b, x \geq 0\},$$

wobei $b \in \mathbf{R}^m$, $c \in \mathbf{R}^n$ gegebene Vektoren und A eine gegebene Matrix vom Typ $[m, n]$ seien. Es sei $S = \{x \in \mathbf{R}^n : Ax = b, x \geq 0\} \neq \emptyset$ und $f^* > -\infty$.

Dann existiert eine optimale Lösung x^* von (K).

Bemerkung.

Die Aussage ist nicht an die (K)-Form der LOA gebunden, sondern gilt für beliebiges polyedrales S und lineares f in (OA).

2.3 Konvexe Funktionen

2.3.1 Der Begriff der konvexen Funktion

Definition.

Es sei $S \subset \mathbf{R}^n$, $f : S \rightarrow \mathbf{R}$, $\bar{x} \in S$. Wir sagen, daß f in \bar{x} bzgl. S eine **konvexe Funktion** ist, wenn aus

$$\begin{cases} y \in S \\ \lambda \in [0, 1] \\ (\lambda\bar{x} + (1 - \lambda)y) \in S \end{cases}$$

stets folgt, daß

$$(\circ) \quad f(\lambda\bar{x} + (1 - \lambda)y) \leq \lambda f(\bar{x}) + (1 - \lambda)f(y).$$

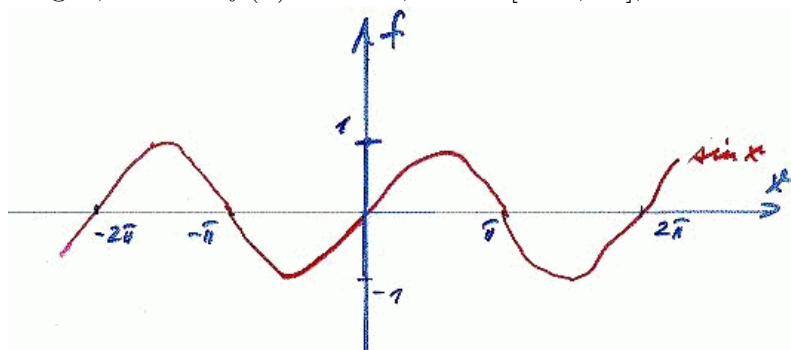
Wenn in (\circ) bei $y \neq \bar{x}$, $\lambda \in (0, 1)$ stets echte Ungleichheit gilt, dann sprechen wir von **strenger Konvexität**.

Wenn f in beliebigem $\bar{x} \in S$ konvex bzgl. S ist, dann sagen wir, daß f auf S konvex ist.

Wenn $F(x) = -f(x)$ in $\bar{x} \in S$ bzgl. S konvex ist, dann sagen wir, daß f in \bar{x} bzgl. S **konkav** ist.

Bemerkung.

1. Man sollte nicht denken, daß eine Funktion entweder konvex oder konkav ist (d.h., es gibt Funktionen, die weder konvex noch konkav sind). So gilt, z.B. für $f(x) := \sin x$, $S = [-2\pi, 2\pi]$,

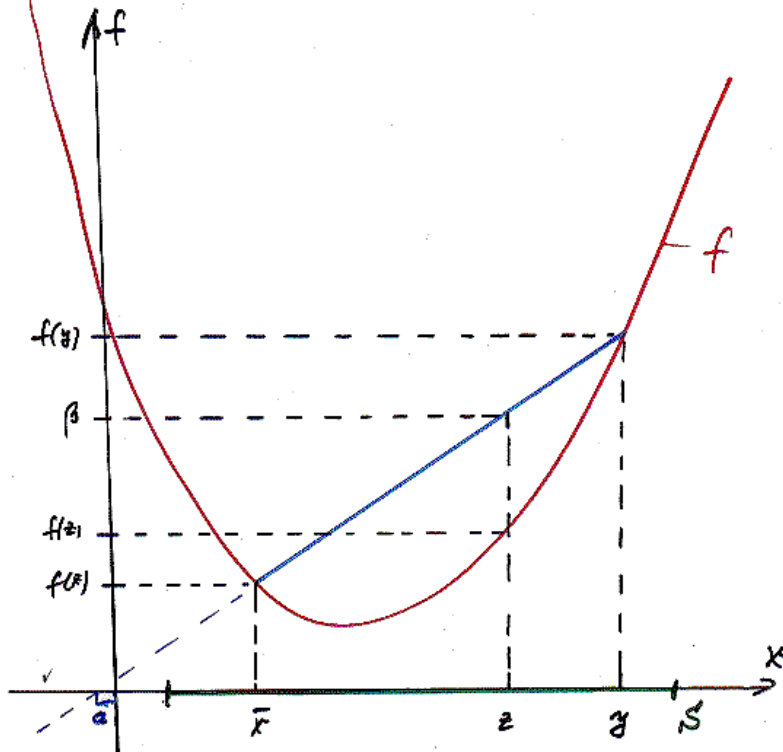


daß f auf S weder konvex noch konkav ist, aber z.B. auf $[\pi, 2\pi]$ konvex und auf $[0, \pi]$ konkav ist (den exakten Nachweis können wir leichter erst zu einem späteren Zeitpunkt führen, z.B. mit Hilfe differentieller Konvexitätskriterien).

2. In unserem Kursus wird meist die Menge S aus der Definition eine konvexe Menge sein. Dann ist die Forderung $(\lambda\bar{x} + (1 - \lambda)y) \in S$ bei $\lambda \in [0, 1]$ automatisch erfüllt.

Wir haben in der obigen Definition absichtlich S als nichtkonvexe Menge erlaubt, um z.B. die Konvexität einer Funktion auf einer diskreten Menge S betrachten zu können.

3. Geometrische Illustration im Fall $n = 1$. Gegeben sei $S \subset \mathbf{R}^1$, $f : S \rightarrow \mathbf{R}$.



Nach dem Strahlensatz ist in der Skizze: $\frac{a+\bar{x}}{f(\bar{x})} = \frac{a+z}{\beta} = \frac{a+y}{f(y)}$, woraus wir z.B. erhalten:

$$f(\bar{x}) = \frac{a + \bar{x}}{a + z} \cdot \beta \quad (\Delta)$$

$$f(y) = \frac{a + y}{a + z} \cdot \beta \quad (\Delta\Delta)$$

Da in der Skizze $z \in [\bar{x}, y]$, so existiert ein $\lambda \in [0, 1]$, so daß $z = \lambda\bar{x} + (1 - \lambda)y$.

Nach der Definition bedeutet die Konvexität von f in \bar{x} bzgl S , daß u.a.

$$f(z) \leq \lambda f(\bar{x}) + (1 - \lambda)f(y) \quad (\square).$$

Wir haben in der Skizze

$$\begin{aligned} \lambda f(\bar{x}) + (1 - \lambda)f(y) &\stackrel{(\Delta)(\Delta\Delta)}{=} \frac{\beta}{a + z} [\lambda(a + \bar{x}) + (1 - \lambda)(a + y)] \\ &= \frac{\beta}{a + z} [\lambda\bar{x} + (1 - \lambda)y + a] \stackrel{z = \lambda\bar{x} + (1 - \lambda)y}{=} \frac{\beta}{a + z} (a + z) \\ &= \beta \end{aligned}$$

Somit bedeutet (\square)

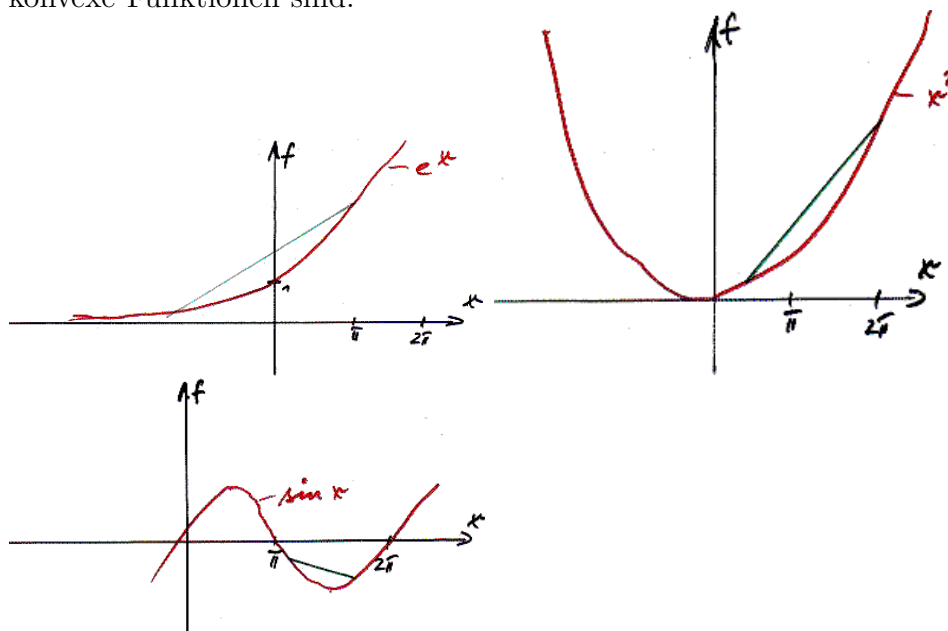
$$f(z) \leq \beta,$$

d.h. bei konvexen Funktionen im \mathbf{R}^1 liegt die Sekante (= Verbindung von $[\bar{x}, f(\bar{x})]$ mit $[y, f(y)]$) nirgendwo unterhalb der Funktion.

4. Ausgehend von der in 3. gegebenen geometrischen Interpretation ist, anschaulich klar, daß z.B. auf $[\pi, 2\pi]$

$$f(x) = e^x, \quad f(x) = x^2, \quad f(x) = \sin x$$

konvexe Funktionen sind:



Einen einfachen formalen Beweis, daß das wirklich so ist, können wir erst später führen (z.B über die zweite Ableitung, vgl. Satz 2.30).

5. Es sei $a \in \mathbf{R}^n$, a - fixiert. Sei $\alpha \in \mathbf{R}$. Wir betrachten $f(x) = \langle a, x \rangle + \alpha$. Offensichtlich ist

$$\begin{aligned} f(\lambda x + (1 - \lambda)y) &= \langle a, \lambda x + (1 - \lambda)y \rangle + \alpha \\ &= \langle a, \lambda v + (1 - \lambda)y \rangle + \lambda\alpha + (1 - \lambda)\alpha \\ &= \lambda[\langle a, y \rangle + \alpha] + (1 - \lambda)[\langle a, y \rangle + \alpha] \\ &= \lambda f(x) + (1 - \lambda)f(y), \forall \lambda \in \mathbf{R}^1. \quad (\square\square) \end{aligned}$$

Also ist f im \mathbf{R}^1 konvex. Wegen der Gleichheit ($\square\square$) ist auch $-f(x)$ konvex, d.h. unser f ist auch konkav..

6. Sei $C \subset \mathbf{R}^n$, C -konvex, $f : C \rightarrow \mathbf{R}$, $\alpha \in \mathbf{R}$. Wir setzen $T_\alpha^f = \{z \in C : f(z) \leq \alpha\}$.
 T_α^f ist die **sogenannte Niveaumenge** von f zum Niveau α .
 Dann gilt: Wenn f auf C konvex ist, dann ist T_α^f eine konvexe Menge.
7. Nach der 6. Bemerkung ist eine konvexe Funktion stets auch quasikonvex.
 Genauer: Ist f in $\bar{x} \in$ bzgl. C konvex (wobei C eine konvexe Menge), dann ist f in \bar{x} bzgl. C auch quasikonvex (vgl. Definition der Quasikonvexität in Abschnitt 1.3).

2.3.2 Erste Eigenschaften konvexer Funktionen

Satz (2.23). (Jensensche Ungleichung)

Sei $C \subset \mathbf{R}^n$, C -konvex, $f : C \rightarrow \mathbf{R}$.

Dann gilt: f ist auf C konvex genau dann, wenn aus

$$\begin{cases} x^1, x^2, \dots, x^k \in C \\ \lambda_i \geq 0, \forall i = 1, 2, \dots, k \\ \sum_{i=1}^k \lambda_i = 1 \end{cases}$$

stets folgt, daß $f(\sum_{i=1}^k \lambda_i x^i) \leq \sum_{i=1}^k \lambda_i f(x^i)$.

Bemerkung.

In der Definition der Konvexität ist $k = 2$. Satz 2.23 verallgemeinert die Ungleichung (o) aus der Definition der Konvexität auf konvexe Linearkombinationen beliebiger (endlicher) Länge.

Satz (2.24).

Sei $S \subset \mathbf{R}^n$. Für $i = 1, 2, \dots, k$ sei $f_i : S \rightarrow \mathbf{R}$ und $\alpha_i \in \mathbf{R}$, $\alpha_i \geq 0$.

Dann gilt: Wenn jedes der f_i , $i = 1, 2, \dots, k$ in $\bar{x} \in S$ konvex bzgl. S ist, dann ist

$$f(x) := \sum_{i=1}^k \alpha_i f_i(x)$$

in \bar{x} bzgl. S konvex.

Bemerkung.

1. Eine nichtnegative Linearkombination konvexer Funktionen ist also wieder konvex.

2. Wenn unter den Bedingungen des Satzes $\exists i_0$, so daß $\alpha_{i_0} > 0$, $f_{i_0}(x)$ in \bar{x} bzgl. S streng konvex, dann ist $f(x)$ streng konvex in \bar{x} bzgl. S .
3. Nach Satz 2.24 ist also

$$f(x) = - \langle a, x \rangle + \alpha + 3 \cdot e^x + 17 \cdot \sin x + 5 \cdot x^2$$

auf $[\pi, 2\pi]$ konvex.

Satz (2.25). (Obere Einhüllende)

Es sei $C \subset \mathbf{R}^n$, C - konvex, $A \subset \mathbf{R}^m$. Es sei $\phi : C \times A \rightarrow \mathbf{R}$, wobei für bel. $y \in A$, y - fixiert, die Funktion $\phi(\cdot, y)$ auf C eine konvexe Funktion (bzgl. x) sei. Dann gilt:

$$f(x) := \sup_{y \in A} \phi(x, y)$$

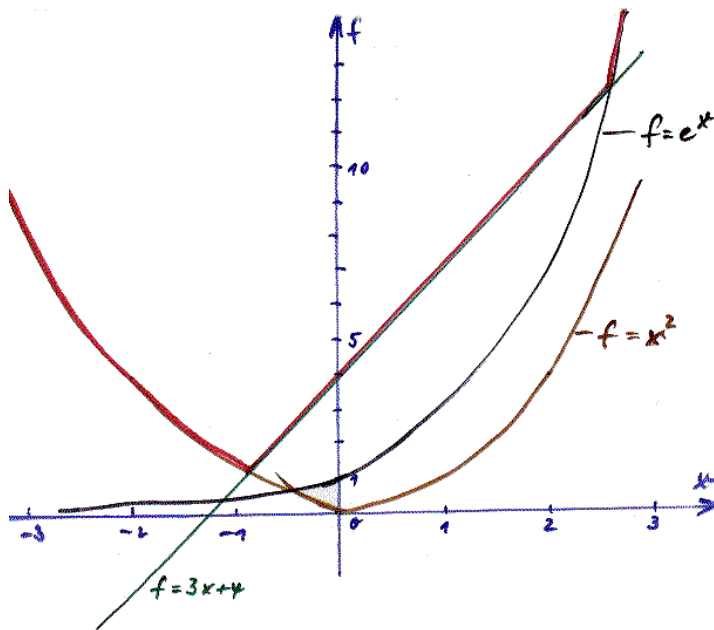
ist auf C konvex.

Bemerkung.

1. I.a. ist $f : C \rightarrow \mathbf{R} \cup \{+\infty\}$.
2. Es sei $A = \{1, 2, \dots, k\}$. Wir setzen $f_i(x) := \phi(x, i)$, $i = 1, 2, \dots, k$.
Der Satz 2.25 bedeutet dann: Wenn f_1, f_2, \dots, f_k auf C konvexe Funktionen sind, dann ist $f(x) = \max_{1 \leq i \leq k} f_i(x)$ auf C konvex.
3. Beispiel aus dem \mathbf{R}^1 : Seien $f_1(x) = x^2$, $f_2(x) = e^x$, $f_3(x) = 3x + 4$.
Nach dem Satz ist

$$f(x) = \max\{x^2, e^x, 3x + 4\} = \begin{cases} x^2 & , x \leq -1 \\ 3x + 4 & , -1 \leq x \leq a \\ e^x & , x \geq a \end{cases}$$

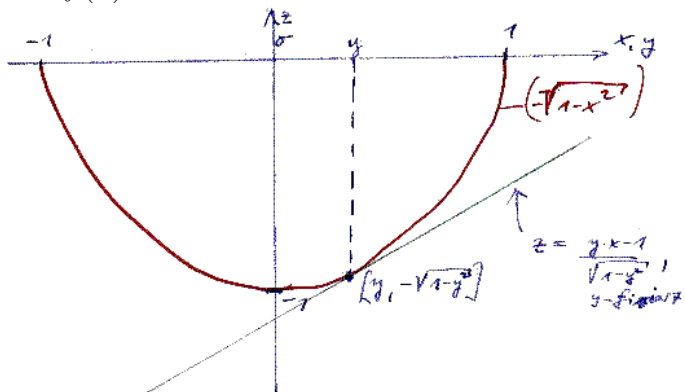
(wobei a Lösung der Gleichung $3x + 4 = e^x$, offenbar ist $2 < a < 3$)
eine konvexe Funktion.



4. Beispiel für $\phi(x, y)$ und f aus dem Satz.

Sei $\phi(x, y) = \frac{1}{\sqrt{1-y^2}}(y \cdot x - 1)$. Sei $A = (-1, 1)$, $C = [-1, 1]$

Bei $y \in A$, y -fixiert, ist $z = \phi(x, y)$ die Gleichung der Tangente im Punkt $[y, -\sqrt{1-y^2}]$ an den unteren Kreisbogen mit der Gleichung $z = f(x) := -\sqrt{1-x^2}$.



Bei fixiertem y ist $\phi(x, y)$ eine affin-lineare Funktion in x und damit bzgl. x eine konvexe Funktion. Nach dem Satz ist $\sup_{y \in (-1, 1)} \phi(x, y)$ eine konvexe Funktion. Die obere Einhüllende der Tangenten ist aber gerade der untere Kreisbogen mit der Gleichung $z = -\sqrt{1-x^2}$.

Also haben wir gezeigt, daß der untere Kreisbogen eine konvexe Funktion ist (was uns durch die Anschauung schon als plausibel erschien).

Satz (2.26). (Superposition konvexer Funktionen)

Es sei $C \subset \mathbf{R}^n$, C - konvex und für $i = 1, 2, \dots, m$ sei $g_i : C \rightarrow \mathbf{R}$ eine auf C konvexe Funktion.

Wir führen die Vektorfunktion $g : C \rightarrow \mathbf{R}^m$ nach

$$g(x) := [g_1(x), g_2(x), \dots, g_m(x)]$$

ein. Weiterhin sei $U \subset \mathbf{R}^m$, U - konvex und die skalare Funktion $\phi : U \rightarrow \mathbf{R}$ habe folgende zwei Eigenschaften:

- ϕ ist auf U konvex.
- ϕ ist auf U bzgl. der Halbordnung des \mathbf{R}^m monoton nichtfallend, d.h. aus $u^1, u^2 \in U$ mit $u^1 \leq u^2$ folgt stets $\phi(u^1) \leq \phi(u^2)$.

Schließlich sei $g(C) \subset U$.

Dann gilt: $f(x) := \phi(g(x))$ ist konvex auf C .

Bemerkung.

1. $g(x) = x^2$ ist im \mathbf{R}^1 konvex, $\phi(u) = e^u$ ist im \mathbf{R}^1 konvex und monoton nichtfallend.
Nach dem Satz ist deshalb $f(x) := e^{x^2}$ im \mathbf{R}^1 konvex.
2. Warum können wir im Satz nicht einfach $U := g(C)$ setzen ?
(Vergewissern Sie sich an einem Beispiel für $m = 2$, daß $g(C)$ i.a. keine konvexe Menge ist, obwohl $g_1(x)$ und $g_2(x)$ auf der konvexen Menge C konvexe Funktionen sind!)

Wie der folgende Satz zeigt, können wir in Satz 2.26 die Forderung der Monotonie an ϕ fallen lassen, wenn g eine affin-lineare Vektorfunktion ist.

Satz (2.27).

Es sei $U \subset \mathbf{R}^m$, U - konvex. Es sei $\phi : U \rightarrow \mathbf{R}$ auf U eine konvexe Funktion.

Schließlich seien $b \in \mathbf{R}^m$ und A eine Matrix vom Typ $[m, n]$.

Wir setzen $C := \{x \in \mathbf{R}^n : (Ax + b) \in U\}$.

Dann ist $f(x) := \phi(Ax + b)$ auf C konvex.

Bemerkung.

Vergewissern Sie sich, daß unter den Bedingungen des Satzes C eine konvexe Menge ist.

2.3.3 Differentielle Konvexitätskriterien

Satz (2.28).

1. Sei $S \subset \mathbf{R}^n$, $f : S \rightarrow \mathbf{R}$. Sei $\bar{x} \in \text{int}(S)$ und f sei in \bar{x} differenzierbar.
Dann gilt:
Wenn f in \bar{x} bzgl. S konvex ist, dann gilt:

$$f(x) \geq f(\bar{x}) + \langle \nabla f(\bar{x}), x - \bar{x} \rangle, \quad \forall x \in S$$

2. Sei $S \subset \mathbf{R}^n$, $f : S \rightarrow \mathbf{R}$, S eine offene Menge, f auf S differenzierbar.
Sei $C \subset S$, C - konvex. Dann gilt:
 f ist auf C konvex genau dann, wenn $f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle$
, $\forall x, y \in C$.

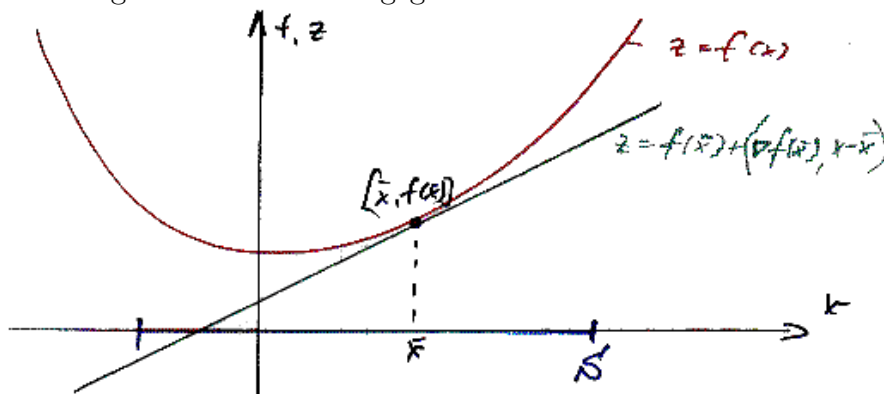
Folgerung (2.28').

Wenn f in $\bar{x} \in \text{int}(S)$ differenzierbar und konvex bzgl. S ist, dann ist f in \bar{x} bzgl. S pseudokonvex (vgl. Definition in Abschnitt 1.2)

Geometrische Illustration: im \mathbf{R}^1

Sei $S \subset \mathbf{R}$, $f : S \rightarrow \mathbf{R}$.

Dann ist durch $z = f(x)$ eine Kurve des \mathbf{R}^2 gegeben. Durch $z = f(\bar{x}) + \langle \nabla f(\bar{x}), x - \bar{x} \rangle$ ist eine Gerade durch den Kurvenpunkt $[\bar{x}, f(\bar{x})]$, genauer, die Tangente an die Kurve gegeben.



Die Forderung aus Satz 2.28 bedeutet: Die Kurve $z = f(x)$ gerät nirgendwo in S unterhalb der Tangente an f in $x = \bar{x}$.

Allgemeiner: ($n > 1$)

Die Hyperfläche mit der Gleichung $z = f(x)$ gerät nirgendwo in S unterhalb der Tangentialhyperebene mit der Gleichung $z = f(\bar{x}) + \langle \nabla f(\bar{x}), x - \bar{x} \rangle$.

Bemerkung.

Konvexe, differenzierbare Funktionen zeichnen sich also dadurch aus, daß bei

ihnen das Restglied in der Taylorreihe

$$f(x) = f(\bar{x}) + \langle \nabla f(\bar{x}), x - \bar{x} \rangle + o(\|x - \bar{x}\|)$$

nichtnegativ ist.

Satz (2.29).

Sei $S \subset \mathbf{R}^n$, $f : S \rightarrow \mathbf{R}$, S sei eine offene Menge und f sei auf S differenzierbar. Sei $C \subset S$, C -konvex. Dann gilt:

f ist auf C genau dann konvex, wenn $\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq 0$, $\forall x, y \in C$.

Bemerkung.

Sei A ein Operator, der aus dem \mathbf{R}^n in den \mathbf{R}^n wirkt $A : \mathbf{R}^n \rightarrow \mathbf{R}^n$. Wir sagen, daß A ein **monotoner Operator** ist wenn $\langle A(x) - A(y), x - y \rangle \geq 0$, $\forall x, y \in \mathbf{R}^n$.

Satz 2.29 sagt also, daß bei einer konvexen Funktion der Gradient ein monotoner Operator ist.

Betrachten Sie auch monotone Funktionen des \mathbf{R}^1 als Beispiel eines monotonen Operators.

Satz (2.30).

1. Sei $S \subset \mathbf{R}^n$, $f : S \rightarrow \mathbf{R}$, sei $\bar{x} \in \text{int}(S)$. Sei f in \bar{x} zweimal differenzierbar. Dann gilt:

Wenn f in \bar{x} bzgl. S konvex ist, dann ist $\nabla^2 f(\bar{x})$ eine positiv semidefinite Matrix.

2. Sei $S \subset \mathbf{R}^n$, $f : S \rightarrow \mathbf{R}$, S - offene Menge des \mathbf{R}^n . Sei $C \subset S$, C -konvex, sei f auf C zweimal differenzierbar. Dann gilt:

f ist auf C genau dann konvex, wenn in jedem Punkt x von C die Matrix $\nabla^2 f(x)$ positiv semidefinit ist.

Folgerung.

1. Sei Q eine quadratische symmetrische Matrix der Ordnung n . Nach dem Satz ist die quadratische Funktion $f(x) = \alpha + \langle c, x \rangle + \langle x, Qx \rangle$ im \mathbf{R}^n genau dann konvex, wenn Q positiv semidefinit ist.

2. Sei $C \subset \mathbf{R}^1$, C - konvex. Sei f eine auf einer Umgebung von C definierte Funktion: $f : U(C) \rightarrow \mathbf{R}$. Sei f auf C zweimal differenzierbar.

Dann ist f auf C konvex $\iff f''(x) \geq 0$, $\forall x \in C$.

2.3.4 Ungleichungssysteme mit konvexen Funktionen

Wir wollen hier eine weitere Anwendung der Trennungssätze erwähnen.

Satz (2.31). (Ky Fan)

Sei $C \subset \mathbf{R}^n$, C -konvex. Seien $f_i : C \rightarrow \mathbf{R}$, $i = 1, 2, \dots, m$. Die f_i , $i = 1, 2, \dots, m$, seien auf C konvex. Außerdem seien f_{m+1}, \dots, f_{m+k} affin-lineare Funktionen im \mathbf{R}^n . Dann gilt:

Wenn das System

$$(+)\begin{cases} f_i(x) < 0, & i = 1, 2, \dots, m \\ f_{m+j}(x) = 0, & j = 1, 2, \dots, k \\ x \in C \end{cases}$$

nicht lösbar ist, dann ist das System (Unbekannte sind die y_i)

$$(++)\begin{cases} \sum_{i=1}^{m+k} y_i f_i(x) \geq 0, & \forall x \in C \\ y_i \geq 0, & i = 1, 2, \dots, m \\ \sum_{i=1}^{m+k} y_i^2 > 0 \end{cases}$$

lösbar.

Satz (2.32). (Farkas - Minkowski)

Sei $C \subset \mathbf{R}^n$, C -konvex, seien $f_i : C \rightarrow \mathbf{R}$, $i = 0, 1, 2, \dots, m$ und diese f_i seien auf C konvex. Schließlich seien f_{m+j} , $j = 1, 2, \dots, k$ affin-lineare Funktionen im \mathbf{R}^n . Dann gilt:

Wenn das System

$$(\circ)\begin{cases} f_0(x) < 0 \\ f_i(x) < 0, & i = 1, 2, \dots, m \\ f_{m+j}(x) \leq 0, & j = 1, 2, \dots, l \\ f_{m+j}(x) = 0, & j = l+1, l+2, \dots, k \\ x \in C \end{cases}$$

nicht lösbar ist, aber das System

$$(\Delta)\begin{cases} f_i(x) < 0, & i = 1, 2, \dots, m \\ f_{m+j}(x) \leq 0, & j = 1, 2, \dots, l \\ f_{m+j}(x) = 0, & j = l+1, \dots, k \\ x \in ri(C) \end{cases}$$

ist lösbar, dann ist auch das System (Unbekannte sind die y_i)

$$(\circ\circ) \begin{cases} f_o(x) + \sum_{i=1}^{m+k} y_i f_i(x) \geq 0, & \forall x \in C \\ y_i \geq 0, & i = 1, 2, \dots, m+k \end{cases}$$

lösbar.

2.4 Konvexe Optimierungsaufgaben

Definition.

Sei $S \subset \mathbf{R}^n$, $f : S \rightarrow \mathbf{R}$. Wir sagen, daß die Optimierungsaufgabe

$$(OA) \begin{cases} f(x) \rightarrow \min \\ x \in S \end{cases}$$

eine **konvexe Optimierungsaufgabe** ist, wenn

$$\begin{cases} S - \text{konvexe Menge} \\ f - \text{eine auf } S \text{ konvexe Funktion} \end{cases}$$

sind.

Bemerkung.

Uns geht es um Aussagen, die generell für konvexe Aufgaben gelten, wir wollen aber auch abgeschwächte Formen der Konvexität betrachten (vgl. hierzu auch die Sätze 1.4.1 und 1.8.2 zur Aufgabe (FO) und (OAGR)).

Satz (2.33).

Wenn die OA

$$(OA) \begin{cases} f(x) \rightarrow \min \\ x \in S \end{cases}$$

eine konvexe OA ist, dann ist jede lokal-optimale Lösung auch global-optimal.

Bemerkung.

1. Da eine global-optimale Lösung immer lokal-optimal ist, so braucht man also bei konvexen Aufgaben nicht zwischen lokal und global zu unterscheiden.
2. Wenn man in Satz 2.33 die Forderung " f konvex auf S " ersetzt durch " f ist im Punkt \bar{x} des lokalen Minimums streng quasikonvex bzgl. S ", so gilt: \bar{x} ist Punkt des globalen Minimums von f über S .

3. Wenn f auf S pseudokonvex ist, S - konvex, dann ist jedes lokale Minimum in (OA) auch global.

Satz (2.34).

Die OA $f^* = \min\{f(x) : x \in S\}$ sei eine konvexe OA. Dann gilt:

1. Die Menge

$$\text{Argmin}\{f(x) : x \in S\} := \{x \in S : f(x) = f^*\}$$

aller optimalen Lösungen ist eine konvexe Menge.

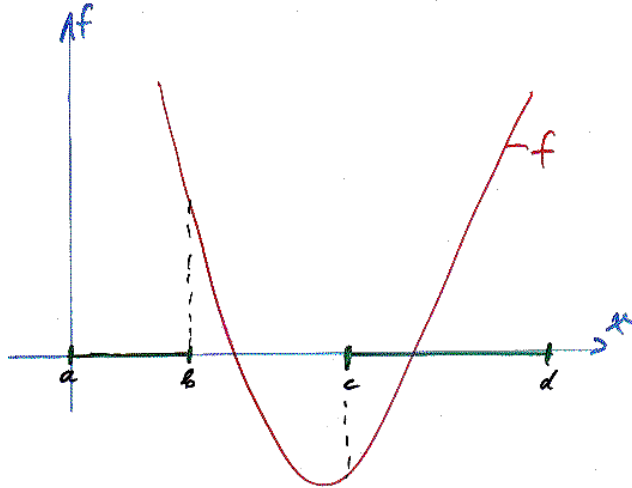
2. Wenn zusätzlich f sogar streng konvex ist, dann besteht $\text{Argmin}\{f(x) : x \in S\}$ aus höchstens einem Punkt.

Folgerung (2.34').

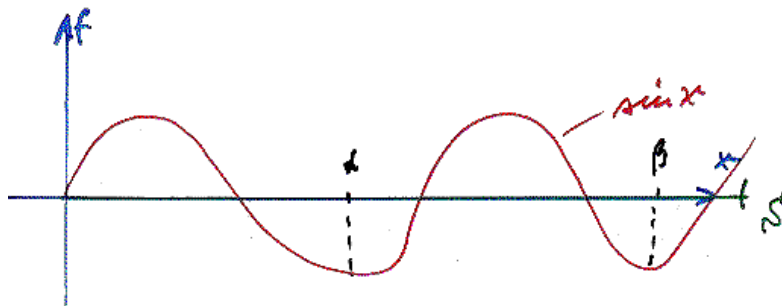
Bei konvexen Aufgaben ist die Optimalmenge eine zusammenhängende Menge.

Bemerkung.

Wenn f auf S nicht konvex und/oder S keine konvexe Menge ist, dann muß weder (1.) noch (2.) des Satzes gelten:



$S = [a, b] \cup [c, d]$, S - nicht konvex, f - streng konvex. Die Menge $\{c, b\}$ der lokal-optimalen Lösungen ist weder konvex noch höchstens einelementig, noch ist lokal gleich global.



f - nicht konvex, S - konvex. Die Menge der lokalen Minima $\{\alpha, \beta\}$ ist gleich der Menge der globalen Minima, aber weder (1.) noch (2.) gelten.

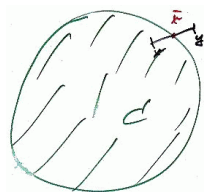
Wir wollen noch den Fall S - konvex, aber f auf S konkav betrachten. Dazu benötigen wir den Begriff des Extrempunktes einer konvexen Menge.

Definition.

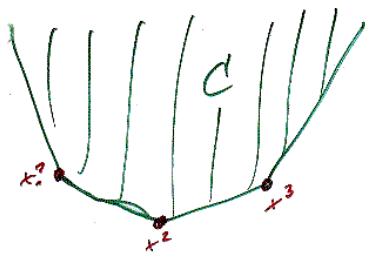
Sei $C \subset \mathbf{R}^n$ konvex. Der Punkt $\bar{x} \in C$ heißt **Extrempunkt** von C , wenn

$$\begin{cases} \bar{x} = \lambda x + (1 - \lambda)y \\ x, y \in C \\ \lambda \in [0, 1] \end{cases} \quad \text{nur bei} \quad \begin{cases} \lambda = 0 & (\text{d.h. } y = \bar{x}) & \text{oder} \\ \lambda = 1 & (\text{d.h. } x = \bar{x}) & \text{oder} \\ 0 < \lambda < 1, & \text{aber } \bar{x} = y = x \end{cases}$$

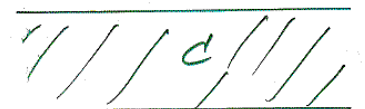
möglich ist (d.h. \bar{x} kann nicht als innerer Punkt eines echten, voll in C liegenden Abschnittes dargestellt werden).



Jeder Punkt der Peripherie ist ein Extrempunkt.



x^1, x^2, x^3 sind alle Extrempunkte von C .



C hat keine Extrempunkte.

Satz (2.35).

Gegeben sei die OA (OA) $f^* = \inf\{f(x) : x \in S\}$, wobei $f \not\equiv \text{const}$ auf S , f auf S quasikonkav, $S \subset \mathbf{R}^n$, S -konvex. Dann gilt:

1. Jede global-optimale Lösung von (OA) gehört zu $cl(S) \setminus ri(S)$ (d.h. ist Punkt des relativen Rands von S).
2. Wenn f sogar streng quasikonkav auf S ist und S einen Extrempunkt besitzt, dann kommen nur Extrempunkte von $cl(S)$ als global-optimale Lösungen in Frage.

Bemerkung.

1. Der Satz hat vielfältige Anwendungen, da er es erlaubt, die Suche nach dem globalen Optimum auf den relativen Rand bzw. die Extrempunkte von S zu beschränken.
2. Da eine konkave Funktion stets quasikonkav ist, gilt (1.) des Satzes auch für konkaves f . Bei konkavem f gilt (2.) in folgender Form:
Wenn S einen Extrempunkt besitzt und die Menge der global-optimalen Lösungen nichtleer ist, dann ist unter den global-optimalen Lösungen wenigstens ein Extrempunkt darunter.

Kapitel 3

Die allgemeine Aufgabe der Mathematischen Optimierung

3.1 Die Aufgabe (MOA)

Im ersten Abschnitt hatten wir bereits einige Klassen von Optimierungsaufgaben betrachtet: Aufgabe des Freien Optimums (FO), die Optimierungsaufgabe mit Gleichungseinschränkungen (OAGR), die Lineare Optimierungsaufgabe (LOA). Wir wollen hier die Aufgabe der Optimierung in endlichdimensionalen Räumen betrachten, welche die obigen Aufgabenstellungen als Spezialfälle enthält.

Definition.

Sei $S \subset \mathbf{R}^n$, $f : S \rightarrow \mathbf{R}$. Wir sagen, daß die Aufgabe

$$(OA) \begin{cases} f(x) \rightarrow \min \\ x \in S \end{cases}$$

eine **Aufgabe der Mathematischen Optimierung** ist, wenn gegeben sind

$$\begin{cases} \text{eine Menge } P \subset \mathbf{R}^n \\ \text{Funktionen } g_i : P \rightarrow \mathbf{R}, \quad i = 1, 2, \dots, m \end{cases}$$

so daß

$$S := \{x \in P : g_i(x) \leq 0, \quad i = 1, 2, \dots, k \\ g_i(x) = 0, \quad i = k + 1, k + 2, \dots, m\}$$

Bemerkung.

1. Wir bezeichnen die angesprochene Aufgabe kurz als MOA:

$$(MOA) \begin{cases} f(x) \rightarrow \min \\ g_i(x) \leq 0 & i = 1, 2, \dots, k \\ g_i(x) = 0 & i = k + 1, k + 2, \dots, m \\ x \in P \end{cases}$$

2. Typische Fälle für P sind $P = \mathbf{R}^n$, $P = \mathbf{R}_+^n$, $P = \{x \in \mathbf{R}^n : a \leq x \leq b\}$.

3. In (MOA) sollen die Fälle $k = m$ (d.h. keine Gleichungsrestriktionen), $k = 0$ (d.h. keine Ungleichungsrestriktionen), $m = 0$ (d.h. nur Menge P als Restriktionsbereich) nicht ausgeschlossen sein.

4. Unser Ziel ist es, notwendige und hinreichende Optimalitätsbedingungen für (MOA) zu finden. Insbesondere soll der sogenannte reguläre Fall diskutiert werden. Später wollen wir (MOA) eine sogenannte duale Aufgabe zuordnen. Die Aussagen in diesem Abschnitt verallgemeinern somit die Aussagen der Abschnitte zu (FO), (OAGR) und (LOA).

3.2 Optimalitätskriterien für (MOA)

Wir betrachten die Aufgabe (MOA) für differenzierbare Funktionen f und $g_i, i = 1, 2, \dots, m$. Es seien folgende Bezeichnungen vereinbart:

$$S = \{x \in P : g_i(x) \leq 0, i = 1, \dots, k \\ g_i(x) = 0, i = k + 1, \dots, m\} - \text{zulässiger Bereich,}$$

$$Y = \{y \in \mathbf{R}^m : y_i \geq 0, \quad i = 1, 2, \dots, k\}$$

$$L(x, y_o, y) = y_o f(x) + \sum_{i=1}^m y_i g_i(x) - \text{zu (MOA) gehörende Lagrangefunktion}$$

$$\nabla_x L(x, y_o, y) = y_o \nabla f(x) + \sum_{i=1}^m y_i \nabla g_i(x) - x - \text{Gradient der Lagrangefunktion}$$

L betrachten wir als definiert für $y_o \geq 0, y \in Y, x \in P$.

Theorem (3.1). (W. Karush 1938, F: John 1948)

Es sei $X_o \subset \mathbf{R}^n, X_o$ offen. Wir betrachten die Aufgabe (MOA) mit $P \subset X_o, P$ -konvex, $f, g_i : X_o \rightarrow \mathbf{R}, \quad i = 1, 2, \dots, m$.

Es sei $x^* \in S$. Die Funktionen f, g_1, \dots, g_k seien in x^* differenzierbar, die Funktionen g_{k+1}, \dots, g_m seien in einer Umgebung von x^* stetig differenzierbar.

Dann gilt:

Wenn x^* lokal-optimale Lösung von (MOA) ist, dann existieren $y_o^* \in \mathbf{R}, y^* \in \mathbf{R}^m$, so daß gilt

$$(STAT) \begin{cases} (y_o^*)^2 + \|y^*\|^2 > 0 \\ y_o^* \geq 0, y^* \in Y \\ y_i^* \cdot g_i(x^*) = 0 \quad i = 1, 2, \dots, k \\ \langle \nabla_x L(x^*, y_o^*, y^*), x - x^* \rangle \geq 0, \quad \forall x \in P \end{cases}$$

Definition.

Jedes $x^* \in S$, zu dem $y_o^* \in \mathbf{R}$ und $y^* \in \mathbf{R}^m$ existieren, so daß das Tripel $[x^*, y_o^*, y^*]$ die Bedingungen (STAT) erfüllt, heißt **stationärer Punkt** von (MOA).

Bemerkung.

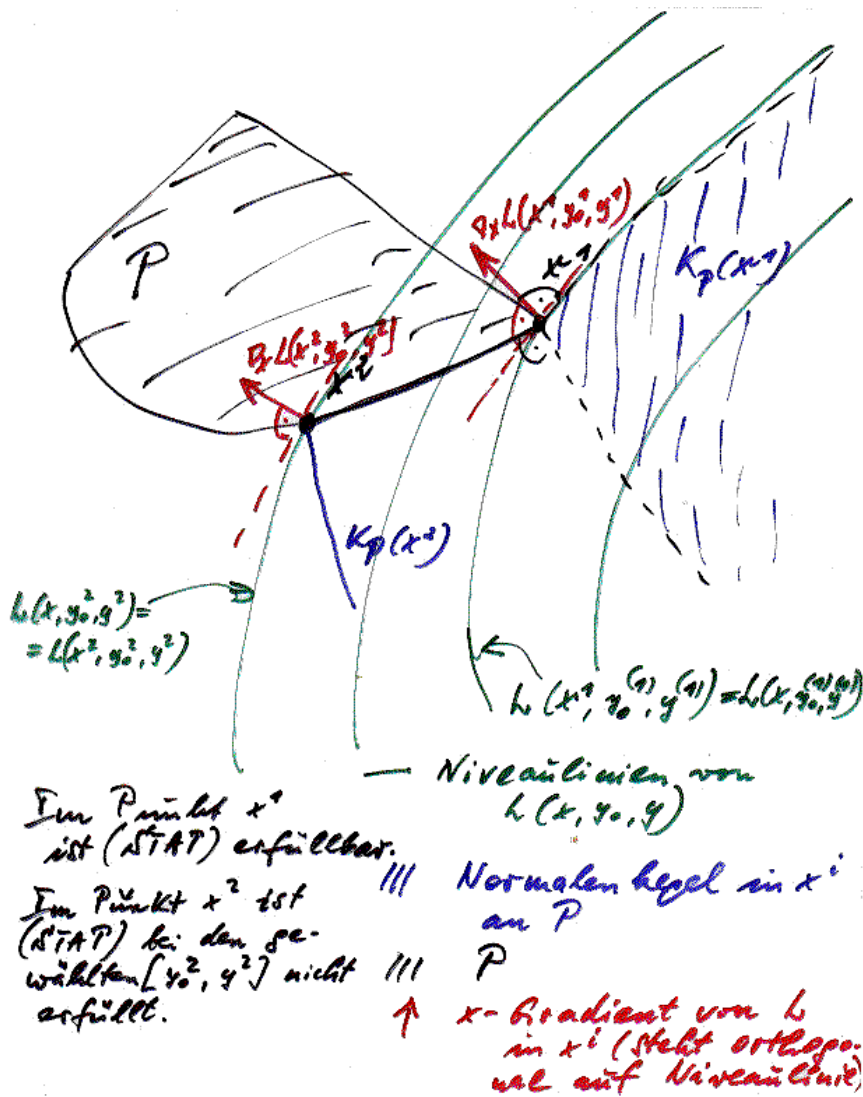
1. Die Bedingungen $y_i^* \cdot g_i(x^*) = 0, i \leq k$, heißen **Komplementaritätsbedingungen**.
2. Wenn wir setzen $I(x^*) := \{i \in \{1, 2, \dots, m\} : g_i(x^*) = 0\}$, dann folgt aus (STAT) und $x^* \in S$, daß $y_i^* = 0$ bei $i \notin I(x^*)$. Deshalb sind die Bedingungen (STAT) äquivalent (in beiden Richtungen) mit

$$\overline{(STAT)} \begin{cases} (y_o^*)^2 + \sum_{i \in I(x^*)} y_i^* > 0 \\ y_o^* \geq 0, \\ y_i^* = 0, \quad i \notin I(x^*) \\ y_i^* \geq 0, \quad (i \leq k) \wedge (i \in I(x^*)) \\ \langle y_o^* \nabla f(x^*) + \sum_{i \in I(x^*)} y_i^* \nabla g_i(x^*), x - x^* \rangle \geq 0, \quad \forall x \in P \end{cases}$$

3. Die in (STAT) vorkommenden Zahlen $y_o^*, y_1^*, \dots, y_m^*$ heißen **Lagrange-sche Multiplikatoren** für (MOA)

Definition.

Sei $P \subset \mathbf{R}^n$. Sei $x^* \in P$. Die Menge $K_P(x^*) := \{a \in \mathbf{R}^n : \langle a, x - x^* \rangle \leq 0, \forall x \in P\}$ heißt **Normalenkegel** in x^* an P .



Bemerkung.

1. Die Behauptung von Theorem 3.1 bedeutet also: Im lokal-optimalen Punkt x^* von (MOA) existieren Lagrangesche Multiplikatoren y_0^*, \dots, y_m^* , so daß $-\nabla_x L(x^*, y_0^*, y^*) \in K_P(x^*)$.
2. Die letzte Zeile von (STAT) enthält i.a. eine unendliche Anzahl von Forderungen ($\langle \cdot, \cdot \rangle \geq 0$ muß für jedes $x \in P$ gelten). Deshalb, wenn wir in der Lage sind, $K_P(x^*)$ durch ein endliches System von Bedingungen zu ersetzen, dann läßt sich auch (STAT) auf ein endliches System reduzieren.
Die nächsten Folgerungen liefern einige Beispiele dafür.

Folgerung (3.2).

Wenn unter den Bedingungen von Theorem 3.1 gilt $x^* \in \text{int}(P)$, dann kann die letzte Zeile von (STAT) ersetzt werden durch $\nabla_x L(x^*, y_o^*, y^*) = 0$.

Bemerkung.

Der angesprochene Fall liegt z.B. vor, wenn $P \equiv \mathbf{R}^n$.

Folgerung (3.3).

Wenn unter den Bedingungen von Theorem 3.1 gilt $P = \{x \in \mathbf{R}^n : a \leq x \leq b\}$, wobei $a, b \in \mathbf{R}^n$ vorgegebene Vektoren mit $a \leq b$, dann kann die letzte Zeile von (STAT) ersetzt werden durch

$$\frac{\partial L(x^*, y_o^*, y^*)}{\partial x_j} = \begin{cases} \geq 0 & , \text{ wenn } x_j^* = a_j \\ \leq 0 & , \text{ wenn } x_j^* = b_j \\ = 0 & , \text{ wenn } a_j < x_j^* < b_j \end{cases}$$

Bemerkung.

Bei $P = \{x \in \mathbf{R}^n : x \geq o\}$ ergibt sich

$$\begin{cases} \nabla_x L(x^*, y_o^*, y^*) \geq 0 \\ \frac{\partial L(x^*, y_o^*, y^*)}{\partial x_j} \cdot x_j^* = 0 \quad \forall j \end{cases}$$

Folgerung (3.4).

Wenn unter den Bedingungen von Theorem 3.1 gilt $P = \{x \in \mathbf{R}^n : Ax = b\}$, wobei $b \in \mathbf{R}^m$, A eine Matrix vom Typ $[m, n]$, dann kann die letzte Zeile von (STAT) ersetzt werden durch: $\exists \lambda^* = [\lambda_1^*, \dots, \lambda_m^*]$, so daß

$$\nabla_x L(x^*, y_o^*, y^*) = \lambda^* A = \sum_{i=1}^m \lambda_i^* a^i,$$

$a^i = i$ -te Zeile von A .

Theorem (3.5). (Hinlänglichkeits von (STAT))

Es sei $X_o \subset \mathbf{R}^n$, X_o offen, $P \subset X_o$. Gegeben sei Aufgabe (MOA). Es sei P konvex. Die Funktionen $f, g_1, \dots, g_m : X_o \rightarrow \mathbf{R}$ seien in $x^* \in S$ differenzierbar. Es sei x^* ein stationärer Punkt von (MOA) (d.h. es existieren y_o^*, y^* , so daß $[x^*, y_o^*, y^*]$ die Bedingungen (STAT) erfüllt).

Wir setzen

$$\begin{aligned} I_1 &= \{i \in \{1, 2, \dots, k\} : g_i(x^*) = 0\}, \\ I_2 &= \{i \in \{k+1, \dots, m\} : y_i^* > 0\}, \\ I_3 &= \{i \in \{k+1, \dots, m\} : y_i^* < 0\}, \end{aligned}$$

Wenn gilt

$$(\Delta) \begin{cases} y_o^* > 0 \\ f \text{ ist in } x^* \text{ bzgl. } P \text{ pseudokonvex} \\ g_i \text{ ist bei } i \in (I_1 \cup I_2) \text{ in } x^* \text{ bzgl. } P \text{ quasikonvex} \\ g_i \text{ ist bei } i \in I_3 \text{ in } x^* \text{ bzgl. } P \text{ quasikonkav} \end{cases}$$

dann ist x^* global-optimale Lösung von (MOA).

Bemerkung.

1. Vermerke, daß für $i \in \{1, 2, \dots, m\} \setminus (I_1 \cup I_2 \cup I_3)$ in (Δ) keine Forderung (außer Differenzierbarkeit) an g_i steht (wie der Beweis zeigt, könnte für diese i die Differenzierbarkeit durch Stetigkeit ersetzt werden).
2. Die Forderungen (Δ) sind z.B. erfüllt, wenn $y_o^* = 1$, f, g_1, \dots, g_k auf P konvex, g_{k+1}, \dots, g_m affin-linear (Menge S ist dann konvex).
3. Abschnitt 3.3 beschäftigt sich mit der Absicherung der Forderung $y_o^* > 0$ (sogenannter regulärer Fall von S).

Theorem 3.5 sichert uns einen globalen Optimumspunkt durch relativ scharfe Forderungen. Wir wollen noch ein Kriterium für ein lokales (sogenanntes strenges lokales Optimum) betrachten. Wir betrachten weiterhin Aufgabe (MOA)

$$\begin{cases} f(x) \rightarrow \min \\ g_i(x) \leq 0 & , i = 1, 2, \dots, k \\ g_i(x) = 0 & , i = k + 1, \dots, m \\ x \in P \end{cases}$$

Zusätzlich zu den oben vereinbarten Bezeichnungen seien noch folgende Notationen festgelegt:

$\nabla_{xx}^2 L(x, y_o, y) := y_o \nabla^2 f(x) + \sum_{i=1}^m y_i \nabla^2 g_i(x)$ – die x -Hessematrix der Lagrangefunktion von (MOA),

$V(x^*) := \{r \in \mathbf{R}^n : \text{es existieren } \lambda > 0 \text{ und } x \in P, \text{ so daß } r = (x - x^*) \cdot \lambda\}$ – für $x^* \in P$ ist das der Kegel der bzgl. P zulässigen Richtungen in x^* ,

$T(x^*) := \{r \in \mathbf{R}^n : \langle \nabla f(x^*), r \rangle \leq 0 \text{ und } \langle \nabla g_i(x^*), r \rangle \leq 0 \text{ für } i \leq k, \text{ bei denen } g_i(x^*) = 0 \text{ und } \langle \nabla g_i(x^*), r \rangle = 0, i = k + 1, \dots, m\}$ – Kegel der Richtungen in $x^* \in S$ die bzgl. der Nebenbedingungen zulässig und bzgl. der Zielfunktion keine Aufstiegsrichtungen sind.

Theorem (3.6). (Hinreichendes Kriterium 2. Ordnung)

Sei $X_o \subset \mathbf{R}^n$, X_o offen. Sei $P \subset X_o$, $f, g_i : X_o \rightarrow \mathbf{R}$, $i = 1, 2, \dots, m$.

Wir betrachten Aufgabe (MOA). Es sei $x^* \in S$. Es seien f, g_1, \dots, g_m in x^* zweimal differenzierbar. Außerdem sei x^* ein stationärer Punkt von (MOA), d.h. zusammen mit gewissen y_o^*, y^* erfüllt das Tripel $[x^*, y_o^*, y^*]$ die Bedingungen (STAT). Schließlich gelte

$$(+)\quad \langle r, \nabla_{xx}^2 L(x^*, y_o^*, y^*) r \rangle >> 0, \forall r \in [cl(V(x^*) \cap T(x^*) \setminus \{0\})].$$

Dann ist x^* lokal-optimale Lösung von (MOA).

Genauer: Es existiert ein $\epsilon > 0$, so daß

$$(\Delta)\quad f(x^*) < f(x) \quad , \forall x \in (U_\epsilon(x^*) \cap S) \setminus \{x^*\}.$$

Bemerkung.

Wir wollen einige Fälle ansprechen, in denen (+) direkt überprüfbar ist (wir setzen: $A := \nabla_{xx}^2 L(x^*, y_o^*, y^*)$):

1. $cl(V(x^*)) = \mathbf{R}^n$ und das mit $T(x^*)$ verbundene System von Gleichungen und Ungleichungen läßt sich explizit nach r auflösen, so daß man $\langle r, Ar \rangle >> 0$ auf $T(x^*) \setminus \{0\}$ direkt überprüfen kann.
2. Über die Eigenwerte von A oder das Sylvestersche Kriterium überprüft man, ob A positiv definit ist. In diesem Fall gilt (+) natürlich.
3. $cl(V(x^*)) \cap T(x^*) = \{0\}$. Dann gilt (+).
4. Es sei $cl(V(x^*)) = \mathbf{R}^n$ und man ist z.B. unter Einsatz von Software in der Lage nachzuweisen, daß die quadratische Optimierungsaufgabe

$$\begin{cases} \langle r, Ar \rangle \rightarrow \min \\ \langle \nabla f(x^*), r \rangle \leq 0 \\ \langle \nabla g_i(x^*), r \rangle \leq 0 & , i \leq k \text{ mit } g_i(x^*) = 0 \\ \langle \nabla g_i(x^*), r \rangle = 0 & , i = k + 1, \dots, m \\ -e \leq r \leq e & , \text{ wobei } e = [1, \dots, 1] \in \mathbf{R}^n \end{cases}$$

einen negativen Optimalwert hat. Dann kann (+) nicht gelten.

3.3 Regularität des zulässigen Bereichs

Wie oben erwähnt, ist die Forderung $y_o^* > 0$ in den Bedingungen (STAT) z.B. wichtig, um die Hinlänglichkeit von (STAT) zu beweisen (vgl. auch Abschnitt

4 über duale Aufgaben). Bedingungen an die Menge S , die das garantieren, werden als Regularitätsbedingungen bezeichnet (CQ - constraint qualification).

Wir wollen vorerst (MOA) über der Annahme $P = \mathbf{R}^n$ betrachten und dafür Regularitätsbedingungen finden. Es sei erwähnt, daß nach Folgerung 3.2 in diesem Fall die letzte Zeile von (STAT) äquivalent mit $\nabla_x L(x^*, y_o^*, y^*) = 0$ ist.

Abschließend zu den einleitenden Bemerkungen heben wir hervor, daß $y_o^* > 0$ in (STAT) äquivalent mit $y_o^* = 1$ ist, denn wir können dann in (STAT) setzen $y_i^* := \frac{y_i^*}{y_o^*}$, $i = 0, 1, \dots, m$.

Definition.

Gegeben sei Aufgabe (MOA) mit $P = \mathbf{R}^n$. Wir sagen, daß der zulässige Bereich S in $x^* \in S$ die **(MFCQ) - Bedingung** erfüllt, wenn gilt:

1. $\{\nabla g_i(x^*)\}_{i=k+1}^m$ sind linear unabhängig
2. $\exists \bar{x} \in \mathbf{R}^n$, so daß

$$\begin{aligned} \langle \nabla g_i(x^*), \bar{x} - x^* \rangle &< 0 && , \forall i \in I_1(x^*) \\ \langle \nabla g_i(x^*), \bar{x} - x^* \rangle &\geq 0 && , \forall i \in \{k+1, \dots, m\} \end{aligned}$$

Hierbei sei $I_1(x^*) = \{i \in \{1, 2, \dots, k\} : g_i(x^*) = 0\}$.

Satz (3.7).

Wir betrachten (MOA) unter der Annahme, daß $P = \mathbf{R}^n$. Wenn S in $x^* \in S$ die (MFCQ)-Bedingung erfüllt, dann ist auf jeder Lösung $[x^*, y_o^*, y^*]$ von (STAT) $y_o^* > 0$.

Definition.

Gegeben sei Aufgabe (MOA) unter der Voraussetzung $P = \mathbf{R}^n$. Wir sagen, daß in $x^* \in S$ die **(LICQ)-Bedingung** erfüllt ist, wenn $\{\nabla g_i(x^*)\}_{i \in I_1(x^*)} \cup \{\nabla g_i(x^*)\}_{i=k+1}^m$ linear unabhängig sind. (Li = linear independence).

Satz (3.8).

Wir betrachten Aufgabe (MOA) unter der Annahme, daß $P = \mathbf{R}^n$. Wenn im stationären Punkt $x^* \in S$ die (LICQ)-Bedingung erfüllt ist, dann ist auf jedem Tripel $[x^*, y_o^*, y^*]$, das (STAT) löst, $y_o^* > 0$.

Bemerkung.

1. Sowohl (MFCQ) als auch (LICQ) sind nicht sehr praktikabel:
 x^* ist unbekannt (d.h. gesucht). Um (LICQ) a-priori abzusichern, müßten wir garantieren, daß die $\{\nabla g_i(x)\}_{i \in I(x)} \cup \{\nabla g_i(x)\}_{i=k+1}^m$ in jedem $x \in S$ linear unabhängig sind. Außerdem ist in der Praxis die lineare Unabhängigkeit faktisch nicht überprüfbar.
2. Unter Benutzung des Satzes von Gordan (Satz 2.21) können sie zeigen, daß aus (LICQ) stets (MFCQ) folgt.

Als praktikabel erweisen sich Regularitätsbedingungen vom Typ (LiNCQ) und (SLATER-CQ). Wir wollen zuerst eine einfache Version von (SLATER-CQ) betrachten und dann praktikablere Varianten definieren.

Definition.

Gegeben sei Aufgabe(MOA) mit $P = \mathbf{R}^n$. Wir sagen, daß S die Bedingung **(SLATER-1-CQ)** erfüllt, wenn

1. Die $g_i, i \geq k + 1$, sind affin-linear (d.h. $g_i(x) = \langle \nabla g_i, x \rangle + \alpha_i, i = k + 1, \dots, m$).
2. Die Vektoren $\{\nabla g_i\}_{i=k+1}^m$ sind linear unabhängig.
3. Die Funktionen $g_i(x), i = 1, 2, \dots, k$ sind in \mathbf{R}^n konvex.
4. Es existiert ein $\bar{x} \in S$ mit $g_i(\bar{x}) < 0, i = 1, 2, \dots, k$.

Bemerkung.

Unter den Voraussetzungen der Definition ist S eine konvexe Menge.

Satz (3.9).

Wir betrachten Aufgabe (MOA) mit $P = \mathbf{R}^n$. Wenn (SLATER-1-CQ) gilt, dann ist auf jeder Lösung $[x^*, y_o^*, y^*]$ von (STAT) $y_o^* > 0$.

Definition.

Wir betrachten Aufgabe (MOA). Wir sagen, daß S die **(LiNCQ)** erfüllt, wenn gilt

1. g_1, \dots, g_m sind affin-linear
2. P ist eine konvexe polyedrale Menge

Definition.

Wir betrachten Aufgabe (MOA). Wir sagen, daß S die **(SLATER-2-CQ)** erfüllt, wenn gilt:

1. P ist konvex und polyedral.
2. Es existiert ein l mit $1 \leq l \leq k$, so daß die g_1, \dots, g_l auf P konvex sind und für wenigstens ein $\bar{x} \in S$ gilt: $g_i(\bar{x}) < 0, \quad i = 1, 2, \dots, l.$
3. $g_{l+1}, \dots, g_k, g_{k+1}, \dots, g_m$ sind affin-linear.

Definition.

Wir betrachten Aufgabe (MOA). Wir sagen, daß S die **(SLATER-3-CQ)** erfüllt, wenn gilt:

1. P ist konvex.
2. Es existiert ein l mit $1 \leq l \leq k$, so daß die g_1, \dots, g_l auf P konvex sind und für wenigstens ein $\bar{x} \in S \cap \text{ri}(P)$ gilt: $g_i(\bar{x}) < 0, \quad i = 1, 2, \dots, l.$
3. $g_{l+1}, \dots, g_k, g_{k+1}, \dots, g_m$ sind affin-linear

Satz (3.10).

Wir betrachten Aufgabe (MOA). Seien f, g_1, \dots, g_m in $x^* \in S$ differenzierbar. Sei x^* eine lokal-optimale Lösung von (MOA). Sei wenigstens eine der Bedingungen (LiNCQ), (SLATER-2-CQ) oder (SLATER-3-CQ) erfüllt. Dann existiert zu x^* wenigstens eine Lösung $[x^*, y_o^*, y^*]$ von (STAT), in der $y_o^* = 1.$

Bemerkung.

Wenn f in x^* bzgl. P noch zusätzlich pseudokonvex ist, dann ist unter den Bedingungen von Satz 3.10 x^* global-optimal (wende Theorem 3.5 an).

Theorem (3.11). (Kuhn-Tucker, 1950)

Gegeben sei Aufgabe (MOA) unter $P = \mathbf{R}^n$. Die Funktionen f, g_1, \dots, g_m seien in $x^* \in S$ differenzierbar. Außerdem seien f, g_1, \dots, g_k in \mathbf{R}^n konvex und g_{k+1}, \dots, g_m seien affin-linear. Sei wenigstens eine der Bedingungen (MFCQ), (LICQ), (LiNCQ), (SLATER-2-CQ) bzw. (SLATER-3-CQ) erfüllt.

Dann gilt: x^* ist global-optimale Lösung von (MOA) genau dann, wenn ein $y^* \in Y$ existiert, so daß

$$(K - T_{MOA}) \begin{cases} \nabla f(x^*) + \sum_{i=1}^m y_i^* \nabla g_i(x^*) = 0 \\ y_i^* \cdot g_i(x^*) = 0, \quad i = 1, 2, \dots, k \\ y_i^* \geq 0, \quad i = 1, 2, \dots, k \\ x^* \in S \end{cases}$$

Kapitel 4

Dualitätstheorie

Wir betrachten in diesem Abschnitt weiterhin die Aufgabe

$$(MOA) \begin{cases} f(x) \rightarrow \min \\ g_i(x) \leq 0 & , i = 1, 2, \dots, k \\ g_i(x) = 0 & , i = k + 1, \dots, m \\ x \in P \end{cases}$$

wobei $P \subset \mathbf{R}^n$. Wir gehen davon aus, daß $f, g_i : P \rightarrow \mathbf{R}$, $i = 1, 2, \dots, m$. Im wesentlichen werden wir den konvexen Fall betrachten, wobei im Gegensatz zum vorherigen Abschnitt die Differenzierbarkeit keine Bedingung ist. Wie oben sei auch hier für den gesamten Abschnitt vereinbart:

$$S = \{x \in P : g_i(x) \leq 0, \quad i = 1, \dots, k \\ g_i(x) = 0, i = k + 1, \dots, m\} - \text{zulässiger Bereich von (MOA).}$$

$Y = \{y \in \mathbf{R}^m : y_i \geq 0, \quad i = 1, 2, \dots, k\}$,
 $L(x, y) = f(x) + \sum_{i=1}^m y_i g_i(x)$ - die mit (MOA) verbundene reguläre Lagrangefunktion,
 L wird auf $P \times Y$ betrachtet.
 $f^* = \inf\{f(x) : x \in S\}$ - Optimalwert von (MOA).

4.1 Die duale Aufgabe

Jeder OA kann man eine duale Aufgabe zuordnen. Das Studium des dualen Objekts erlaubt tieferes Verständnis der Natur des ursprünglichen Objekts, das wir in diesem Zusammenhang primales Objekt nennen. Die Theorie der Beziehungen zwischen der primalen und der dualen Aufgabe

(die sogenannte Dualitätstheorie) ist eng mit der Theorie der Optimalitätskriterien verbunden. Wir wählen einen Zugang zum dualen Objekt über die reguläre Lagrangefunktion. Man kann auch andere Wege gehen und dabei andere duale Ausgaben erhalten: Fenchel-duale, über die Empfindlichkeitsfunktion usw. Um zur dualen Aufgabe zu gelangen, wollen wir die primale Aufgabe (MOA) erst in einer anderen Form aufschreiben.

Eingeführt haben wir sie als $f^* = \inf\{f(x) : x \in S\}$.

Da aber bei $x \in S$ und $y \in Y$ gilt

$$\begin{aligned} y_i \cdot g_i(x) &= 0, & i &= k+1, \dots, m \\ y_i \cdot g_i(x) &\leq 0, & i &= 1, 2, \dots, k \end{aligned}$$

und da $y = 0 \in Y$, so ist $\sup_{y \in Y} \sum_{i=1}^m y_i g_i(x) = 0$ (bei $x \in S$).

Andererseits, wenn $x \in (P \setminus S)$, dann ist wenigstens eine der folgenden zwei Beziehungen wahr:

- $\exists i_o \in \{k+1, \dots, m\}$ mit $g_{i_o}(x) \neq 0$,
- $\exists i_1 \in \{1, 2, \dots, m\}$ mit $g_{i_1}(x) > 0$

In beiden Fällen ist $\sup_{y \in Y} \sum y_i g_i(x) = +\infty$ (bei $x \in (P \setminus S)$). Deshalb haben wir

$$\begin{aligned} \Psi(x) &=_{Def} \sup\{L(x, y) : y \in Y\} \\ &= \sup\{f(x) + \sum_{i=1}^m y_i g_i(x) : y \in Y\} \\ &= f(x) + \sup\{\sum_{i=1}^m y_i g_i(x) : y \in Y\} \\ &= \begin{cases} f(x), & x \in S \\ +\infty, & x \in (P \setminus S). \end{cases} \end{aligned}$$

Deshalb ist (MOA) die Aufgabe (wenn $S \neq \emptyset$) $f^* = \inf\{\Psi(x) : x \in P\}$, d.h. die Aufgabe $f^* = \inf_{x \in P} \sup_{y \in Y} L(x, y)$.

Definition.

Es sei $\phi(y) = \inf_{x \in P} L(x, y)$ und $Q = \{y \in Y : \phi(y) > -\infty\}$. Die Aufgabe

$$\phi^* = \sup\{\phi(y) : y \in Q\} \quad (MOA_D)$$

heißt **duale Aufgabe** zu (MOA), genauer: (MOA_D) ist die **Lagrange-Duale** zu (MOA).

Bemerkung.

1. Bei $S = \emptyset$ setzen wir: $f^* = +\infty$. Bei $Q = \emptyset$ setzen wir: $\phi^* = -\infty$.
2. (MOA_D) ist damit die Aufgabe $\phi^* = \sup_{y \in Y} \inf_{x \in P} L(x, y)$. In dieser Form erkennen wir Symmetrieeigenschaften zur obigen Form $f^* = \inf_{x \in P} \sup_{y \in Y} L(x, y)$ von (MOA).
3. Aber (MOA) können wir in expliziter Form aufschreiben

$$(MOA) \begin{cases} f(x) \rightarrow \min \\ x \in S \end{cases}$$

(weil wir davon ausgegangen waren, daß uns f und S explizit bekannt sind), wogegen uns das bei (MOA_D) nur in wenigen Spezialfällen gelingen wird, da wir ϕ und Q nur algorithmisch kennen und i.a. nicht explizit durch eine endliche Formel beschreiben können (Ausnahme, z.B. LOA).

4. Noch ein paar Worte zu den Unterschieden zwischen der expliziten Form von (MOA) und der algorithmisch gegebenen (und damit impliziten) Form von (MOA_D) .

Es seien f, g_i, P explizit gegeben. Um in (MOA) für ein x zu ermitteln, ob x zulässig ist und den Zielfunktionswert zu berechnen, brauchen wir nur die Funktionswerte der uns explizit gegebenen Funktionen $g_i(x), i = 1, \dots, m$ sowie $f(x)$ berechnen, die in die Nebenbedingungen eingehenden Gleichungen und Ungleichungen überprüfen und prüfen, ob x zur explizit gegebenen Menge P gehört.

Um in (MOA_D) für ein y zu ermitteln, ob es zulässig ist und den Zielfunktionswert zu berechnen, müssen wir i.a. unter Einsatz eines Lösungsverfahrens (Algorithmus) zuerst die Optimierungsaufgabe $\phi(y) = \inf_{x \in P} L(x, y)$ lösen.

5. Rein formal bestehen zwischen (MOA) und (MOA_D) natürlich keine Unterschiede: f und ϕ sind Funktionen, S und Q sind Mengen, beides sind Optimierungsaufgaben.
6. Wir vermerken, daß i.a. $\phi : Y \rightarrow \mathbf{R} \cup \{-\infty\}$.

7. Abschließend sei vermerkt, daß sich die duale Aufgabe ändern kann, wenn man die (explizite) analytische Beschreibung der primalen Aufgabe ändert (ohne sie inhaltlich zu ändern), z.B. wenn man eine Redefinition von P vornimmt (z.B. $g_1(x) \leq 0$ zur Definition von P hinzufügt) oder Folgerungen aus dem System

$$\begin{cases} g_i(x) \leq 0 & , i = 1, \dots, k \\ g_i(x) = 0 & , i = k + 1, \dots, m \end{cases}$$

zum System hinzufügt.

Satz (4.1).

Unabhängig von der Natur der Menge P sowie der Funktionen $f, g_i, \quad i = 1, \dots, m$ gilt stets, daß (MOA_D) eine konvexe Aufgabe ist.

Satz (4.2). (Schwache Dualität)

1. Bei $x \in S$ und $y \in Y$ gilt stets $f(x) \geq \phi(y)$.
2. Wenn $S \neq \emptyset$ und $Q \neq \emptyset$, dann ist $-\infty < \phi^* \leq f^* < +\infty$.

Bemerkung.

1. Da $Q \subset Y$, so gilt $f(x) \geq \phi(y)$ um so mehr für zulässige Lösungen von (MOA) und (MOA_D) , d.h. für $x \in S, y \in Q$.
2. Wenn wenigstens eine der Mengen S und Q leer ist, dann gilt $f^* \geq \phi^*$ auch, denn $S = \emptyset \implies f^* = +\infty \implies f^* \geq \phi^*$ gilt, bei $Q = \emptyset \implies \phi^* = -\infty \implies f^* \geq \phi^*$ gilt.
3. Nach dem Beweis gilt

$$f(x) \geq \phi^*, \quad \forall x \in S \quad f^* \geq \phi(y), \quad \forall y \in Q.$$

Damit liefert der Zielfunktionswert einer zulässigen Lösung von (MOA) eine obere Schranke für den optimalen Zielfunktionswert von (MOA_D) . Analoges folgt aus der zweiten Ungleichung für f^* .

4. Nach dem Beweis haben wir außerdem $f(x) \geq f^* \geq \phi^* \geq \phi(y), \forall x \in S, \forall y \in Q$.

Damit läßt sich z.B. die Güte einer zulässigen Lösung x von (MOA) abschätzen (Güte gemessen an der Entfernung des Zielfunktionswertes vom optimalen Zielfunktionswert), denn wenn uns für ein $\bar{y} \in Q$ der Wert $\phi(\bar{y})$ bekannt ist und wir für $x \in S$ den Wert $f(x)$ berechnet haben, dann gilt $0 \leq f(x) - f^* \leq f(x) - \phi(\bar{y})$, d.h. der uns dann bekannte Wert $f(x) - \phi(\bar{y})$ liefert eine obere Schranke für die Optimumsentfernung von x .

5. Wenn für das Paar $x^* \in S, y^* \in Q$ zulässiger Lösungen von (MOA) und (MOA_D) gilt: $f(x^*) = \phi(y^*)$, dann sind (MOA) und (MOA_D) lösbar, ja es sind sogar x^* und y^* optimale Lösungen von (MOA) bzw. (MOA_D) , denn nach der 4. Bemerkung ist

$$\begin{aligned} f(x^*) &\geq f^* \geq \phi^* \geq \phi(y^*) = f(x^*) \\ \implies 0 &\geq f^* - f(x^*) \geq \phi^* - \phi(y^*) \geq 0 \\ \implies f(x^*) &= f^*, \phi(y^*) = \phi^* \\ \implies &\begin{cases} x^* \text{ ist optimale Lösung von } (MOA), \\ y^* \text{ ist optimale Lösung von } (MOA_D) \end{cases} \end{aligned}$$

Wir vermerken, daß in diesem Fall $f^* = \phi^*$ gilt.

6. Nach dem Satz ist $f^* \geq \phi^*$ in jedem Fall. Wenn $f^* > \phi^*$ gilt, spricht man von einer **Dualitätslücke**. Wenn es gelingt, für ein Aufgabenpaar (MOA) und (MOA_D) zu beweisen, daß $f^* = \phi^*$, dann sagt man, daß **starke Dualität** vorliegt.

Jeder denkbare Fall einer Dualitätslücke

$$\begin{aligned} +\infty = f^* > \phi^* > -\infty, & \quad +\infty > f^* > \phi^* = -\infty, \\ +\infty > f^* > \phi^* > -\infty, & \quad +\infty = f^* > \phi^* = -\infty \end{aligned}$$

ist möglich (d.h. läßt sich durch Beispiele belegen). Wir betrachten in 8.) ein Beispiel, auf dem $+\infty > f^* > \phi^* > -\infty$ gilt.

Zuvor sei vermerkt, daß auch $+\infty = f^* = \phi^*$ bzw. $-\infty = f^* = \phi^*$ möglich sind (d.h. durch Beispiele belegbar). Die 7. Bemerkung zeigt, welche spezielle Situation dann vorliegen muß.

7. Wenn $f^* = -\infty$, dann ist $Q = \emptyset$. Wenn $\phi^* = +\infty$, dann ist $S = \emptyset$.

8. Es sei $f(x) = x - 1$, $g(x) = x^2 - 1$, $P = \mathbf{R}_+^1$ und wir betrachten

$$\begin{cases} f(x) \rightarrow \min \\ g(x) = 0 \\ x \in P. \end{cases}$$

Diese Aufgabe hat $x = +1$ als einzige zulässige Lösung. Damit ist $f^* = 0$. Wir haben

$$\begin{aligned} \phi(y) &= \inf\{L(x, y) : x \in P\} = \inf\{(x - 1) + y(x^2 - 1) : x \geq 0\} \\ &= \begin{cases} -\infty, & y < 0 \\ -1 - y, & y \geq 0 \end{cases} \end{aligned}$$

Damit ist $Q = \mathbf{R}_+^1$ und $\phi^* = \sup\{\phi(y) : y \in Q\} = -1$, d.h. $0 = f^* > \phi^* = -1$. Begründung für die Werte von ϕ :

- Bei $y = 0$ ist $x - 1 + y(x^2 - 1) = x - 1$ und offensichtlich ist $\inf\{x - 1 : x \geq 0\} = -1$.
- Bei $y > 0$ ist $x - 1 + y(x^2 - 1)$ als Funktion von x eine nach oben geöffnete Parabel. Sie hat ihren kleinsten Wert im Scheitelpunkt der Parabel. Der Scheitelpunkt muß $0 = (x - 1 + y(x^2 - 1))' = 1 + 2yx$ erfüllen und liegt damit für $y > 0$ bei negativen x -Werten $\implies \inf\{(x - 1) + y(x^2 - 1) : x \geq 0\} = 0 - 1 + y(0 - 1) = -1 - y$.
- Bei $y < 0$ ist $x - 1 + y(x^2 - 1)$ als Funktion von x eine nach unten geöffnete Parabel. $\implies \inf\{(x - 1) + y(x^2 - 1) : x \geq 0\} = -\infty$.

9. Nach den eingangs des Abschnittes geführten Motivationsbetrachtungen zum Einführen der dualen Aufgabe bedeutet starke Dualität (d.h. $f^* = \phi^*$), daß

$$\inf_{x \in P} \sup_{y \in Y} L(x, y) = \sup_{y \in Y} \inf_{x \in P} L(x, y).$$

4.2 Starke Dualität

Definition.

Es sei $P \neq \emptyset$. Wie oben bezeichne f^* den Optimalwert von (MOA), genauer: $f^* = \inf_{x \in P} \sup_{y \in Y} L(x, y)$.

Der Vektor $y^* \in Y$ heißt **Kuhn-Tucker-Vektor** der Aufgabe (MOA), wenn

$$f^* \leq f(x) + \sum_{i=1}^m y_i^* g_i(x), \quad \forall x \in P \quad (\circ)$$

gilt.

Bemerkung.

Bei $S = \emptyset$ haben wir: $f^* = +\infty$. Wenn $P \neq \emptyset$, dann ist folglich (\circ) nur möglich, wenn $S \neq \emptyset$. Mit anderen Worten:

Wenn (MOA) einen Kuhn-Tucker-Vektor besitzt und $P \neq \emptyset$, dann ist $f^* < +\infty$ und $S \neq \emptyset$.

Satz (4.3). (Starke Dualität)

Aufgabe (MOA) besitze wenigstens einen Kuhn-Tucker-Vektor. Dann gilt:

1. Wenn $f^* > -\infty$, dann ist die Menge $Y^* =_{Def} \{y \in Q : \phi(y) = \phi^*\}$ der optimalen Lösungen von (MOA_D) nicht leer und es gilt:
 - (i) $+\infty > f^* = \phi^* > -\infty$,
 - (ii) $Y^* =$ Menge der Kuhn-Tucker-Vektoren von (MOA) .
2. Wenn $Q \neq \emptyset$, dann ist $f^* > -\infty$.
Wenn $Q = \emptyset$, dann ist $f^* = -\infty$.

Bemerkung.

1. Unter der Bedingungen des Satzes wird die starke Dualität, die Endlichkeit des gemeinsamen Optimalitätswertes und die Annahme des Optimalwertes in (MOA_D) gewährleistet, nicht aber die Annahme des Optimalwertes in (MOA) (Optimalwert f^* von (MOA) ist endlich, aber die Optimalmenge kann leer sein).
2. Nach Bemerkung 7.) zu Satz 4.2. folgt aus $f^* = -\infty$, daß $Q = \emptyset$. Satz 4.3. zeigt, daß bei Existenz eines Kuhn-Tucker-Vektors aus $Q = \emptyset$ auch $f^* = -\infty$ folgt. Generell können wir aus $Q = \emptyset$ aber nicht schlußfolgern, daß $f^* = -\infty$. Wie in Bemerkung 6.) zu Satz 4.2. bemerkt wurde, kann bei $Q = \emptyset$ (d.h. $\phi^* = -\infty$) auch $f^* = +\infty$ (d.h. $S = \emptyset$) oder $-\infty < f^* < +\infty$ sein. Das läßt sich durch Beispiele belegen (siehe Bemerkung unter Folgerung 4.8).
3. Die Existenz eines Kuhn-Tucker-Vektors von (MOA) läßt sich auf die Existenz eines Sattelpunktes der mit (MOA) verbundenenen regulären Lagrangefunktion zurückführen: $[x^*, y^*] \in P \times Y$ heißt **Sattelpunkt** von $L(x, y)$, wenn

$$(SP) \begin{cases} L(x^*, y) \leq L(x^*, y^*), & \forall y \in Y \\ L(x^*, y^*) \leq L(x, y^*), & \forall x \in P \end{cases}$$

Man kann leicht zeigen: Wenn $[x^*, y^*]$ ein Sattelpunkt von L ist, dann gilt

$$\begin{cases} x^* \text{ ist optimale Lösung von } (MOA) \\ y^* \text{ ist optimale Lösung von } (MOA_D) \\ \sum_{i=1}^m y_i^* g_i(x^*) = 0 \\ L(x^*, y^*) = f(x^*) = f^* \\ y^* \text{ ist Kuhn-Tucker-Vektor von } (MOA) \end{cases}$$

Wir wollen mit Satz 4.4. einen anderen Weg der Absicherung der Existenz eines Kuhn-Tucker-Vektors gehen, weil

- der Weg über den Sattelpunkt in der Praxis nur etwas bringt, wenn wir praktikable Kriterien haben, die die Existenz eines Sattelpunktes garantieren. Das aber führt wieder zu Satz 4.4.
- der Weg über den Sattelpunkt leicht über die Aussage von Satz 4.3. hinausschießt: Er sichert auch die Annahme des Optimums in (MOA).

Starke Dualität ist (wenn wir keine Differenzierbarkeitsforderungen erheben) nach Satz 4.3 mit der Existenz eines Kuhn-Tucker-Vektors verbunden. Letzteres hängt eng mit der Regularität des zulässigen Bereichs zusammen (vgl. $y_o^* = 1$ beim Kuhn-Tucker-Vektor). Wir werden das nur für die Regularitätsbedingung (SLATER-1-CQ) näher betrachten.

Satz (4.4).

Wenn in Aufgabe (MOA) f eine konvexe Funktion ist und (SLATER-1-CQ) erfüllt ist, dann besitzt (MOA) einen Kuhn-Tucker-Vektor.

4.3 Die Duale zu einer LOA

Wir wollen nur die (LOA) in (K)-Form betrachten:

$$(K) \begin{cases} \langle c, x \rangle \rightarrow \min \\ Ax = b \\ x \geq 0 \end{cases}$$

wobei $c \in \mathbf{R}^n, b \in \mathbf{R}^m, A$ - Matrix vom Typ $[m, n]$. (K) ist eine Aufgabe vom Typ (MOA).

Wir setzen $f(x) = \langle c, x \rangle, P = \mathbf{R}_+^n, k = 0, g(x) = Ax - b, Y = \mathbf{R}^m$. Dann ist

$$\begin{cases} L(x, y) = \langle c, x \rangle + \langle y, Ax - b \rangle = - \langle y, b \rangle + \langle c + yA, x \rangle \\ \phi(y) = \inf_{x \in P} L(x, y) = - \langle y, b \rangle + \inf \{ \langle c + yA, x \rangle : x \geq 0 \} \end{cases}$$

Was können wir über $\phi(y)$ aussagen?

- 1.Fall: Für fixiertes $y \in \mathbf{R}^m$ sei $c + yA \geq 0$. Dann ist $\langle c + yA, x \rangle \geq 0, \forall x \geq 0 \implies \inf \{ \langle c + yA, x \rangle : x \geq 0 \} = 0 \implies \phi(y) = - \langle y, b \rangle$.
- 2.Fall: Für fixiertes $y \in \mathbf{R}^m$ sei $c + yA \not\geq 0$. Dann ist $\inf \{ \langle c + yA, x \rangle : x \geq 0 \} = -\infty \implies \phi(y) = -\infty$.

Damit haben wir:

$$\phi(y) = \begin{cases} - \langle y, b \rangle & , \text{wenn } c + yA \geq 0 \\ -\infty & , \text{wenn } c + yA \not\geq 0 \end{cases}$$

Im Falle der Aufgabe (K) ist also

$$Q =_{Def} \{y \in Y : \phi(y) > -\infty\} = \{y \in \mathbf{R}^m : c + yA \geq 0\},$$

und damit ergibt sich als duale Aufgabe

$$(K_D) \quad \phi^* = \sup\{\langle -y, b \rangle : c + yA \geq 0\} = \sup\{\langle y, b \rangle : yA \leq c\}.$$

Bemerkung.

1. (K_D) ist also die Aufgabe

$$\begin{cases} \langle y, b \rangle \rightarrow \min \\ c + yA \geq 0 \end{cases}$$

und damit eine LOA.

2. Im Gegensatz zu vielen Aufgaben (MOA) ist die Duale zu (K) in expliziter Form aufschreibbar. Das gilt nicht nur für (K), sondern für alle LOA.
3. Auf analoge Weise, wie wir oben (K_D) aus (K) nach der Definition der Dualen hergeleitet haben, läßt sich die Duale zu einer beliebigen LOA erhalten. So ist

$$\begin{aligned} \phi^* = -\min\{ & \langle u, b \rangle + \langle v, d \rangle : c_j + \langle u, A_j \rangle + \langle v, B_j \rangle \geq 0, \\ & j = 1, \dots, t \text{ und } c_j + \langle u, A_j \rangle + \langle v, B_j \rangle = 0, \\ & j = t + 1, \dots, n \quad , v \geq 0\} \end{aligned}$$

die duale Aufgabe zu

$$f^* = \min\{\langle c, x \rangle : Ax = b, Bx \leq d, x_j \geq 0, \quad j = 1, \dots, t\}$$

(A_j, B_j bezeichne die j -te Spalte von A bzw. B).

4. Aus der dritten Bemerkung lassen sich folgende Merkgeln dazu ableiten, wie (K_D) aus (K) entsteht:

- rechte Seite wird zum Zielfunktionsvektor
- negativer Zielfunktionsvektor wird zur rechten Seite
- Nebenbedingungsmatrix wird zur transponierten Nebenbedingungsmatrix
- Nichtnegativitätsforderungen auf die Variablen werden zu Ungleichungsforderungen in den Nebenbedingungen

4.4 Optimalitäts- und Lösbarkeitskriterien für LOA

Wir wollen uns wieder auf LOA in der (K)-Form beschränken:

$$(K) \quad f^* = \inf\{\langle c, x \rangle : x \in S\},$$

wobei $S = \{x \in \mathbf{R}^n : Ax = b, x \geq 0\}$. Im Abschnitt 1.4 hatten wir mit Satz 1.9 bereits ein Optimalitätskriterium für Basislösungen von (K) bewiesen. Wir wollen hier zeigen, daß es sich auf beliebige zulässige Lösungen von (K) übertragen läßt.

Satz (4.5). (Optimalitätskriterien)

Es sei x^* zulässige Lösung von Aufgabe (K). Dann sind folgende 3 Behauptungen äquivalent:

- (i) x^* ist optimale Lösung von (K)
- (ii) Es existiert eine zulässige Lösung y^* der dualen Aufgabe (K_D) , für die $\langle c + y^*A, x^* \rangle = 0$.
- (iii) Es existiert eine zulässige Lösung y^* der dualen Aufgabe (K_D) , für die $\langle c, x^* \rangle + \langle y^*, b \rangle = 0$.

Bemerkung.

1. Wie aus dem Beweis folgt, ist unter den Bedingungen (ii) und (iii) y^* eine optimale Lösung von (K_D) .
2. Nach den Standardumformungen, wie wir sie im Abschnitt 1.4 zur Simplexmethode kennengelernt haben, kann man Aufgabe (K_D) eine äquivalente Aufgabe in (K)-Form zuordnen. Schließlich zeigt man, daß die Duale zu dieser Aufgabe (d.h die Duale von der Dualen) eine Aufgabe ist, die äquivalent zur Ausgangsaufgabe (K) ist. Aus diesen beiden Ergebnissen ist klar, daß ein symmetrisches Resultat zu Satz 4.5 gelten muß.

Satz (4.5').

Es sei y^* eine zulässige Lösung von Aufgabe (K_D) . Dann sind folgende 3 Behauptungen äquivalent:

- (i) y^* ist optimale Lösung von (K_D)
- (ii) Es existiert eine zulässige Lösung x^* von Aufgabe (K), für die $\langle c + y^*A, x^* \rangle = 0$

- (iii) Es existiert eine zulässige Lösung x^* von Aufgabe (K), für die $\langle c, x^* \rangle + \langle y^*, b \rangle = 0$.

Satz (4.6). (Lösbarkeitskriterien)

Wir betrachten die zueinander dualen Aufgaben

$$(K) \quad f^* = \inf\{\langle c, x \rangle : x \in S\} \text{ und}$$

$$(K_D) \quad \phi^* = \sup\{-\langle y, b \rangle : y \in Q\}$$

mit $S = \{x \in \mathbf{R}^n : Ax = b, x \geq 0\}$, $Q = \{y \in \mathbf{R}^m : c + yA \geq 0\}$. Dann sind folgende Aussagen äquivalent:

- (i) (K) besitzt eine optimale Lösung.
- (ii) (K_D) besitzt eine optimale Lösung.
- (iii) $S \neq \emptyset, Q \neq \emptyset$
- (iv) $Q \neq \emptyset, \phi^* = \sup\{\langle -y, b \rangle : y \in Q\} < +\infty$
- (v) $S \neq \emptyset, f^* = \inf\{\langle c, x \rangle : x \in S\} > -\infty$.

Folgerung (4.7).

Aufgabe (K) ist genau dann lösbar, wenn Aufgabe (K_D) lösbar ist.

Folgerung (4.8). (Starke Dualität in LOA's)

Wenn wenigstens eine der Bedingungen (i) - (v) von Satz 4.6 erfüllt ist, dann gilt

- (i) $+\infty > f^* = \phi^* > -\infty$
- (ii) $\forall x \in S$ und $\forall y \in Q$ gilt $\langle c, x \rangle \geq f^* = \phi^* \geq -\langle y, b \rangle$.

Satz (4.9).

Für die Aufgaben (K) und (K_D) gilt:

- (i) Wenn $S \neq \emptyset$, aber $Q = \emptyset$, dann ist $f^* = \phi^* = -\infty$.
- (ii) Wenn $S = \emptyset$, aber $Q \neq \emptyset$, dann ist $f^* = \phi^* = +\infty$.
- (iii) Wenn $f^* = -\infty$, dann ist $Q = \emptyset$.
- (iv) Wenn $\phi^* = +\infty$, dann ist $S = \emptyset$.

Bemerkung.

1. Der Fall $S = \emptyset$ und $Q = \emptyset$ (d.h. $f^* = +\infty, \phi^* = -\infty$) ist, wie das folgende Beispiel zeigt, möglich (vergewissern Sie sich durch analytische oder geometrische Betrachtungen):

$$(K) \begin{cases} -x_1 - 3x_2 \rightarrow \min \\ x_1 - 2x_2 + x_3 = 1 \\ -x_1 + 2x_2 + x_4 = -3 \\ x_j \geq 0, \quad j = 1, 2, 3, 4 \end{cases} \quad (K_D) \begin{cases} -y_1 + 3y_2 \rightarrow \max \\ -1 + y_1 - y_2 \geq 0 \\ -3 - 2y_1 + 2y_2 \geq 0 \\ y_1 \geq 0, y_2 \geq 0 \end{cases}$$

2. Wie die Folgerungen und Sätze zeigen, ist außer in dem Fall der 1. Bemerkung (d.h. bei $S = \emptyset, Q = \emptyset$) in der Linearen Optimierung stets $f^* = \phi^*$, d.h. bei LOA sind die Fälle:
- $+\infty > f^* > \phi^* = -\infty,$
 - $+\infty = f^* > \phi^* > -\infty,$
 - $+\infty > f^* > \phi^* > -\infty,$
- nicht möglich.