III Optimierung

Viele der interessanteren ökonomischen Optimierungsprobleme hängen von einer Vielzahl von Variablen ab. Ein Verbraucher wählt Mengen von verschiedenen Gütern, um einen möglichst großen Nutzen zu erreichen. Ein Unternehmen versucht die Kosten für Lohn, Lagerhaltung, Maschineneinsatz, Transport etc. für ein vorgegebenes Produktionsziel möglichst gering zu halten. Die mathematische Behandlung solcher Aufgaben erfordert die Bestimmung globaler bzw. lokaler Extrema von Funktionen mehrerer Variablen. Dabei können zusätzlich Nebenbedingungen/Restriktionen in Form von Gleichungen und Ungleichungen auftreten (begrenzte Ressourcen, Mindestproduktionsmenge, Budgetbeschränkung etc.). Allgemein stellt sich somit die Aufgabe:

Bestimme Minimum bzw. Maximum von $f(x_1, x_2, \dots, x_n)$,

so dass
$$\left.\begin{aligned} g_1(x_1, x_2, \dots, x_n) &= b_1 \\ &\vdots \\ g_r(x_1, x_2, \dots, x_n) &= b_r \end{aligned}\right\} \text{Gleichungsrestriktionen}$$
$$\left.\begin{aligned} h_1(x_1, x_2, \dots, x_n) &\le c_1 \\ &\vdots \\ h_s(x_1, x_2, \dots, x_n) &\le c_s \end{aligned}\right\} \text{Ungleichungsrestriktionen}$$

Bevor wir in diese Thematik einsteigen, stellen wir einige Grundlagen aus der Analysis mehrerer Variablen zusammen.

III.1 Grundlagen der Analysis mehrerer Variablen

Definition 3.1.1: Eine reelle Funktion f von n Variablen ist eine Zuordnung, die jedem geordneten n-Tupel $\vec{x}^T = (x_1, x_2, \dots, x_n)$ aus einer Menge $\mathbb{D}_f \subseteq \mathbb{R}^n$ eindeutig eine reelle Zahl, den Funktionswert zuordnet. $\mathbb{D}_f$ heißt Definitionsbereich von f, die Menge aller Funktionswerte $\mathbb{W}_f$ heißt Wertebereich von f. ($f: \mathbb{D}_f \subseteq \mathbb{R}^n \longrightarrow \mathbb{W}_f \subseteq \mathbb{R}$, $(x_1, x_2, \dots, x_n) \longmapsto f(x_1, x_2, \dots, x_n)$ oder auch kurz $\vec{x} \longmapsto f(\vec{x})$.)

Bezeichnung 3.1.2: Ist f eine Funktion von n Variablen $x_1, x_2, \ldots, x_n$, so bezeichnen wir häufig den Wert von f an einer Stelle $\vec{x}^T = (x_1, x_2, \ldots, x_n)$ mit $z = f(\vec{x})$. $x_1, x_2, \ldots, x_n$ heißen dann unabhängige Variable oder Argumente und z abhängige Variable von f. In den Wirtschaftswissenschaften verwendet man häufig auch die Bezeichnungen exogene Variablen für die unabhängigen Variablen und endogene Variable für die abhängige Variable.

Bemerkung 3.1.3: Wichtig bei der Definition des Funktionsbegriffs ist die Forderung nach der Eindeutigkeit der Zuordnung.

Wir vereinbaren, dass der Definitionsbereich aus allen $\vec{x} \in \mathbb{R}^n$ bestehen soll, für die die Funktion berechnet werden kann, es sei denn, dass ein anderer Definitionsbereich durch die Aufgabenstellung vorgegeben ist.

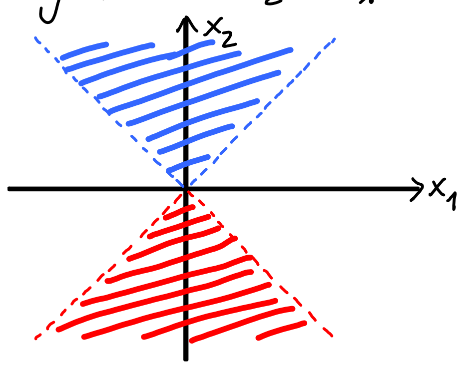
Beispiel 3.1.3: In einem landwirtschaftlichen Betrieb hängt die Produktion, d.h. die Anzahl produzierter Einheiten vom investierten Kapital K, dem Arbeitseinsatz L und der verwendeten Anbaufläche T ab. Häufig wird ein solcher Zusammenhang durch eine Cobb-Douglas-Funktion, d.h. eine Funktion des Typs $Y = A \cdot K^a \cdot L^b \cdot T^c$ mit positiven Konstanten A, a, b, c beschrieben. Hier sind K, L, T die unabhängigen Variablen, Y ist die abhängige Variable. Der Definitionsbereich ist $\mathbb{D}_Y = \{(K, L, T) \in \mathbb{R}^3 : K \geq 0, L \geq 0, T \geq 0\}$.

Beispiel 3.1.4: Bestimmung des Definitionsbereiches von $f(x_1, x_2) = \ln(x_2^2 - x_1^2)$.

Da der Logarithmus nur für positive Argumente definiert ist, muss gelten: $x_2^2 - x_1^2 > 0 \iff x_2^2 > x_1^2$

$$\iff |x_2| > |x_1|$$

$$\iff x_2 > |x_1| \lor x_2 < -|x_1|$$



Affin-lineare Funktionen und Polynome

Definition 3.1.5: Eine Funktion $f: \mathbb{R}^n \to \mathbb{R}$, $f(\vec{x}) = \sum_{i=1}^{n} a_i x_i + a_0$ mit beliebigen reellen Konstanten $a_0, a_1, \dots, a_{n-1}$ heißt affin-linear. Falls $a_0 = 0$ ist, heißt sie lineare Funktion.

Beispiel 3.1.6:

$f(x_1, x_2, x_3, x_4) = \sqrt{\pi}\, x_1 - 2x_2 + \sqrt{3}\, x_3 - 7x_4 + 1$ ist affin-linear.

$g(x_1, x_2, x_3, x_4) = x_1 - \sqrt{2}\, x_2 + 4 \cdot \sqrt{7}\, x_3 - \pi x_4$ ist linear.

Definition 3.1.7: Eine Summe von Ausdrücken der Form

$c \cdot x_1^{k_1} \cdot x_2^{k_2} \cdot \ldots \cdot x_n^{k_n}$ mit nichtnegativen, ganzzahligen Potenzen $k_1, k_2, \dots, k_n$ heißt Polynom. Der Grad des Polynoms ist die maximal vorkommende Summe der auftretenden Potenzen.

Beispiel 3.1.8: $f(x_1, x_2, x_3) = x_1^3 x_2^7 + x_1^2 x_2^2 x_3 + x_2 x_3$ ist Polynom vom Grad 10.

$g(x_1, x_2, x_3, x_4) = -\sqrt{2}\, x_1 + \sqrt{3}\, x_2 - \pi x_3 + x_4 - \pi$ ist Polynom vom Grad 1.

Niveaumengen

In den Grundvorlesungen zur Analysis von Funktionen mehrerer Variablen werden üblicherweise im Zusammenhang mit Visualisierungsmöglichkeiten von Funktionen mit zwei Variablen sogenannte Niveaulinien/Höhenlinien behandelt. Dies sind Linien, auf denen die Funktion einen konstanten Wert annimmt. In den Anwendungen haben diese häufig spezielle Namen wie Isokostenlinien, die die möglichen Kombinationen von Einsatzmengen zweier Produktionsfaktoren wie z.B. Kapital und Arbeitskraft darstellen, die die gleichen Kosten verursachen oder Isoquanten, die die möglichen Kombinationen zweier Produktionsfaktoren darstellen, die den gleichen Output erzeugen. Im Folgenden wollen wir die Begriffe allgemeiner fassen.

Definition 3.1.9: Sei $f: D_f \subseteq \mathbb{R}^n \to \mathbb{R}$ eine Funktion und $c \in \mathbb{R}$.

Dann heißt $N(f,c) = \{\vec{x} \in D_f : f(\vec{x}) = c\}$ Niveaumenge,

$L(f,c) = \{\vec{x} \in D_f : f(\vec{x}) \leq c\}$ untere Niveaumenge und

$U(f,c) = \{\vec{x} \in D_f : f(\vec{x}) \geq c\}$ obere Niveaumenge der Funktion f zum Niveau c.

Beispiel 3.1.10: Sei $f(x_1,x_2) = \sqrt{x_1^2 + x_2^2}$. Da $x_1^2 + x_2^2 \geq 0$ für alle $(x_1,x_2) \in \mathbb{R}^2$ und $\sqrt{x_1^2+x_2^2} \geq 0$, ist $D_f = \mathbb{R}^2$, $W_f = [0,\infty)$.

Zur Bestimmung der Niveaumengen zu einem Niveau c mit $c \geq 0$ sind diejenigen Wertepaare (x_1,x_2) zu bestimmen, für die

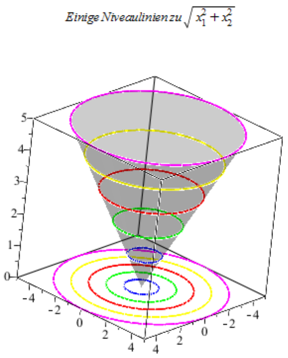
$$f(x_1,x_2) = c \iff \sqrt{x_1^2+x_2^2} = c \underset{\text{da } c \geq 0}{\iff} x_1^2 + x_2^2 = c^2$$

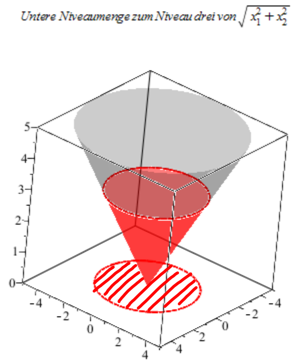
gilt. Die Niveaumengen beschreiben also Kreise um den Ursprung mit Radius c. Für $c \geq 0$ gilt also:

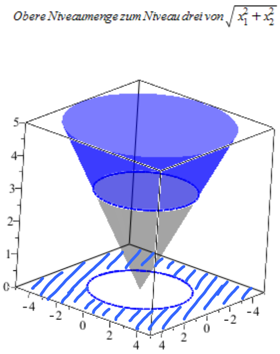
$$N(f,c) = \{(x_1,x_2) \in \mathbb{R}^2 : x_1^2 + x_2^2 = c^2\}$$

$$L(f,c) = \{(x_1,x_2) \in \mathbb{R}^2 : 0 \leq x_1^2 + x_2^2 \leq c^2\}$$

$$U(f,c) = \{(x_1,x_2) \in \mathbb{R}^2 : c^2 \leq x_1^2 + x_2^2\}$$





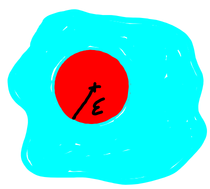


Abgeschlossene, offene, konvexe Mengen

Während bei Funktionen von einer Variablen häufig Intervalle als Teilmengen der reellen Zahlen betrachtet werden, die entweder abgeschlossen, offen oder halboffen sind, ist die Situation bei mehreren Variablen häufig komplizierter. Wir stellen daher einige grundlegende Begriffe zur Verfügung.

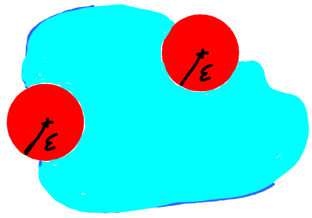
Definition 3.1.11: Sei S eine Teilmenge des $\mathbb{R}^n$.

1) Ein Punkt $(x_1^*,\dots,x_n^*) \in \mathbb{R}^n$ heißt innerer Punkt von S, wenn es ein $\varepsilon > 0$ gibt, so dass alle Punkte von $U_\varepsilon(\vec{x}^*) = \{\vec{x} \in \mathbb{R}^n : |\vec{x} - \vec{x}^*| < \varepsilon\}$ in S liegen.

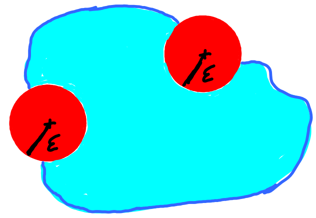


2) S heißt offen, wenn sie nur aus inneren Punkten besteht.

3) Ein Punkt $(x_1^*, \dots, x_n^*) \in \mathbb{R}^n$ heißt Randpunkt von S, wenn $U_\varepsilon(\vec{x}^*)$ für jedes $\varepsilon > 0$ sowohl Punkte von S als auch Punkte, die nicht zu S gehören, enthält.



4) S heißt abgeschlossen, wenn das Komplement von S in $\mathbb{R}^n$, d.h. $\mathbb{R}^n \setminus S$ offen ist. Das bedeutet, dass jeder Randpunkt von S zu S gehört.



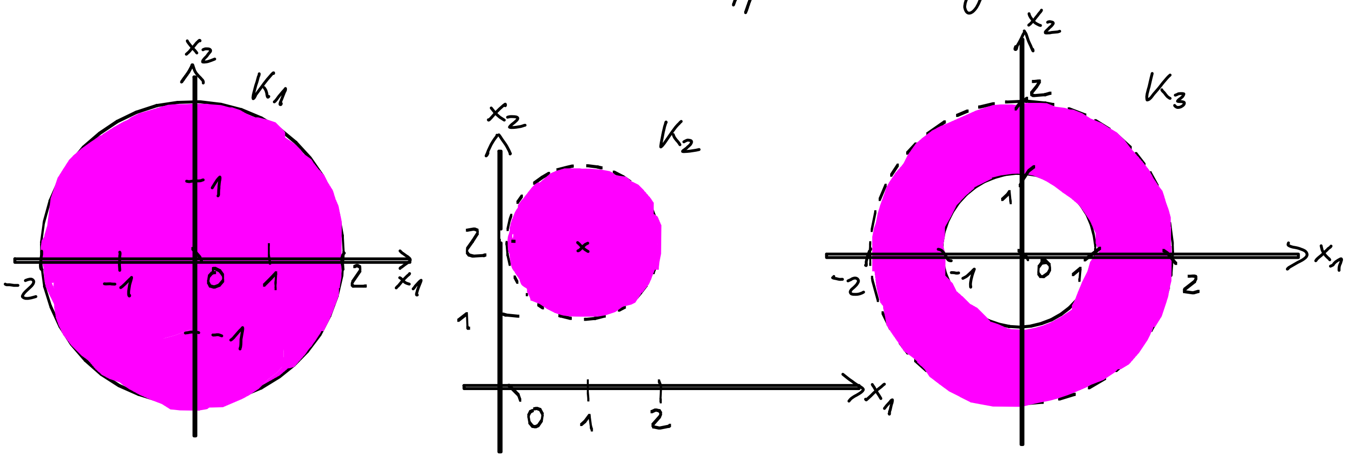
5) S heißt beschränkt, wenn es eine Kugel (hinreichend groß) gibt, die S enthält.

Beispiel 3.1.12:

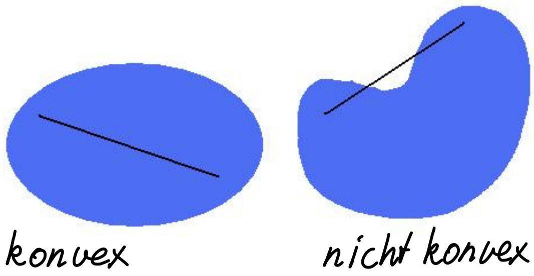
1) $K_1 = \{(x_1,x_2) \in \mathbb{R}^2 : x_1^2 + x_2^2 \le 4\}$ ist abgeschlossen.

2) $K_2 = \{(x_1,x_2) \in \mathbb{R}^2 : (x_1-1)^2 + (x_2-2)^2 < 1\}$ ist offen.

3) $K_3 = \{(x_1,x_2) \in \mathbb{R}^2 : 1 \le x_1^2 + x_2^2 < 4\}$ ist weder offen noch abgeschlossen.



Definition 3.1.13: Eine Menge $S \subseteq \mathbb{R}^n$ heißt konvex, wenn für alle $\vec{x}, \vec{y} \in S$ und jedes $\lambda \in [0,1]$ gilt: $\lambda \vec{x} + (1-\lambda)\vec{y} \in S$.

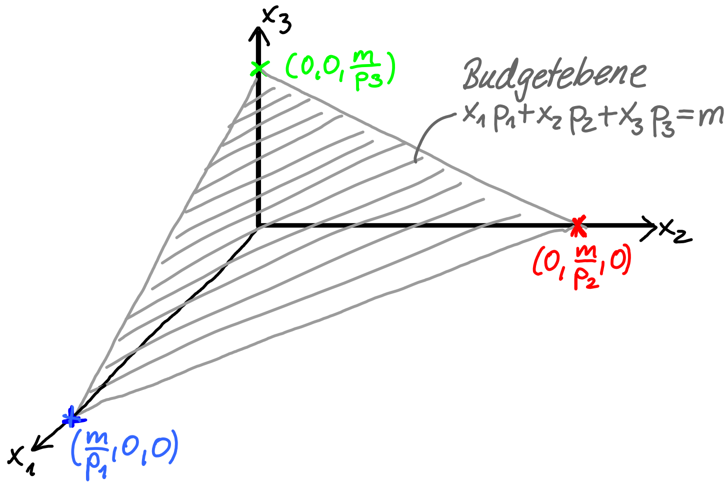


Anschaulich (im $\mathbb{R}^2, \mathbb{R}^3$) bedeutet Konvexität, dass mit je zwei Punkten der Menge S auch die komplette Verbindungsstrecke der beiden Punkte in S liegt.

In $\mathbb{R}$ ist jedes Intervall, egal ob abgeschlossen, offen oder halboffen, konvex.

Beispiel 3.1.14: Ein Verbraucher kauft Mengen x_1, x_2, x_3 von drei verschiedenen Gütern zu Preisen p_1, p_2, p_3 je Mengeneinheit. Dafür hat er ein Budget m zur Verfügung. Die Budgetmenge ist

$$B = \{(x_1, x_2, x_3) : x_1 \geq 0,\ x_2 \geq 0,\ x_3 \geq 0,\ x_1 p_1 + x_2 p_2 + x_3 p_3 \leq m\}$$



Die Budgetmenge bildet anschaulich einen dreiminensionalen konvexen Körper, der durch die Koordinatenebenen und die Budgetebene begrenzt ist.

Grundlegende Begriffe der Differentialrechnung

In diesem Abschnitt werden grundlegende Begriffe wie Stetigkeit, partielle Ableitungen, Richtungsableitungen etc. für Funktionen mehrerer Variablen zusammengestellt.

Definition 3.1.15: $z^* \in \mathbb{R}$ heißt Grenzwert von $f(\vec{x})$ für $\vec{x}$ gegen $\vec{x}^*$, wenn für alle Folgen $\{x_{i,m}\}_{m \in \mathbb{N}}$ mit $(x_{1,m}, x_{2,m}, \dots, x_{n,m}) \in D_f$ und $\lim\limits_{m \to \infty} x_{i,m} = x_i^*$ gilt, dass $\lim\limits_{m \to \infty} f(x_{1,m}, \dots, x_{n,m}) = z^*$ ist.

Man schreibt dann $\lim\limits_{\vec{x} \to \vec{x}^*} f(\vec{x}) = z^*$.

$f(\vec{x})$ heißt stetig im Punkt $\vec{x}^*$, wenn $\vec{x}^* \in D_f$ und $\lim\limits_{\vec{x} \to \vec{x}^*} f(\vec{x}) = f(\vec{x}^*)$ gilt.

Beispiel 3.1.16: Wir betrachten die sogenannte Parabelfalte

$$f(x_1, x_2) = \begin{cases} \dfrac{2x_1^2 x_2}{x_1^4 + x_2^2}, & x_1 \neq 0 \vee x_2 \neq 0 \\ 0, & x_1 = 0 \wedge x_2 = 0 \end{cases}$$

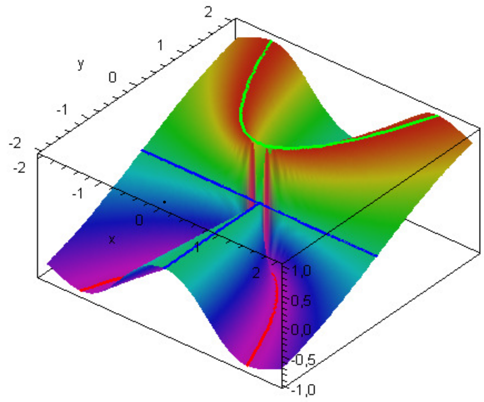
Offensichtlich ist hier die Stelle $(x_1^*, x_2^*) = (0,0)$ problematisch.

Wir untersuchen die Funktionswerte zu speziellen Linien im Definitionsbereich, die durch $(0,0)$ verlaufen. Speziell gilt:

- für $x_1 = 0$: $f(0, x_2) = 0$
- für $x_2 = 0$: $f(x_1, 0) = 0$
- für $x_2 = x_1^2$ mit $x_1 \neq 0$: $f(x_1, x_1^2) = \dfrac{2x_1^2 \cdot x_1^2}{x_1^4 + (x_1^2)^2} = 1$
- für $x_2 = -x_1^2$ mit $x_1 \neq 0$: $f(x_1, -x_1^2) = \dfrac{2 \cdot x_1^2 \cdot (-x_1^2)}{x_1^4 + (-x_1^2)^2} = -1$

Somit gilt z.B. $f(x_1, x_1^2) \longrightarrow 1$ für $x_1 \to 0$
oder auch $f(x_1, -x_1^2) \longrightarrow -1$ für $x_1 \to 0$
aber nach Definition der Funktion $f(0,0) = 0$, d.h. die Funktion ist unstetig an der Stelle $(x_1^*, x_2^*) = (0,0)$.

Parabelfalte



Die Überlegungen zeigen auch, dass es im Allgemeinen nicht reicht, das Verhalten in Richtung der Koordinatenachsen zu überprüfen.

Die Bestimmung momentaner Änderungsraten bzgl. einer Variablen x_i bedeutet die Berechnung der partiellen Ableitung nach dieser Variablen. Dabei werden alle anderen Variablen als konstant betrachtet und nach x_i differenziert.

Definition 3.1.17: Sei $f: D_f \subseteq \mathbb{R}^n \to \mathbb{R}$ und $\vec{x}^* \in D_f$.

1) f heißt an der Stelle $\vec{x}^*$ partiell nach x_i differenzierbar, wenn der Grenzwert
$$\lim_{h \to 0} \frac{f(\vec{x}^* + h \cdot \vec{e}_i) - f(\vec{x}^*)}{h} = \lim_{h \to 0} \frac{f(x_1, \ldots, x_{i-1}, x_i + h, x_{i+1}, \ldots, x_n) - f(x_1, \ldots, x_n)}{h}$$
existiert. Der Grenzwert wird dann mit $f_{x_i}(\vec{x}^*)$ oder mit $\frac{\partial f}{\partial x_i}(\vec{x}^*)$ bezeichnet und heißt 1. partielle Ableitung von f nach x_i an der Stelle $\vec{x}^*$.
Ist f an jeder Stelle $\vec{x} \in D_f$ partiell nach x_i differenzierbar, so heißt f_{x_i} bzw. $\frac{\partial f}{\partial x_i}$ erste partielle Ableitung von f nach x_i.

2) Ist f an allen Stellen $\vec{x} \in D_f$ nach jeder Variablen x_i, $i = 1, 2, \ldots, n$, partiell differenzierbar, so heißt die Funktion partiell differenzierbar. Sind zusätzlich die partiellen Ableitungen f_{x_i}, $i = 1, 2, \ldots, n$, stetig, so heißt f stetig

partiell differenzierbar.

3) Partielle Ableitungen höherer Ordnung ergeben sich iterativ als partielle Ableitungen niedrigerer Ordung.

4) Mit $C^k(U), k \geq 1$, bezeichnet man die Menge aller f, für die alle partiellen Ableitungen bis k-ter Ordnung existieren und stetig sind.

Regeln 3.1.18: Zusammenstellung der wichtigsten Ableitungsregeln.

1) Faktoren, die unabhängig von x_i sind, bleiben beim Differenzieren nach x_i erhalten.

2) Die partielle Ableitung einer Summe (Differenz) ist die Summe (Differenz) der partiellen Ableitungen

3) Produktregel: $\frac{\partial}{\partial x_i}\{u(\vec{x})\cdot v(\vec{x})\} = u_{x_i}(\vec{x})\, v(\vec{x}) + u(\vec{x})\cdot v_{x_i}(\vec{x})$

4) Quotientenregel: $\frac{\partial}{\partial x_i}\left\{\frac{u(\vec{x})}{v(\vec{x})}\right\} = \frac{u_{x_i}(\vec{x})\, v(\vec{x}) - u(\vec{x})\, v_{x_i}(\vec{x})}{[v(\vec{x})]^2}$

5) Einfache Kettenregel: $f(\vec{x}) = u(v(\vec{x}))$, $v: \mathbb{D}_v \subseteq \mathbb{R}^n \to \mathbb{R}$, $u: \mathbb{D}_u \subseteq \mathbb{R} \to \mathbb{R}$

$$f_{x_i}(\vec{x}) = \underbrace{\frac{du}{dv}(v(\vec{x}))}_{\text{"Äußere Ableitung"}} \cdot \underbrace{\frac{\partial v}{\partial x_i}(\vec{x})}_{\text{"Innere Ableitung"}}$$

Beispiel 3.1.19: $f(x_1, x_2, x_3) = x_1^2 \cdot x_2^3 \cdot \ln(x_1 + x_2 x_3)$

$$f_{x_1}(x_1, x_2, x_3) = x_2^3\left\{2x_1 \cdot \ln(x_1 + x_2 x_3) + x_1^2 \cdot \frac{1}{x_1 + x_2 x_3}\right\}$$
$$= x_1 x_2^3 \left\{2\ln(x_1 + x_2 x_3) + \frac{x_1}{x_1 + x_2 x_3}\right\}$$
$$f_{x_2}(x_1, x_2, x_3) = x_1^2\left\{3x_2^2 \ln(x_1 + x_2 x_3) + x_2^3 \cdot \frac{1}{x_1 + x_2 x_3} \cdot x_3\right\}$$
$$= x_1^2 x_2^2 \left\{3\ln(x_1 + x_2 x_3) + \frac{x_2 x_3}{x_1 + x_2 x_3}\right\}$$
$$f_{x_3}(x_1, x_2, x_3) = x_1^2 \cdot x_2^3 \cdot \frac{1}{x_1 + x_2 x_3} \cdot x_2 = \frac{x_1^2 x_2^4}{x_1 + x_2 x_3}$$

In vielen Fällen kann bei partiellen Ableitungen höherer Ordnung die Reihenfolge, in der die partiellen Ableitungen berechnet werden, vertauscht werden. Der folgende Satz gibt eine hinreichende Bedingung an, die für unsere Zwecke in der Regel erfüllt ist.

Satz von Schwarz 3.1.20: Bei einer gemischten partiellen Ableitung k-ter Ordnung darf die Reihenfolge der einzelnen Differentiationsschritte vertauscht werden, wenn die partiellen Ableitungen k-ter Ordnung stetig sind.

Beispiel 3.1.21: $f(x,y) = \dfrac{x^2 y}{x+y^2}$

$$f_x = \frac{x^2y + 2xy^3}{(x+y^2)^2}, \quad f_y = \frac{x^3 - x^2y^2}{(x+y^2)^2}$$

$$f_{xx} = \frac{2y^5}{(x+y^2)^3}, \quad f_{xy} = \frac{x^3 + 3x^2y^2 - 2xy^4}{(x+y^2)^3} = f_{yx}, \quad f_{yy} = \frac{2x^2y^3 - 6x^3y}{(x+y^2)^3}$$

Partielle Ableitungen erster bzw. zweiter Ordnung werden häufig in einem Vektor bzw. in einer Matrix zusammengefasst.

Definition 3.1.22: Sei $f: D_f \subseteq \mathbb{R}^n \to \mathbb{R}$ zweimal partiell differenzierbar. Dann ist der Gradient von f definiert durch

$$\nabla f(\vec{x}) = \operatorname{grad} f(\vec{x}) = \begin{pmatrix} f_{x_1}(\vec{x}) \\ f_{x_2}(\vec{x}) \\ \vdots \\ f_{x_n}(\vec{x}) \end{pmatrix}$$

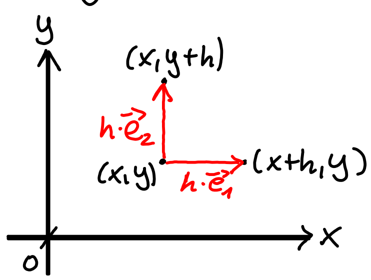
und die Hesse-Matrix von f durch

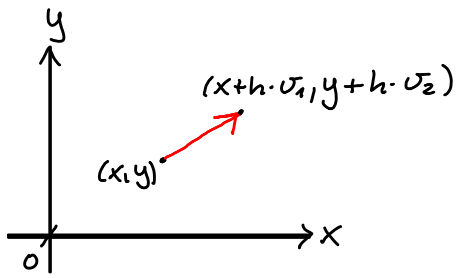
$$H_f(\vec{x}) = \begin{pmatrix} f_{x_1x_1}(\vec{x}) & f_{x_1x_2}(\vec{x}) & \cdots & f_{x_1x_n}(\vec{x}) \\ f_{x_2x_1}(\vec{x}) & f_{x_2x_2}(\vec{x}) & \cdots & f_{x_2x_n}(\vec{x}) \\ \vdots & \vdots & & \vdots \\ f_{x_nx_1}(\vec{x}) & f_{x_nx_2}(\vec{x}) & \cdots & f_{x_nx_n}(\vec{x}) \end{pmatrix}$$

Beispiel 3.1.23: $f(x,y) = \sin(xy)$

$$\nabla f(x,y) = \begin{pmatrix} y\cos(xy) \\ x\cos(xy) \end{pmatrix}, \quad H_f(x,y) = \begin{pmatrix} -y^2\sin(xy) & -xy\sin(xy) + \cos(xy) \\ -xy\sin(xy) + \cos(xy) & -x^2\sin(xy) \end{pmatrix}$$

Während man bei partiellen Ableitungen momentane Änderungsraten in Richtung der Koordinatenachsen, d.h. in Richtung der kartesischen Einheitsvektoren $\vec{e}_i = (0, \dots, 0, \overset{i\text{-te Stelle}}{1}, 0, \dots, 0)^T$ berechnet, werden im allgemeinen Fall der Richtungsableitung beliebige Richtungen betrachtet.





Definition 3.1.24: Sei $f: D_f \subseteq \mathbb{R}^n \to \mathbb{R}$ und $\vec{v} \in \mathbb{R}^n$ mit $|\vec{v}| = 1$. Dann ist die Richtungsableitung von f in Richtung $\vec{v}$ definiert durch

$$\lim_{h \to 0} \frac{f(\vec{x} + h\vec{v}) - f(\vec{x})}{h} = \partial_{\vec{v}} f(\vec{x}), \text{ falls der Grenzwert existiert.}$$

Wählt man speziell für $\vec{v}$ einen der kartesichen Einheitsvektoren $\vec{e}_i$, so erhält man die partiellen Ableitungen als spezielle Richtungsableitungen, d.h. $\partial_{\vec{e}_i} f(\vec{x}) = f_{x_i}(\vec{x})$.

Der folgende Satz ist für praktische Rechnungen von großer Bedeutung.

Satz 3.1.25: Sei $f: D_f \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}$ stetig partiell differenzierbar und $\vec{v} \in \mathbb{R}^n$ mit $|\vec{v}| = 1$. Dann gilt

$$\partial_{\vec{v}} f(\vec{x}) = \langle \vec{v}, \nabla f(\vec{x}) \rangle = \sum_{i=1}^{n} v_i f_{x_i}(\vec{x})$$

Beispiel 3.1.26: $f(x,y) = \ln(y^2 - x)$, $\vec{v} = \frac{1}{5}\begin{pmatrix} 4 \\ 3 \end{pmatrix}$.

Es gilt $|\vec{v}| = \frac{1}{5}\sqrt{16+9} = 1$.

Mit $\nabla f(x,y) = \begin{pmatrix} -\frac{1}{y^2-x} \\ \frac{2y}{y^2-x} \end{pmatrix} = \frac{1}{y^2-x} \cdot \begin{pmatrix} -1 \\ 2y \end{pmatrix}$

ist
$$\partial_{\vec{v}} f(x,y) = \left\langle \frac{1}{5}\begin{pmatrix} 4 \\ 3 \end{pmatrix}, \frac{1}{y^2-x}\begin{pmatrix} -1 \\ 2y \end{pmatrix} \right\rangle = \frac{6y-4}{5(y^2-x)}$$

An der Stelle $(x^*, y^*) = (-1, 2)$ gilt z.B. $f_x(-1,2) = -\frac{1}{5}$, $f_y(-1,2) = \frac{4}{5}$, $\partial_{\vec{v}}(-1,2) = \frac{8}{25}$.

Der folgende Satz beinhaltet zwei wichtige Eigenschaften des Gradienten.

Satz 3.1.27: Sei $f \in C^1(A)$, A offene Menge, $\vec{x}^* \in A$ mit $\nabla f(\vec{x}^*) \neq \vec{0}$. Dann gilt:

1) $\nabla f(\vec{x}^*)$ bzw. $-\nabla f(\vec{x}^*)$ zeigt in Richtung des steilsten Anstiegs bzw. Abstiegs von f.
2) $\nabla f(\vec{x}^*)$ ist orthogonal zu der Niveaumenge $N(f, f(\vec{x}^*)) = \{x \in D_f : f(\vec{x}) = f(\vec{x}^*)\}$.

III.2 Einführung zur Optimierung

Eine Optimierungsaufgabe besteht in der Minimierung bzw. Maximierung einer reellen Funktion, der Zielfunktion auf einer gegebenen Menge, der Menge zulässiger Punkte. Beispiele für solche Aufgabenstellungen sind:

- Produktionsplanung unter Beachtung von Maschinenkapazitäten, verfügbaren Arbeitskräften, etc., so dass der Gewinn maximal wird,
- Transport bestimmter Mengen eines Produktes von verschiedenen Produktions- und Lagerstätten zu verschiedenen Verbrauchern, so dass die Transportkosten minimal werden.

Zum Einstieg behandeln wir zunächst ein Beispiel, das sich geometrisch lösen lässt.

Beispiel 3.2.1: In einem landwirtschaftlichen Betrieb sollen Weizen und Kartoffeln produziert werden. Für 1 Morgen Anbaufläche benötigt man zur Erzeugung von

- Kartoffeln: 5 € Anbaukosten, 2h Arbeitszeit
- Weizen : 10 € Anbaukosten, 10h Arbeitszeit

Der Reingewinn für Kartoffeln beträgt 20€ pro Morgen, für Weizen 60 € pro Morgen. Für die Gesamtproduktion stehen 1200 Morgen, 7000 € und 5200 h Arbeitszeit zur Verfügung.

Die Produktion soll nun so gestaltet werden, dass der Reingewinn maximal wird.

Wir beschreiben die Aufgabenstellung zunächst durch ein mathematisches Modell. Dazu bezeichnen wir mit x_1 die Anbaufläche für Kartoffeln und mit x_2 die für Weizen.

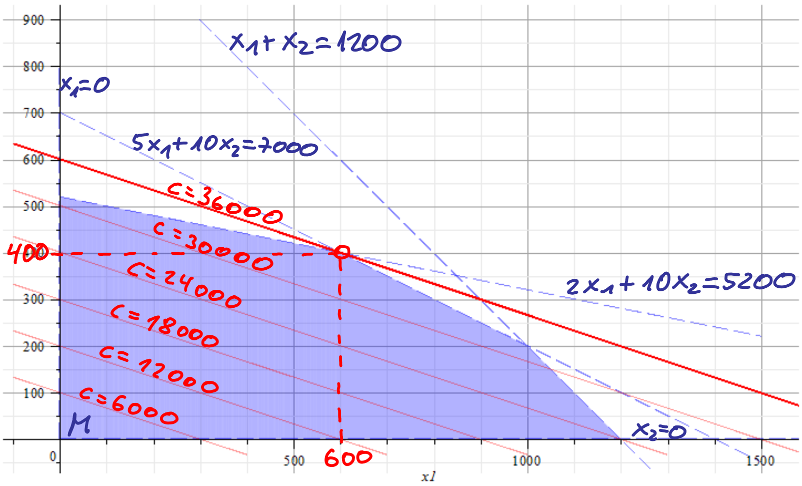
Ziel ist es nun, die Zielfunktion (Gewinnfunktion)

$$f(x_1, x_2) = 20x_1 + 60x_2$$

zu maximieren, so dass die Restriktionen (Nebenbedingungen)

$$\begin{aligned} x_1 + x_2 &\leq 1200 && \text{(maximale Anbaufläche)} \\ 5x_1 + 10x_2 &\leq 7000 && \text{(maximale Anbaukosten)} \\ 2x_1 + 10x_2 &\leq 5200 && \text{(maximale Arbeitsstunden)} \\ x_1 \geq 0,\ & x_2 \geq 0 \end{aligned}$$

erfüllt sind. Die Restriktionen beschreiben die Menge $M \subseteq \mathbb{R}^2$ der zulässigen Punkte. M kann hier in der x_1x_2-Ebene dargestellt werden. Die Zielfunktion beschreibt hier eine Ebene im $\mathbb{R}^3$. Gesucht ist das Maximum über dem zulässigen Bereich M. Um dies geometrisch zu bestimmen, zeichnet man Niveaulinien $20x_1 + 60x_2 = c$ der Zielfunktion für verschiedene c. In diesem Beispiel sind dies parallele Geraden. Das Maximum der Zielfunktion wird da angenommen, wo die Niveaulinie mit maximal möglichem c, den zulässigen Bereich gerade noch berührt.

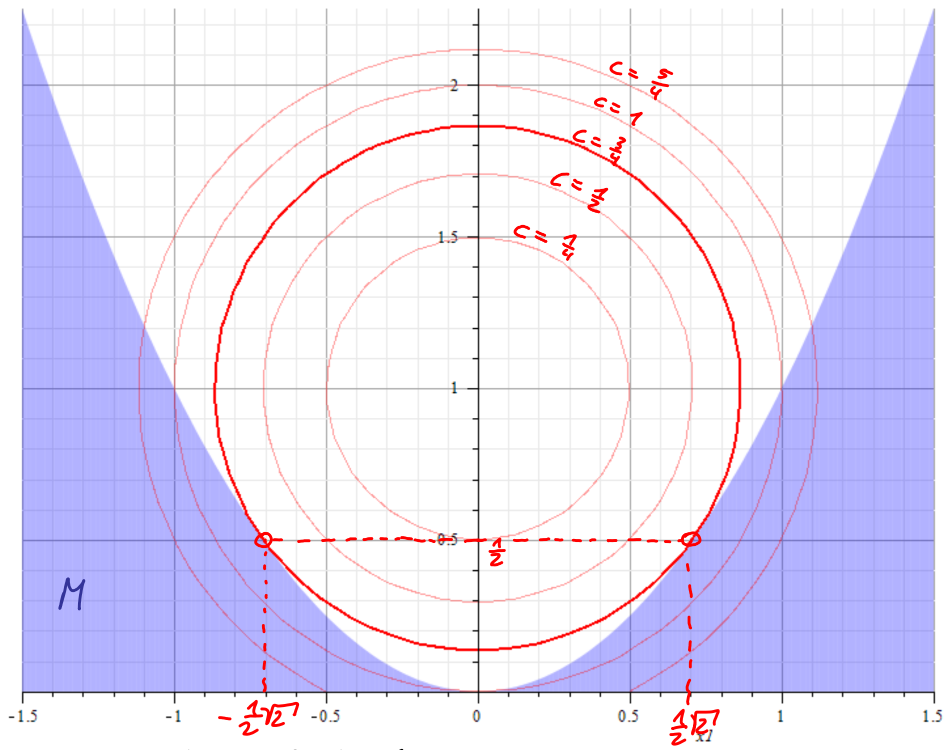


600 Morgen Kartoffeln und 400 Morgen Weizen ergeben also unter Einhaltung der Restriktionen einen maximalen Reingewinn von 36 000 €.

Das folgende Beispiel lässt sich zwar noch anschaulich erklären, zeigt aber auch bereits, dass man allein mit graphischen Methoden nicht auskommt.

Beispiel 3.2.2: Wir suchen das Minimum von $f(x_1, x_2) = x_1^2 + (x_2 - 1)^2$ unter den Restriktionen $x_2 \leq x_1^2$ und $x_2 \geq 0$.

Die Niveaulinien von f sind gegeben durch $x_1^2 + (x_2-1)^2 = c$, d.h. für $c > 0$ Kreise um $(0,1)$ mit Radius $\sqrt{c}$! Mit wachsendem c wird f größer. Zu bestimmen ist also derjenige Kreis, der den zulässigen Bereich berührt. Das Minimum von f unter Einhaltung der Restriktionen wird also für



$x_1 = \pm\frac{1}{2}\sqrt{2}$, $x_2 = \frac{1}{2}$ angenommen und beträgt $f(\pm\frac{1}{2}\sqrt{2}, \frac{1}{2}) = \frac{3}{4}$.

Praktische Problemstellungen beinhalten oft eine Vielzahl von Variablen und Restriktionen, so dass man für die Lösungen entsprechende mathematische Methoden benötigt. Wir geben zunächst eine allgemeine mathematische Formulierung der Aufgabenstellung an.

Optimierungsaufgabe 3.2.3: Sei $f: \mathbb{D}_f \subseteq \mathbb{R}^n \to \mathbb{R}$, $\mathbb{D}_f$ offene Menge und $M \subseteq \mathbb{D}_f$. Dann ist die Aufgabe:

$$(P)\begin{cases} \min f(\vec{x}), \\ \text{so dass } \vec{x} \in M. \end{cases}$$

Die Funktion f heißt Zielfunktion oder Kostenfunktion, M heißt zulässige Menge und die Elemente von M zulässige Punkte.

Ist $M = \mathbb{D}_f$, so spricht man von einem unrestringierten Optimierungsproblem. Wird M durch Nebenbedingungen (in Form von Gleichungen oder Ungleichungen) definiert, so spricht man von einem restringierten Optimierungsproblem oder Optimierungsproblem mit Restriktionen (Nebenbedingungen).

Bemerkung 3.2.4: Es gibt natürlich auch Problemstellungen, in denen nicht Minima, sondern Maxima gesucht werden. Da dies aber äquivalent dazu ist, Minima von $g = -f$ zu bestimmen, werden häufig nur Minimierungsaufgaben betrachtet.

Wir legen nun fest, was unter lokalen bzw. globalen Minima zu verstehen ist.

Definition 3.2.5: $\vec{x}^* \in M$ heißt lokales Minimum von f auf M oder lokale Lösung von (P), falls es ein $\varepsilon > 0$ gibt, so dass

$$f(\vec{x}) \geq f(\vec{x}^*) \quad \text{für alle } \vec{x} \in M \cap U_\varepsilon(\vec{x}^*).$$

$\vec{x}^* \in M$ heißt striktes lokales Minimum von f auf M oder strikte lokale Lösung von (P), wenn es ein $\varepsilon > 0$ gibt, so dass

$$f(\vec{x}) > f(\vec{x}^*) \quad \text{für alle } \vec{x} \in M \cap U_\varepsilon(\vec{x}^*), \vec{x} \neq \vec{x}^*.$$

$\vec{x}^* \in M$ heißt globales Minimum von f auf M oder globale Lösung von (P), wenn

$$f(\vec{x}) \geq f(\vec{x}^*) \text{ für alle } \vec{x} \in M.$$

$\vec{x}^* \in M$ heißt striktes globales Minimum von f auf M oder strikte globale Lösung von (P), wenn

$$f(\vec{x}) > f(\vec{x}^*) \text{ für alle } \vec{x} \in M, \vec{x} \neq \vec{x}^*.$$

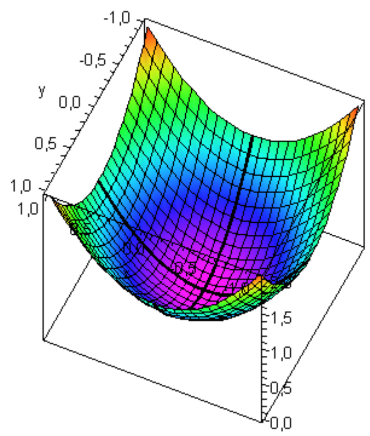
$f(\vec{x}^*)$ heißt Optimalwert, wenn $\vec{x}^*$ lokale oder globale Lösung von (P) ist.

III.3 Unrestringierte Optimierung

In diesem Kapitel betrachten wir Optimierungsprobleme ohne Nebenbedingungen und beschäftigen uns mit Fragen der Existenz von Minima, sowie mit notwendigen und hinreichenden Kriterien unter geeigneten Differenzierbarkeitsvoraussetzungen an die Zielfunktion f. Zum Teil handelt es sich dabei um Verallgemeinerungen von Funktionen mit 2 Variablen auf Funktionen mehrerer Variablen.

Anschaulich ist zunächst klar, dass bei einer differenzierbaren Funktion einer Variablen die Tangente an die Kurve an einem Minimum im Inneren des Definitionsbereichs parallel zur x_1-Achse verläuft.

Ebenso ist die Tangentialebene an einem Minimum im Inneren des Definitionsbereichs einer Funktion von zwei Variablen parallel zur x_1x_2-Ebene. Das bedeutet, dass dort insbesondere die partiellen Ableitungen (Steigungen in Richtung der Koordinatenachsen) Null sein müssen.



Dieser Sachverhalt gilt analog auch für Funktionen von n Variablen, die momentanen Änderungsraten müssen notwendigerweise für jede Variable Null sein, wenn ein Minimum vorliegen soll.

Satz 3.3.1: Notwendige Bedingung 1. Ordnung

Sei $f: \mathbb{D}_f \subseteq \mathbb{R}^n \to \mathbb{R}$, $\vec{x}^* \in D$, $D \subseteq \mathbb{D}_f$ offen und f in $\vec{x}^*$ differenzierbar. Besitzt f an der Stelle $\vec{x}^*$ ein lokales Minimum, dann gilt

$$\nabla f(\vec{x}^*) = \vec{0}.$$

Durch Bestimmung aller $\vec{x}^*$, für die $\nabla f(\vec{x}^*) = \vec{0}$ gilt, erhält man also alle möglichen Kandidaten im Inneren von $\mathbb{D}_f$ für Minimal-, bzw. allgemeiner Extremalstellen. Punkte $\vec{x}^*$ mit $\nabla f(\vec{x}^*) = \vec{0}$ heißen auch stationäre Punkte. Man spricht von notwendiger Bedingung 1. Ordnung, weil partielle Ableitungen erster Ordnung betrachtet werden.

Beispiel 3.3.2: Wir bestimmen die stationären Punkte von

$$f(x_1,x_2,x_3) = x_1^2 + 2x_2^2 + 2x_3^2 + 2x_1x_2 - 2x_2x_3 - 2x_1 + 2x_2 - 4x_3 + 5$$

$$\nabla f(\vec{x}) = \begin{pmatrix} 2x_1 + 2x_2 - 2 \\ 4x_2 + 2x_1 - 2x_3 + 2 \\ 4x_3 - 2x_2 - 4 \end{pmatrix}$$

$$\nabla f(\vec{x}) = \vec{0} \iff \left\{ \begin{array}{rcl} x_1 + x_2 & = & 1 \\ x_1 + 2x_2 - x_3 & = & -1 \\ -x_2 + 2x_3 & = & 2 \end{array} \right\}$$

Dies ist ein lineares Gleichungssystem. Die eindeutig bestimmte Lösung ist $x_1 = 3$, $x_2 = -2$, $x_3 = 0$. Die Funktion besitzt also nur einen stationären Punkt, nämlich $P(3,-2,0)$. Ob f dort ein Minimum besitzt, ist ohne weitere Untersuchungen nicht zu entscheiden.

Besitzt eine Funktion einer Variablen an einer Stelle eine waagerechte Tangente und ist die Funktion in einem Bereich um diese Stelle strikt konvex, so liegt ein Minimum vor. Auch die dick eingezeichneten Kurven im Bild auf S. 14, die durch das Minimum verlaufen sind strikt konvex. Offenbar spielt die Konvexität im Zusammenhang mit der Untersuchung von Minimalstellen eine wichtige Rolle. Wir werden daher zunächst die Definition auf Funktionen von n Variablen verallgemeinern und Kriterien angeben, falls die Funktion geeignete Differenzierbarkeitsvoraussetzungen erfüllt.

Definition 3.3.3: Sei f definiert auf einer konvexen Menge S. Dann heißt f

1) $\left\{\begin{matrix}\text{konvex}\\ \text{konkav}\end{matrix}\right\}$ auf S, wenn $f(\lambda\vec{x}+(1-\lambda)\vec{y})\left\{\begin{matrix}\leq\\ \geq\end{matrix}\right\}\lambda f(\vec{x})+(1-\lambda)f(\vec{y})$

für alle $\vec{x},\vec{y}\in S$ und $\lambda\in[0,1]$,

2) $\left\{\begin{matrix}\text{strikt konvex}\\ \text{strikt konkav}\end{matrix}\right\}$ auf S, wenn $f(\lambda\vec{x}+(1-\lambda)\vec{y})\left\{\begin{matrix}<\\ >\end{matrix}\right\}\lambda f(\vec{x})+(1-\lambda)f(\vec{y})$

für alle $\vec{x},\vec{y}\in S$ mit $\vec{x}\neq\vec{y}$ und $\lambda\in(0,1)$,

Für Funktionen von einer bzw. zwei Variablen bedeutet strikte Konvexität anschaulich, dass die Verbindungsstrecke zwischen zwei beliebigen verschiedenen Punkten stets oberhalb des Funktionsgraphen verläuft.

Beispiel 3.3.4: Wir zeigen, dass die Funktion $f(x)=x^2-1$ auf der konvexen Menge $S=D_f=\mathbb{R}$ strikt konvex ist.

$$(*)\quad f(\lambda x+(1-\lambda)y)<\lambda f(x)+(1-\lambda)f(y)$$

$$\iff [\lambda x+(1-\lambda)y]^2-1<\lambda[x^2-1]+(1-\lambda)[y^2-1]$$

$$\iff \lambda^2x^2+2\lambda(1-\lambda)xy+(1-\lambda)^2y^2-1<\lambda x^2-\lambda+(1-\lambda)y^2-(1-\lambda)$$

$$\iff 0<\underbrace{[\lambda-\lambda^2]}_{=\lambda(1-\lambda)}x^2-2\lambda(1-\lambda)xy+\underbrace{[(1-\lambda)-(1-\lambda)^2]}_{=(1-\lambda)[1-(1-\lambda)]=\lambda(1-\lambda)}y^2$$

$$\iff 0<\lambda(1-\lambda)(x-y)^2$$

Da $\lambda(1-\lambda)>0$ für jedes $\lambda\in(0,1)$ und $(x-y)^2>0$ für $x\neq y$, ist die Ungleichung $(*)$ für alle $\lambda\in(0,1)$, $x,y\in\mathbb{R}$ mit $x\neq y$ erfüllt. Also ist f strikt konvex auf $\mathbb{R}$.

Das Beispiel zeigt, dass das Nachprüfen der Konvexität an Hand der Definition schon bei einfachen Funktionen aufwändig ist. Im Falle geeigneter Differenzierbarkeitseigenschaften gibt es einfachere Kriterien. Bei einer Funktion einer Variablen wissen wir, dass sie strikt konvex ist, wenn ihre zweite Ableitung positiv ist. Bei Funktionen mehrerer Variablen wird die Rolle der zweiten Ableitung von der Hesse-Matrix übernommen.

Aus der positiven (negativen) Definitheit der Hesse-Matrix folgt, dass die Funktion strikt konvex (konkav) ist.

Satz 3.3.5: Sei f eine C^2-Funktion auf einer offenen, konvexen Menge $S \subseteq \mathbb{R}^n$ und $\vec{x}^* \in S$ ein stationärer Punkt von f. Dann gilt:

Ist die Hessematrix von f an der Stelle $\vec{x}^*$ ($H_f(\vec{x}^*)$) positiv definit, dann besitzt f an der Stelle $\vec{x}^*$ ein striktes lokales Minimum.

Ist die Hessematrix von f an der Stelle $\vec{x}^*$ negativ definit, dann besitzt f an der Stelle $\vec{x}^*$ ein striktes lokales Maximum.

Zur Überprüfung der positiven Definitheit dient das Hurwitz-Kriterium (vgl. auch Definition 1.4.11)

Satz 3.3.6: Seien D_r, $r = 1, 2, \ldots, n$, die führenden Hauptminoren einer symmetrischen $n \times n$-Matrix A. Dann gilt:

A positiv definit $\Leftrightarrow D_r > 0$ für alle $r = 1, 2, \ldots, n$

A negativ definit $\Leftrightarrow (-1)^r D_r > 0$ für alle $r = 1, 2, \ldots, n$

Beispiel 3.3.7: $f(x_1, x_2, x_3) = x_2^3 + x_1^2 + 3x_3^2 - 2x_1x_2 - x_2$

Wir bestimmen zunächst die stationären Punkte.

$$\operatorname{grad} f(\vec{x}) = \begin{pmatrix} 2x_1 - 2x_2 \\ 3x_2^2 - 2x_1 - 1 \\ 6x_3 \end{pmatrix}$$

$$\operatorname{grad} f(\vec{x}) = 0 \Leftrightarrow \left\{ \begin{array}{l} 1)\ 2x_1 - 2x_2 = 0 \\ 2)\ 3x_2^2 - 2x_1 - 1 = 0 \\ 3)\ 6x_3 = 0 \end{array} \right\}$$

3) liefert: $x_3 = 0$ 1) liefert: $x_1 = x_2$

$x_1 = x_2$ in 2) liefert: $3x_1^2 - 2x_1 - 1 = 0 \Leftrightarrow x_1^2 - \frac{2}{3}x_1 - \frac{1}{3} = 0$

$$\Leftrightarrow x_1 = \frac{1}{3} \pm \sqrt{\frac{1}{9} + \frac{1}{3}}$$

$$\Leftrightarrow x_1 = -\frac{1}{3} \vee x_1 = 1$$

Die stationären Punkte sind also: $P(-\frac{1}{3}, -\frac{1}{3}, 0)$, $Q(1, 1, 0)$.

Die Hesse-Matrix ist gegeben durch $-\cancel{6}^2\frac{1}{3}$

$$H_f(\vec{x}) = \begin{pmatrix} 2 & -2 & 0 \\ -2 & 6x_2 & 0 \\ 0 & 0 & 6 \end{pmatrix}, \quad H_f(-\tfrac{1}{3}, -\tfrac{1}{3}, 0) = \begin{pmatrix} 2 & -2 & 0 \\ -2 & -2 & 0 \\ 0 & 0 & 6 \end{pmatrix}, \quad H_f(1,1,0) = \begin{pmatrix} 2 & -2 & 0 \\ -2 & 6 & 0 \\ 0 & 0 & 6 \end{pmatrix}$$

Da $2 > 0$, $\begin{vmatrix} 2 & -2 \\ -2 & -2 \end{vmatrix} = -8 < 0$ ist $H_f(-\frac{1}{3}, -\frac{1}{3}, 0)$ weder positiv noch negativ definit. Satz 3.3.5 liefert also keine Aussage.

Da $2 > 0$, $\begin{vmatrix} 2 & -2 \\ -2 & 6 \end{vmatrix} = 8 > 0$, $\begin{vmatrix} 2 & -2 & 0 \\ -2 & 6 & 0 \\ 0 & 0 & 6 \end{vmatrix} = 48 > 0$, ist $H_f(1,1,0)$ nach Satz 3.3.6 positiv definit. Nach Satz 3.3.5 besitzt f an der Stelle $(1,1,0)$ ein lokales Minimum $f(1,1,0) = -1$.

Bei den hinreichenden Bedingungen haben wir gesehen, dass das Konvexitätsverhalten eine wichtige Rolle spielt. Wir untersuchen daher im Folgenden das Verhalten konvexer (konkaver) Funktionen.

Satz 3.3.8: Sei $f: D_f \subseteq \mathbb{R}^n \to \mathbb{R}$, D_f konvex, eine C^2-Funktion. Dann gilt:

1) $H_f(\vec{x})$ positiv semidefinit für alle $\vec{x} \in D_f$ $\Longleftrightarrow$ f konvex
2) $H_f(\vec{x})$ positiv definit für alle $\vec{x} \in D_f$ $\Longrightarrow$ f strikt konvex
3) $H_f(\vec{x})$ negativ semidefinit für alle $\vec{x} \in D_f$ $\Longleftrightarrow$ f konkav
4) $H_f(\vec{x})$ negativ definit für alle $\vec{x} \in D_f$ $\Longrightarrow$ f strikt konkav

Wir ergänzen Satz 3.3.6 noch um Kriterien für Semidefinitheit.

Satz 3.3.9: Sei A eine symmetrische $n \times n$-Matrix. Dann gilt:

A positiv semidefinit $\Longleftrightarrow$ $\tilde{D}_r \geq 0$ für alle Hauptminoren $\tilde{D}_r$ von A und alle $r = 1, 2, \ldots, n$

A negativ semidefinit $\Longleftrightarrow$ $(-1)^r \tilde{D}_r \geq 0$ für alle Hauptminoren $\tilde{D}_r$ von A und alle $r = 1, 2, \ldots, n$

Der Vorteil konvexer (konkaver) Funktionen zeigt sich in dem folgenden Satz.

Satz 3.3.10: Sei $f: D_f \subseteq \mathbb{R}^n \to \mathbb{R}$, D_f konvex. Dann gilt:

1) Ist f konvex, dann ist jedes lokale Minimum von f globales Minimum.
2) Ist f konkav, dann ist jedes lokale Maximum von f globales Maximum.
3) Ist f strikt konvex und besitzt f ein Minimum, dann ist dies ein striktes globales Minimum und eindeutig bestimmt.
4) Ist f strikt konkav und besitzt f ein Maximum, dann ist dies ein striktes globales Maximum und eindeutig bestimmt.

Im Folgenden betrachten wir nun einige Beispiele mit anwendungsorientiertem Hintergrund.

Beispiel 3.3.11: Ein Unternehmen verkauft ein Produkt auf zwei verschiedenen Märkten 1 und 2 zu Preisen P_1 und P_2, wobei $P_1 = a_1 - b_1 Q_1$, $P_2 = a_2 - b_2 Q_2$ mit Konstanten $a_1, a_2, b_1, b_2 > 0$ sein soll und Q_1, Q_2 die nachgefragten Mengen bezeichnen. Als einfaches Modell für die Kosten betrachten wir die Funktion $K(Q_1, Q_2) = \alpha(Q_1 + Q_2)$, $\alpha > 0$.

Der Gesamtgewinn beträgt dann:

$$\begin{aligned} G(Q_1, Q_2) &= P_1 Q_1 + P_2 Q_2 - \alpha(Q_1 + Q_2) \\ &= (a_1 - b_1 Q_1) Q_1 + (a_2 - b_2 Q_2) Q_2 - \alpha(Q_1 + Q_2) \end{aligned}$$

Ziel ist es, diejenigen Werte für $Q_1, Q_2 > 0$ zu bestimmen, für die die Gewinnfunktion maximal wird. Dies ist äquivalent dazu, $Q_1, Q_2 > 0$ zu bestimmen, für die $f(Q_1, Q_2) = -G(Q_1, Q_2)$ minimal wird.

Dabei ist $D_f = \{(Q_1, Q_2) \in \mathbb{R}^2 : Q_1 > 0 \wedge Q_2 > 0\}$ eine konvexe Menge.

Als notwendige Bedingung für ein lokales Extremum ermitteln wir:

$$\operatorname{grad} f(Q_1, Q_2) = \begin{pmatrix} 2b_1 Q_1 - a_1 + \alpha \\ 2b_2 Q_2 - a_2 + \alpha \end{pmatrix} = \vec{0} \iff \begin{pmatrix} 2b_1 & 0 \\ 0 & 2b_2 \end{pmatrix} \begin{pmatrix} Q_1 \\ Q_2 \end{pmatrix} = \begin{pmatrix} a_1 - \alpha \\ a_2 - \alpha \end{pmatrix}$$

Die eindeutig bestimmte Lösung dieses linearen Gleichungssystems und somit einziger Kandidat für eine lokale Extremalstelle ist:

$$(Q_1^*, Q_2^*) = \left(\frac{a_1 - \alpha}{2b_1}, \frac{a_2 - \alpha}{2b_2} \right), \text{ falls } a_1 > \alpha \text{ und } a_2 > \alpha.$$

Weiter ist $H_f(Q_1, Q_2) = \begin{pmatrix} 2b_1 & 0 \\ 0 & 2b_2 \end{pmatrix}$, d.h. $f_{Q_1 Q_1}(Q_1, Q_2) = 2b_1 > 0$ und

$\Delta_f(Q_1, Q_2) = 4b_1^2 > 0$. Nach dem Hurwitz-Kriterium ist die Hesse-Matrix positiv definit, d.h. f strikt konvex auf D_f.

Insgesamt besitzt somit f an der Stelle (Q_1^*, Q_2^*) ein eindeutig bestimmtes globales Minimum.

$$f(Q_1^*, Q_2^*) = (b_1 Q_1^* - a_1 + \alpha) Q_1^* + (b_2 Q_2^* - a_2 + \alpha) Q_2^*$$
$$= \left(\frac{a_1-\alpha}{2} - a_1 + \alpha\right) \cdot \frac{a_1-\alpha}{2b_1} + \left(\frac{a_2-\alpha}{2} - a_2 + \alpha\right) \cdot \frac{a_2-\alpha}{2b_2}$$
$$= -\frac{(a_1-\alpha)^2}{4b_1} - \frac{(a_2-\alpha)^2}{4b_2}$$

Der maximal mögliche Gewinn beträgt somit

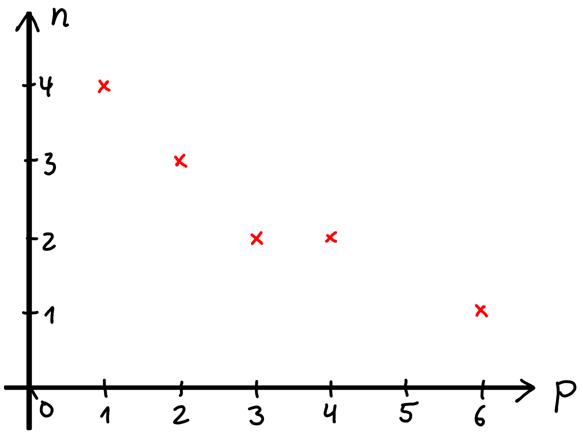
$$G(Q_1^*, Q_2^*) = -f(Q_1^*, Q_2^*) = \frac{(a_1-\alpha)^2}{4b_1} + \frac{(a_2-\alpha)^2}{4b_2}$$

für die Mengen $Q_1^* = \frac{a_1-\alpha}{2b_1}$, $Q_2^* = \frac{a_2-\alpha}{2b_2}$ zu den Preisen $P_1^* = \frac{a_1+\alpha}{2}$, $P_2^* = \frac{a_2+\alpha}{2}$.

Eine wichtige Anwendung im Zusammenhang mit diesem Abschnitt ist die lineare Regression, von der Sie vielleicht im Bereich Statistik schon einmal gehört haben. Bevor wir uns einer allgemeinen Lösung der Problemstellung zuwenden, betrachten wir die Aufgabenstellung und mögliche Lösungsansätze an einem Beispiel.

Beispiel 3.3.12: Eine Marktanalyse hat folgende Daten $(p_i; d_i)$, $i = 1, 2, \ldots, 5$, für die Nachfrage nach einem Produkt in Abhängigkeit vom Preis ergeben.

i	1	2	3	4	5
p_i	1	2	3	4	6
d_i	4	3	2	2	1



Die Skizze legt die Vermutung nahe, dass sich die Nachfrage in Abhängigkeit vom Preis durch eine affin-lineare Funktion der Form $D(p) = \alpha + \beta \cdot p$ beschreiben lässt. Nun liegen aber offensichtlich nicht alle Punkte auf ein und derselben Geraden. Da es sich um Messdaten handelt, die in der Regel nicht exakt sondern fehlerbehaftet sind, kann man dies auch nicht erwarten. Man versucht daher, eine Gerade so zu bestimmen, dass die Abweichungen zu den Messdaten "möglichst gering" ausfallen. Ein gängiges Kriterium für Optimalität (eine möglichst gute Gerade) ist die Minimierung der Summe der Fehlerquadrate. Man sucht also $\alpha, \beta \in \mathbb{R}$, so dass $f(\alpha,\beta) = \sum_{i=1}^{5} (D(p_i) - d_i)^2 = \sum_{i=1}^{5} (\alpha + \beta p_i - d_i)^2$ minimal wird.

Zunächst gilt: $f_\alpha(\alpha,\beta) = \sum_{i=1}^{5} 2(\alpha + \beta p_i - d_i)$, $f_\beta(\alpha,\beta) = \sum_{i=1}^{5} 2p_i(\alpha+\beta p_i - d_i)$,
d.h. nach Einsetzen der Daten aus der Tabelle

$$f_\alpha(\alpha,\beta) = 2(\alpha+\beta-4) + 2(\alpha+2\beta-3) + 2(\alpha+3\beta-2) + 2(\alpha+4\beta-2) + 2(\alpha+6\beta-1)$$
$$= 2(5\alpha + 16\beta - 12)$$

$$f_\beta(\alpha,\beta) = 2(\alpha+\beta-4) + 4(\alpha+2\beta-3) + 6(\alpha+3\beta-2) + 8(\alpha+4\beta-2) + 12(\alpha+6\beta-1)$$
$$= 2(16\alpha + 66\beta - 30)$$

Die notwendige Bedingung für ein lokales Extremum ist somit:

$$\operatorname{grad} f(\alpha,\beta) = \vec{0} \iff \left\{\begin{array}{l} 5\alpha + 16\beta - 12 = 0 \\ 16\alpha + 66\beta - 30 = 0 \end{array}\right\} \iff \begin{pmatrix} 5 & 16 \\ 16 & 66 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} 12 \\ 30 \end{pmatrix}$$

Die eindeutig bestimmte Lösung des Gleichungssystems ist $\alpha^* = \frac{156}{37}$, $\beta^* = -\frac{21}{37}$

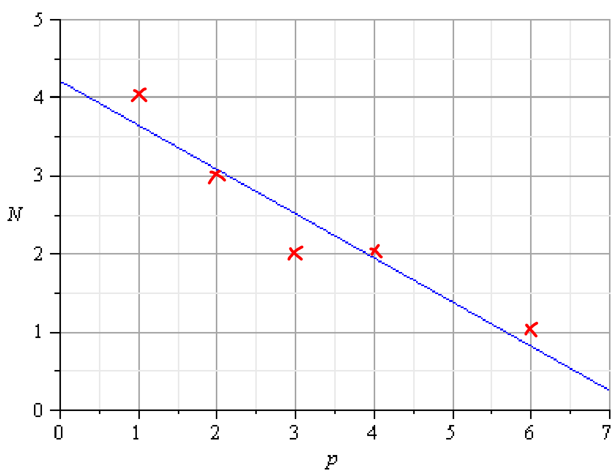
Weiter gilt:

$$H_f(\alpha,\beta) = \begin{pmatrix} 10 & 32 \\ 32 & 132 \end{pmatrix}, \quad \Delta_f(\alpha,\beta) = 296 > 0.$$

Da außerdem $f_{\alpha\alpha}(\alpha,\beta) = 10 > 0$, folgt, dass f strikt konvex ist auf der konvexen Menge $D_f = \mathbb{R}^2$. f nimmt also an der eindeutig bestimmten Stelle (α^*, β^*) sein globales Minimum an.

Die affin-lineare Funktion, die die Summe der Fehlerquadrate minimiert, ist also eindeutig bestimmt durch

$$D(p) = \frac{156}{37} - \frac{21}{37} p.$$



Lineare Regression 3.3.13: Den Vorgang der Minimierung der Summe der Fehlerquadrate bei der Approximation von Messdaten durch eine affin-lineare Funktion bezeichnet man mit linearer Regression oder auch allgemeiner mit Approximation im quadratischen Mittel.

Um nun ein allgemeines Konzept entwickeln zu können, benötigen wir einige Notationen.

Gegeben seien m Messdaten (x_i, y_i), $i = 1, 2, \ldots, m$. Der funktionale Zusammenhang zwischen den x- und y-Werten soll durch eine affin-lineare Funktion $Y(x) = \alpha + \beta x$ so beschrieben werden, dass die Summe der Fehlerquadrate $f(\alpha, \beta)$ minimal wird. Es ist

$$f(\alpha, \beta) = \sum_{i=1}^{m} (Y(x_i) - y_i)^2 = \sum_{i=1}^{m} (\alpha + \beta x_i - y_i)^2$$

Die ersten partiellen Ableitungen sind:

$$f_\alpha(\alpha, \beta) = \sum_{i=1}^{m} 2(\alpha + \beta x_i - y_i) = 2\left(m\alpha + \beta \sum_{i=1}^{m} x_i - \sum_{i=1}^{m} y_i\right)$$

$$f_\beta(\alpha, \beta) = \sum_{i=1}^{m} 2(\alpha + \beta x_i - y_i) \cdot x_i = 2\left(\alpha \sum_{i=1}^{m} x_i + \beta \sum_{i=1}^{m} x_i^2 - \sum_{i=1}^{m} x_i y_i\right)$$

Unter Verwendung der in der Statistik üblichen Abkürzungen $\bar{x} = \frac{1}{m} \sum_{i=1}^{m} x_i$, $\bar{y} = \frac{1}{m} \sum_{i=1}^{m} y_i$ für die Mittelwerte ergibt sich daraus:

$$f_\alpha(\alpha, \beta) = 2(m\alpha + m\beta\bar{x} - m\bar{y}), \quad f_\beta(\alpha, \beta) = 2\left(m\alpha\bar{x} + \beta \sum_{i=1}^{m} x_i^2 - \sum_{i=1}^{m} x_i y_i\right)$$

Die notwendige Bedingung $\operatorname{grad} f(\alpha, \beta) = \vec{0}$ führt also auf das lineare Gleichungssystem:

1) $\alpha + \beta \bar{x} = \bar{y}$

2) $\alpha m \bar{x} + \beta \sum_{i=1}^{m} x_i^2 = \sum_{i=1}^{m} x_i y_i$

mit der eindeutig bestimmten Lösung

$$\alpha^* = \frac{\bar{y} \sum_{i=1}^{m} x_i^2 - \bar{x} \sum_{i=1}^{m} x_i y_i}{\sum_{i=1}^{m} x_i^2 - m\bar{x}^2}, \quad \beta^* = \frac{\sum_{i=1}^{m} x_i y_i - m\bar{x}\bar{y}}{\sum_{i=1}^{m} x_i^2 - m\bar{x}^2}$$

Für die Hesse-Matrix von f gilt:

$$H_f(\alpha,\beta) = \begin{pmatrix} 2m & 2m\bar{x} \\ 2m\bar{x} & 2\sum_{i=1}^{m} x_i^2 \end{pmatrix}, \quad \Delta_f(\alpha,\beta) = 4m\left\{\sum_{i=1}^{m} x_i^2 - m\bar{x}^2\right\}$$

Um eine Aussage über das Vorzeichen von $\Delta_f(\alpha,\beta)$ machen zu können, zeigen wir, dass der Ausdruck in der geschweiften Klammer mit $\sum_{i=1}^{m}(x_i-\bar{x})^2$ übereinstimmt. Es gilt:

$$\begin{aligned}\sum_{i=1}^{m}(x_i-\bar{x})^2 &= \sum_{i=1}^{m}\left(x_i^2 - 2\bar{x}x_i + \bar{x}^2\right) \\ &= \sum_{i=1}^{m} x_i^2 - 2\bar{x}\underbrace{\sum_{i=1}^{m} x_i}_{=m\cdot\bar{x}} + \bar{x}^2\underbrace{\sum_{i=1}^{m} 1}_{=m} \\ &= \sum_{i=1}^{m} x_i^2 - 2m\bar{x}^2 + m\bar{x}^2 = \sum_{i=1}^{m} x_i^2 - m\bar{x}^2.\end{aligned}$$

Also gilt $\Delta_f(\alpha,\beta) = 4m\sum_{i=1}^{m}(x_i-\bar{x})^2$. Die Determinante der Hesse-Matrix ist somit genau dann größer als Null, wenn mindestens zwei der x_i verschieden sind. Für den praktischen Gebrauch kann man dies aber ohne Einschränkung voraussetzen. (Überlegen Sie einmal selbst, wie Messdaten (x_i, y_i) im $\mathbb{R}^2$ verteilt wären, bei denen alle x_i gleich sind!).
Da außerdem $f_{\alpha\alpha}(\alpha,\beta) = 2m > 0$ ist, ist insgesamt die Hesse-Matrix für alle $(\alpha,\beta)\in\mathbb{R}^2$ positiv definit. Also ist f strikt konvex auf ihrem konvexen Definitionsbereich $D_f = \mathbb{R}^2$. f nimmt somit für die oben angegebenen α^*, β^* ihr globales Minimum an.

Bemerkung 3.3.14: Durch Einführung weiterer Größen aus der Statistik lassen sich α^*, β^* in kürzerer Form angeben.
Wir bezeichnen mit

$$s_x^2 = \frac{1}{m-1}\sum_{i=1}^{m}(x_i-\bar{x})^2,\ s_y^2 = \frac{1}{m-1}\sum_{i=1}^{m}(y_i-\bar{y})^2,\ s_{xy} = \frac{1}{m-1}\sum_{i=1}^{m}(x_i-\bar{x})(y_i-\bar{y}),\ r = \frac{s_{xy}}{s_x s_y}$$

die Schätzwerte für die Varianzen, die Kovarianz und den Korrelationskoeffizienten. Dann gilt: $\alpha^* = \bar{y} - r\cdot\frac{s_y}{s_x}\cdot\bar{x}$, $\beta^* = r\cdot\frac{s_y}{s_x}$

Zusammengefasst haben wir also folgende Ergebnisse:

Sind (x_i, y_i), $i = 1, 2, \ldots, m$ Messdaten, wobei mindestens zwei der x_i verschieden sein sollen. Dann ist die lineare Regressionsgerade eindeutig bestimmt und gegeben durch die Gleichung

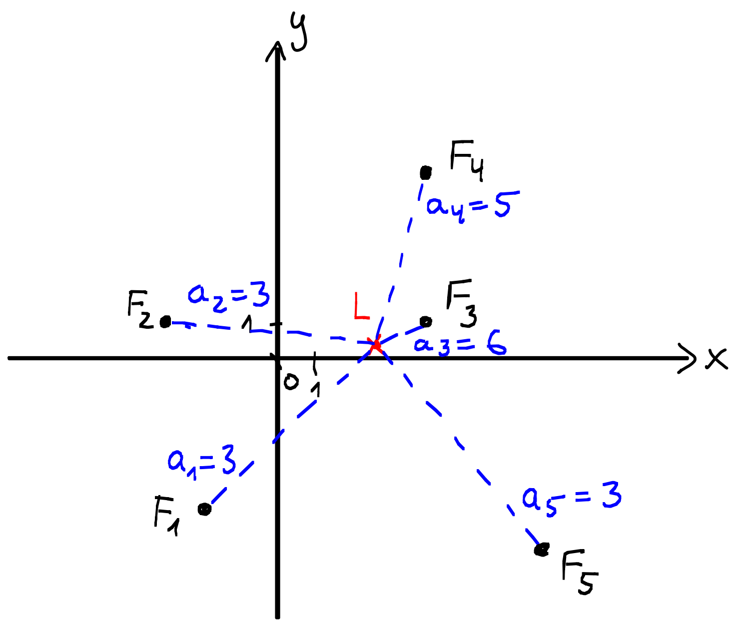
$$y(x) = \left(\bar{y} - r \cdot \frac{s_y}{s_x} \cdot \bar{x}\right) + r \cdot \frac{s_y}{s_x} \cdot x$$

Nach den obigen Erläuterungen braucht man zur Bestimmung der Koeffizienten der Regressionsgeraden nur das lineare Gleichungssystem zu lösen, das sich aus den notwendigen Bedingungen ergibt.

Beispiel 3.3.15: Standortoptimierung

Ein Unternehmen hat 5 Produktionsstätten F_i an verschiedenen Orten, deren Lage durch die Koordinaten x_i und y_i gegeben sind als:

i	1	2	3	4	5
(x_i, y_i)	$(-2,-4)$	$(-3,1)$	$(4,1)$	$(4,5)$	$(7,-5)$



Es wird nun nach einem optimalen Standort (x,y) für ein Ersatzteillager L gesucht. Die zu erwartenden Mengen, die pro Planungsperiode an die Produktionsstätten F_i auszuliefern sind, sind gegeben durch:

$a_1 = a_2 = 3$, $a_3 = 6$, $a_4 = 5$, $a_5 = 3$

Wir gehen davon aus, dass die Transportkosten von L zu F_i linear mit der Transportmenge a_i und quadratisch mit der Entfernung

$d_i = \left\| \begin{pmatrix} x-x_i \\ y-y_i \end{pmatrix} \right\| = \sqrt{(x-x_i)^2+(y-y_i)^2}$ zwischen L und F_i zunehmen, d.h.

wir nehmen an, dass die Transportkosten durch

$$f(x,y) = \sum_{i=1}^{5} a_i d_i^2 = \sum_{i=1}^{5} a_i \{(x-x_i)^2+(y-y_i)^2\}$$

gegeben sind. Der optimale Standort (x,y) ist nun so zu bestimmen, dass die Fahrtkosten minimal werden.

Als notwendige Bedingungen erhalten wir:

$$f_x(x,y) = \sum_{i=1}^{5} 2a_i(x-x_i) = 0 \iff x = \frac{\sum_{i=1}^{5} a_i x_i}{\sum_{i=1}^{5} a_i} = 2.5$$

$$f_y(x,y) = \sum_{i=1}^{5} 2a_i(y-y_i) = 0 \iff y = \frac{\sum_{i=1}^{5} a_i y_i}{\sum_{i=1}^{5} a_i} = 0.35$$

Für die Hesse-Matrix erhalten wir

$$H_f(x,y) = \begin{pmatrix} 2\sum_{i=1}^{5} a_i & 0 \\ 0 & 2\sum_{i=1}^{5} a_i \end{pmatrix} = \begin{pmatrix} 40 & 0 \\ 0 & 40 \end{pmatrix}.$$

Mit $f_{xx}(x,y) = 20 > 0$ und $\Delta_f(x,y) = 1600 > 0$ schließen wir, dass f strikt konvex ist auf der konvexen Menge $\mathbb{R}^2$.

Die Koordinaten für den optimalen Lagerstandort sind somit $(x^*,y^*) = (2.5, 0.35)$.

Hängen die Transportkosten nicht quadratisch, sondern linear von der Entfernung ab, lässt sich leider aus den notwendigen Bedingungen keine explizite Darstellung für x^* und y^* ablesen. Für solche Probleme werden dann iterative Verfahren verwendet, die z.B. die Lösung des oben betrachteten quadratischen Problems als Startnäherung verwenden.

III.4 Optimierung mit Gleichungsrestriktionen

Das in 3.2.3 allgemein gestellte Optimierungsproblem behandeln wir in diesem Abschnitt in der speziellen Form, dass die Nebenbedingungen (Beschreibung der Menge M der zulässigen Punkte) in Form von Gleichungen für die Variablen gegeben sind, d.h. wir behandeln die Aufgabe

$$\min f(\vec{x}),$$

$$\text{so dass} \quad g_1(\vec{x})=0, \ldots, g_m(\vec{x})=0 \quad (m<n)$$

Lagrangesches Multiplikatorverfahren

Bevor wir dieses Verfahren allgemein angeben, wollen wir an einem Beispiel anschaulich erläutern, worauf dieses Verfahren beruht.

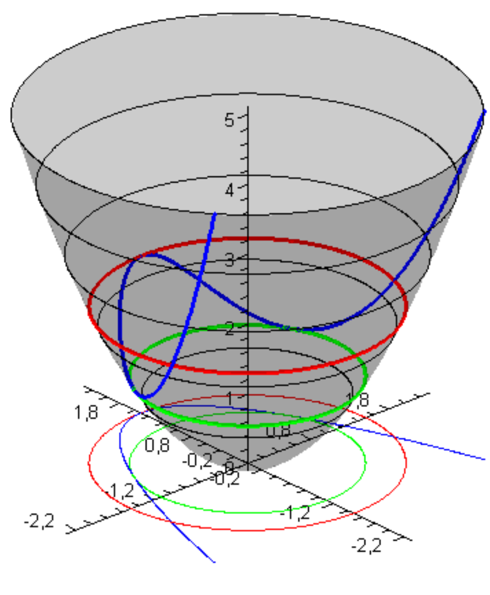
Beispiel 3.4.1: Wir betrachten das Problem

$$\min f(x_1,x_2) = x_1^2 + x_2^2$$

$$\text{so dass } g(x_1,x_2) = x_2 + x_1^2 - \tfrac{3}{2} = 0$$

zunächst an Hand einer Graphik.

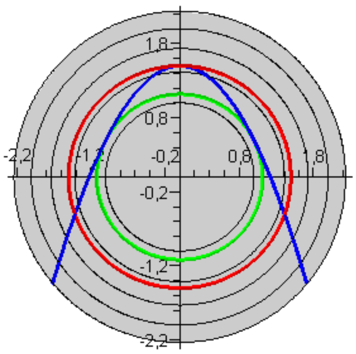
Die Funktion f beschreibt die grau schattierte Fläche im $\mathbb{R}^3$.



Extrema mit Restriktionenen

Zur Verdeutlichung sind einige Niveaulinien von f schwarz eingezeichnet. Durch $g(x_1,x_2)=0$ wird die in der x_1x_2-Ebene eingezeichnete blaue Parabel (die Menge der zulässigen Punkte) beschrieben. Schränkt man f auf diejenigen (x_1,x_2) ein, die die Restriktion erfüllen, so erhält man die auf der Fläche in blau hervorgehobene Kurve. Längs dieser Kurve sollen nun lokale Extrema bestimmt werden. Die lokalen Minima und das lokale Maximum sind in der Graphik deutlich zu erkennen. Die zugehörigen Niveaulinien sind auf der Fläche und in der x_1x_2-Ebene in grün bzw. in rot dargestellt. Man sieht, dass die blaue Kurve auf der Fläche die rote bzw. die grüne Niveaulinie berührt.

Wir betrachten die Situation in der x_1x_2-Ebene noch einmal genauer.



Extrema mit Restriktionenen

Die Restriktionskurve und die zu den lokalen Extremalstellen gehörenden Niveaulinien berühren sich an den lokalen Extremalstellen. Anders ausgedrückt, haben diese Niveaulinien und die Restriktionskurve an den Extremalstellen jeweils dieselbe Steigung.

Setzt man die Steigungen der implizit gegebenen Kurven $f(x_1,x_2)=c$ und $g(x_1,x_2)=0$ gleich so erhält man:

$$\text{(Steigung Niveaulinie)}\quad -\frac{f_{x_1}(x_1,x_2)}{f_{x_2}(x_1,x_2)} = -\frac{g_{x_1}(x_1,x_2)}{g_{x_2}(x_1,x_2)}\quad \text{(Steigung Restriktion)},$$

bzw. anders aufgeschrieben

$$-\frac{f_{x_1}(x_1,x_2)}{g_{x_1}(x_1,x_2)} = -\frac{f_{x_2}(x_1,x_2)}{g_{x_2}(x_1,x_2)}.$$

Diese Gleichheit kann man nun auch so beschreiben, dass für ein $\lambda\in\mathbb{R}$ die beiden Gleichungen

$$-\frac{f_{x_1}(x_1,x_2)}{g_{x_1}(x_1,x_2)} = \lambda \quad \text{und} \quad -\frac{f_{x_2}(x_1,x_2)}{g_{x_2}(x_1,x_2)} = \lambda$$

bzw. anders ausgedrückt

$$f_{x_1}(x_1,x_2)+\lambda g_{x_1}(x_1,x_2)=0 \quad \text{und} \quad f_{x_2}(x_1,x_2)+\lambda g_{x_2}(x_1,x_2)=0$$

erfüllt sein müssen.

Zusammen mit der Restriktion ergeben sich für unser Beispiel die drei Gleichungen:

1) $f_{x_1}(x_1,x_2)+\lambda g_{x_1}(x_1,x_2)=2x_1+2\lambda x_1=0 \iff x_1(1+\lambda)=0$

2) $f_{x_2}(x_1,x_2)+\lambda g_{x_2}(x_1,x_2)=2x_2+\lambda \quad =0 \iff \lambda=-2x_2$

3) $\qquad g(x_1,x_2)=x_2+x_1^2-\frac{3}{2} \quad =0$

Wegen 1) $x_1(1+\lambda)=0 \iff x_1=0 \vee \lambda=-1$, unterscheiden wir zwei Fälle:

1\. Fall: $\underline{x_1=0}$

Einsetzen in 3) liefert $\underline{x_2=\frac{3}{2}}$

Einsetzen in 2) liefert $\underline{\lambda=-3}$

2\. Fall: $\underline{\lambda=-1}$

Einsetzen in 2) liefert $\underline{x_2=\frac{1}{2}}$

Einsetzen in 3) liefert $\underline{x_1=1 \vee x_1=-1}$

Die Punkte $P_1(0,\frac{3}{2})$, $P_2(1,\frac{1}{2})$, $P_3(-1,\frac{1}{2})$ kommen also als Kandidaten für Extremalstellen in Frage.

Es gilt: $f(0,\frac{3}{2})=\frac{9}{4}$ ist lokales Maximum, $f(1,\frac{1}{2})=f(-1,\frac{1}{2})=\frac{5}{4}$ sind lokale Minima von f unter der gegebenen Restriktion.

Bevor wir die Methode, die uns notwendige Bedingungen für lokale Extrema bei Optimierungsproblemen mit Nebenbedingungen liefert, in allgemeiner Form angeben, schauen wir die obigen Bedingungen noch einmal formal aus einem anderen Blickwinkel an.

Zu dem Problem aus unserem Beispiel definieren wir eine Hilfsfunktion (Lagrangefunktion) durch

$$L(x_1,x_2;\lambda)=f(x_1,x_2)+\lambda g(x_1,x_2).$$

Die Forderung $\operatorname{grad} L(x_1,x_2;\lambda)=\vec{0}$ führt auf

$$L_{x_1}(x_1,x_2;\lambda)=f_{x_1}(x_1,x_2)+\lambda g_{x_1}(x_1,x_2)=0$$
$$L_{x_2}(x_1,x_2;\lambda)=f_{x_2}(x_1,x_2)+\lambda g_{x_2}(x_1,x_2)=0$$
$$L_{\lambda}(x_1,x_2;\lambda)=g(x_1,x_2)=0$$

Dies sind genau dieselben Gleichungen, die wir aus der zunächst anschaulichen Argumentation hergeleitet haben.
Mit Hilfe der Lagrangefunktion lässt sich nun die Lagrangesche Multiplikatormethode sehr allgemein formulieren.

Notwendige Bedingung 3.4.2: Es seien $f, g_1, g_2, \ldots, g_m : \mathbb{R}^n \to \mathbb{R}$, $m < n$, C^1-Funktionen und f besitze an einem inneren Punkt $\vec{x}^*$ von $D_f \cap D_{g_1} \cap \ldots \cap D_{g_m}$ ein lokales Extremum unter den Restriktionen $g_i(\vec{x}^*) = 0$, $i = 1, \ldots, m$.
Der Rang (Anzahl linear unabhängiger Zeilen- bzw. Spalten) der sogenannten Jacobi-Matrix zu den Funktionen $g_1, \ldots, g_m$ an der Stelle $\vec{x}^*$ (Matrix der ersten partiellen Ableitungen von $g_1, \ldots, g_m$ an der Stelle $\vec{x}^*$)

$$J(\vec{x}^*) = \begin{pmatrix} \frac{\partial g_1}{\partial x_1}(\vec{x}^*) & \cdots & \frac{\partial g_1}{\partial x_n}(\vec{x}^*) \\ \vdots & & \\ \frac{\partial g_m}{\partial x_1}(\vec{x}^*) & \cdots & \frac{\partial g_m}{\partial x_n}(\vec{x}^*) \end{pmatrix} = \begin{pmatrix} [\nabla g_1(\vec{x}^*)]^T \\ \vdots \\ [\nabla g_m(\vec{x}^*)]^T \end{pmatrix}$$

sei m. Dann existieren m Zahlen $\lambda_1, \ldots, \lambda_m$ (die sogenannten Lagrange-Multiplikatoren), so dass für $L(\vec{x}; \lambda_1, \ldots, \lambda_m) = f(\vec{x}) + \sum_{i=1}^{m} \lambda_i g_i(\vec{x})$ gilt:

$$\frac{\partial L}{\partial x_i}(\vec{x}^*) = \frac{\partial f}{\partial x_i}(\vec{x}^*) + \sum_{j=1}^{m} \lambda_j \frac{\partial g_j}{\partial x_i}(\vec{x}^*) = 0 \quad , i = 1, \ldots, n.$$

Bemerkung 3.4.3: 1) Die Nebenbedingungen lassen sich auch in der Form $L_{\lambda_j}(\vec{x}; \lambda_1, \ldots, \lambda_m) = 0$, $j = 1, \ldots, m$, angeben.

2) Praktisch bedeutet Satz 3.4.2, dass man als Kandidaten für die Lösung des Optimierungsproblems diejenigen $(\vec{x}; \lambda_1, \ldots, \lambda_m)$ bestimmt, für die gilt:

I) $\frac{\partial L}{\partial x_i}(\vec{x}) = 0 \quad , i = 1, \ldots, n$

II) $g_j(\vec{x}) = 0 \quad , j = 1, \ldots, m$

III) $Rg(J(\vec{x})) = m$

3) Die Bedingung, dass die Jacobi-Matrix maximalen Rang besitzt, garantiert, dass die Restriktionen $g_i(\vec{x}) = 0$, $i = 1, \ldots, m$, lokal nach m

der n Variablen auflösbar sind. Dabei bedeutet lokal, dass die Auflösbarkeit gewährleistet ist, wenn $\vec{x}$ in einer Umgebung von $\vec{x}^*$ liegt.

Beispiel 3.4.4: Wir betrachten das Problem

$$\min f(x_1,x_2,x_3,x_4) = x_2^2 + 2x_3^2 + x_2x_3 - x_1 + 4x_2 - x_3 - 2x_4 + 5$$

$$\text{so dass } \quad g_1(x_1,x_2,x_3,x_4) = x_1 - x_2 + x_3 - x_4 + 1 = 0$$
$$g_2(x_1,x_2,x_3,x_4) = 2x_1 - 4x_2 + 2x_3 + x_4 - 4 = 0$$

Die zugehörige Lagrangefunktion lautet:

$$L(x_1,x_2,x_3,x_4;\lambda_1,\lambda_2) = x_2^2 + 2x_3^2 + x_2x_3 - x_1 + 4x_2 - x_3 - 2x_4 + 5$$
$$+ \lambda_1(x_1 - x_2 + x_3 - x_4 + 1) + \lambda_2(2x_1 - 4x_2 + 2x_3 + x_4 - 4)$$

Die partiellen Ableitungen sind:

$$L_{x_1}(x_1,x_2,x_3,x_4;\lambda_1,\lambda_2) = -1 + \lambda_1 + 2\lambda_2$$
$$L_{x_2}(x_1,x_2,x_3,x_4;\lambda_1,\lambda_2) = 2x_2 + x_3 + 4 - \lambda_1 - 4\lambda_2$$
$$L_{x_3}(x_1,x_2,x_3,x_4;\lambda_1,\lambda_2) = 4x_3 + x_2 - 1 + \lambda_1 + 2\lambda_2$$
$$L_{x_4}(x_1,x_2,x_3,x_4;\lambda_1,\lambda_2) = -2 - \lambda_1 + \lambda_2$$
$$L_{\lambda_1}(x_1,x_2,x_3,x_4;\lambda_1,\lambda_2) = x_1 - x_2 + x_3 - x_4 + 1$$
$$L_{\lambda_2}(x_1,x_2,x_3,x_4;\lambda_1,\lambda_2) = 2x_1 - 4x_2 + 2x_3 + x_4 - 4$$

In diesem Beispiel ergibt das Nullsetzen der partiellen Ableitungen ein lineares Gleichungssystem (in Matrix-Vektor-Form)

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 2 \\ 0 & 2 & 1 & 0 & -1 & -4 \\ 0 & 1 & 4 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & -1 & 1 \\ 1 & -1 & 1 & -1 & 0 & 0 \\ 2 & -4 & 2 & 1 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ \lambda_1 \\ \lambda_2 \end{pmatrix} = \begin{pmatrix} 1 \\ -4 \\ 1 \\ 2 \\ -1 \\ 4 \end{pmatrix},$$

das sich z.B. mit dem Gauß-Algorithmus lösen lässt. Die eindeutige Lösung ist $(\vec{x}^*; \lambda_1^*, \lambda_2^*) = (-\frac{2}{21}, -\frac{4}{7}, \frac{1}{7}, \frac{34}{21}; -1, 1)$. Wir haben somit einen möglichen Kandidaten für ein lokales Minimum bestimmt.

Für die Jacobi-Matrix von g_1, g_2 gilt:

$$J(\vec{x}) = \begin{pmatrix} 1 & -1 & 1 & -1 \\ 2 & -4 & 2 & 1 \end{pmatrix} = J(\vec{x}^*).$$

Offensichtlich ist $Rg(J(x_1^*, x_2^*, x_3^*, x_4^*)) = 2$, da z.B. der 1. und 4. Spaltenvektor linear unabhängig sind. Die Rangbedingung ist also erfüllt.

Ergänzung 3.4.5: Gauß-Algorithmus (schematisch)

1	-1	1	-1	0	0	-1	①
2	-4	2	1	0	0	4	(-2)
0	2	1	0	-1	-4	-4	
0	1	4	0	1	2	1	
0	0	0	0	1	2	1	
0	0	0	0	-1	1	2	(1)
	-2	0	3	0	0	6	②
	2	1	0	-1	-4	-4	(1)
	1	4	0	1	2	1	($\frac{1}{2}$)
				1	2	1	⑤
					3	3	⑥
		1	3	-1	-4	2	③
		4	$\frac{3}{2}$	1	2	4	(-4)
			$-\frac{21}{2}$	5	18	-4	④

Rückwärtsauflösen

aus ⑥: $3\lambda_2 = 3 \Leftrightarrow \lambda_2 = 1$

aus ⑤: $\lambda_1 = 1 - 2\lambda_2$

mit $\lambda_2 = 1$: $\lambda_1 = -1$

aus ④: $-\frac{21}{2} x_4 = -4 - 5\lambda_1 - 18\lambda_2$

mit $\lambda_2 = 1$, $\lambda_1 = -1$: $x_4 = \frac{34}{21}$

aus ③: $x_3 = 2 - 3x_4 + \lambda_1 + 4\lambda_2$

mit $\lambda_2 = 1, \lambda_1 = -1, x_4 = \frac{34}{21}$:

$x_3 = \frac{1}{7}$

aus ②: $2x_2 = -6 + 3x_4$

mit $x_4 = \frac{34}{21}$: $x_2 = -\frac{4}{7}$

aus ①: $x_1 = -1 + x_2 - x_3 + 4$

mit $x_2 = -\frac{4}{7}, x_3 = \frac{1}{7}, x_4 = \frac{34}{21}$:

$x_1 = -\frac{2}{21}$

Um nun entscheiden zu können, ob stationäre Punkte der Lagrangefunktion tatsächlich Minimierer bzw. Maximierer des Optimierungsproblems sind, benötigen wir hinreichende Bedingungen. Dabei spielt die Konvexität der Lagrangefunktion eine Rolle. Dies drücken wir mit Hilfe von Definitheitseigenschaften der Hesse-Matrix von L aus.

Zunächst stellen wir einen Zusammenhang zwischen den Minima der Lagrangefunktion und der Zielfunktion auf dem zulässigen Bereich her. Sei nun $\vec{x}^*$ ein stationärer Punkt der Lagrangefunktion mit mit zugehörigen λ_i^*, $i=1,\dots,m$. Dann bezeichnen wir mit $L^*(\vec{x})$ die reduzierte Lagrangefunktion, die sich durch Einsetzen der festen Werte λ_i^* für λ_i, $i=1,\dots,m$, in die Lagrangefunktion ergibt.

Dann gilt:

$\vec{x}^*$ (lokales) Minimum von $L^* \Leftrightarrow \vec{x}^*$ (lokales) Minimum von f auf der Menge der zulässigen Punkte

denn:

$$L^*(\vec{x}^*) \le L^*(\vec{x}) \Leftrightarrow f(\vec{x}^*) + \sum_{i=1}^{m} \lambda_i^* \underbrace{g_i(\vec{x}^*)}_{=0} \le f(\vec{x}) + \sum_{i=1}^{m} \lambda_i^* \underbrace{g_i(\vec{x})}_{=0}$$

da $\vec{x}^*$ stationärer Punkt von L

$$\Leftrightarrow f(\vec{x}^*) \le f(\vec{x})$$

Damit lässt sich nun folgendes hinreichende Kriterium formulieren:

Satz 3.4.6: Seien f, g_i, $i=1,\dots,m$, C^2-Funktionen und $(\vec{x}^*, \lambda_1^*,\dots,\lambda_m^*)$ ein stationärer Punkt von $L(\vec{x};\lambda_1,\dots,\lambda_m) = f(\vec{x}) + \sum_{i=1}^{m} \lambda_i g_i(\vec{x})$. Dann gilt: Ist die Hesse-Matrix von L^* an der Stelle $\vec{x}^*$ positiv definit, dann besitzt f an der Stelle $\vec{x}^*$ auf der Menge der zulässigen Punkte ein striktes lokales Minimum (bei negativer Definitheit striktes lokales Maximum.

Ist L^* für alle $\vec{x}$ konvex (konkav), so handelt es sich um ein globales Minimum (Maximum).

Beispiel 3.4.7: $\min f(x_1,x_2,x_3) = x^2+y^2+z^2$

so dass $x+2y+z-30=0$

$2x-y-3z-10=0$

Die zugehörige Lagrangefunktion ist

$$L(x_1,x_2,x_3;\lambda_1,\lambda_2) = x^2+y^2+z^2+\lambda_1(x+2y+z-30)+\lambda_2(2x-y-3z-10)$$

und es gilt:

$$\operatorname{grad} L(\vec{x};\lambda_1,\lambda_2)=\begin{pmatrix} 2x+\lambda_1+2\lambda_2 \\ 2y+2\lambda_1-\lambda_2 \\ 2z+\lambda_1-3\lambda_2 \\ x+2y+z-30 \\ 2x-y-3z-10 \end{pmatrix}.$$

Wir bestimmen die stationären Punkte der Lagrangefunktion.

$$\operatorname{grad} L(\vec{x};\lambda_1,\lambda_2)=\vec{0} \Longleftrightarrow \left\{\begin{array}{l} 2x+\lambda_1+2\lambda_2=0 \\ 2y+2\lambda_1-\lambda_2=0 \\ 2z+\lambda_1-3\lambda_2=0 \\ x+2y+z=30 \\ 2x-y-3z=10 \end{array}\right\}$$

Die Lösung dieses linearen Gleichungssystems ist eindeutig bestimmt.
Sie lautet: $x^*=10, y^*=10, z^*=0; \lambda_1^*=-12, \lambda_2^*=-4.$
Wir untersuchen nun die Definitheitseigenschaften der Hesse-Matrix von

$$L^*(x,y,z)=x^2+y^2+z^2+12(x+2y+z-30)+4(2x-y-3z-10)$$
$$=x^2+y^2+z^2+20x+20y-400.$$

Für x,y,z beliebig ist $H_{L^*}(x,y,z)=\begin{pmatrix} 2&0&0\\0&2&0\\0&0&2\end{pmatrix}$ positiv definit.

Damit ist $(x^*,y^*,z^*)=(10,10,0)$ globales Minimum von f unter den angegebenen Restriktionen mit $f(10,10,0)=200$.

Man beachte, dass es sich in Satz 3.4.6 um hinreichende Bedingungen handelt. Wenn diese nicht erfüllt sind, lassen sich mit diesem Satz keine Aussagen machen. Es gibt hinreichende Kriterien, die schwächer sind, die wir im Rahmen dieser Vorlesung aber nicht weiter behandeln.
In praktischen Anwendungen kann man häufig durch genauere Überlegungen weitere Aussagen machen.

Beispiel 3.4.8: Ein Verbraucher bestimmt seinen Nutzen mit einer Cobb-Douglas Nutzenfunktion $U(x_1,x_2,x_3)=x_1^2x_2^3x_3$ mit $x_1,x_2,x_3\geq 0$. Er unterliegt der Budgetbeschränkung $x_1+x_2+x_3=12$. Wie soll er x_1,x_2,x_3 wählen?

Das zugehörige Optimierungsproblem lautet:

$$\max\ U(x_1, x_2, x_3) = x_1^2\, x_2^3\, x_3$$

so dass $g(x_1, x_2, x_3) = 12 - x_1 - x_2 - x_3 = 0$

Lagrangefunktion:

$$L(x_1, x_2, x_3; \lambda) = x_1^2 x_2^3 x_3 + \lambda(12 - x_1 - x_2 - x_3)$$

Die partiellen Ableitungen der Lagrange-Funktion lauten:

$$L_{x_1}(x_1, x_2, x_3; \lambda) = 2x_1 x_2^3 x_3 - \lambda$$

$$L_{x_2}(x_1, x_2, x_3; \lambda) = 3x_1^2 x_2^2 x_3 - \lambda$$

$$L_{x_3}(x_1, x_2, x_3; \lambda) = x_1^2 x_2^3 - \lambda$$

$$L_{\lambda}(x_1, x_2, x_3; \lambda) = 12 - x_1 - x_2 - x_3$$

Wir erhalten somit die Bedingungen:

1) $2x_1 x_2^3 x_3 = \lambda$
2) $3x_1^2 x_2^2 x_3 = \lambda$
3) $x_1^2 x_2^3 \quad = \lambda$
4) $x_1 + x_2 + x_3 = 12$

Wir können in diesem Beispiel ohne Einschränkung voraussetzen, dass x_1, x_2, x_3 alle von Null verschieden ist, da sonst der Nutzen gleich Null und somit sicher nicht maximal ist.

Gleichsetzen von 1) und 2) liefert: $2x_1 x_2^3 x_3 = 3x_1^2 x_2^2 x_3 \quad | : 2x_1 x_2^2 x_3 \neq 0$

I) $x_2 = \frac{3}{2} x_1$

Gleichsetzen von 1) und 3) liefert: $2x_1 x_2^3 x_3 = x_1^2 x_2^3 \quad | : 2x_1 x_2^3 \neq 0$

II) $x_3 = \frac{1}{2} x_1$

Einsetzen von I) und II) in 4) liefert: $x_1 + \frac{3}{2}x_1 + \frac{1}{2}x_1 = 12 \iff \underline{x_1 = 4}$

Einsetzen von $x_1 = 4$ in I) und II) liefert: $\underline{x_2 = 6}$, $\underline{x_3 = 2}$

Einsetzen von $x_1 = 4$, $x_2 = 6$, $x_3 = 2$ in 1), 2) bzw. 3) liefert: $\underline{\lambda = 3456}$

Für die Jacobi-Matrix gilt: $J(x_1, x_2, x_3) = (1, 1, 1) = J(x_1^*, x_2^*, x_3^*)$, also ist die Rangbedingung $Rg(J(x_1^*, x_2^*, x_3^*)) = 1$ erfüllt.

Die einzige mögliche (lokale) Extremalstelle von f unter der Budgetbeschränkung ist somit $(x_1^*, x_2^*, x_3^*) = (4, 6, 2)$.
Da der durch die Budgetbeschränkung beschriebene Bereich abgeschlossen und beschränkt ist, besitzt f dort ein absolutes Maximum. Da f auf dem Rand der zulässigen Menge 0 ist, ist an der Stelle $(4,6,2)$ der Nutzen unter der Budgetbeschränkung maximal mit $U(4,6,2) = 4^2 \cdot 6^3 \cdot 2 = 6912$.

Wir werden nun das letzte Beispiel etwas allgemeiner betrachten, um damit einige weitere Begriffe und Zusammenhänge erläutern zu können.

Beispiel 3.4.9: Die Aufgabenstellung sei wie im letzten Beispiel. Für die Budgetbeschränkung verwenden wir allerdings allgemeiner die Gleichung

$$x_1 + x_2 + x_3 = m \quad \text{mit dem Einkommen } m > 0.$$

Das zugehörige Optimierungsproblem lautet dann

$$\max\ U(x_1, x_2, x_3) = x_1^2 x_2^3 x_3$$

$$\text{so dass} \quad g(x_1, x_2, x_3) = m - x_1 - x_2 - x_3 = 0$$

Es ist zu erwarten, dass die Optimallösung vom Einkommen m abhängig ist, was durch die folgende Rechnung bestätigt wird.
Lagrangefunktion:

$$L(x_1, x_2, x_3; \lambda) = x_1^2 x_2^3 x_3 + \lambda(m - x_1 - x_2 - x_3)$$

Die partiellen Ableitungen der Lagrangefunktion sind:

$$L_{x_1}(x_1, x_2, x_3; \lambda) = 2x_1 x_2^3 x_3 - \lambda$$

$$L_{x_2}(x_1, x_2, x_3; \lambda) = 3x_1^2 x_2^2 x_3 - \lambda$$

$$L_{x_3}(x_1, x_2, x_3; \lambda) = x_1^2 x_2^3 - \lambda$$

$$L_{\lambda}(x_1, x_2, x_3; \lambda) = m - x_1 - x_2 - x_3$$

Wir erhalten somit die Bedingungen:
1) $2x_1 x_2^3 x_3 = \lambda$
2) $3x_1^2 x_2^2 x_3 = \lambda$
3) $x_1^2 x_2^3 = \lambda$
4) $x_1 + x_2 + x_3 = m$

Das Lösen des nichtlinearen Gleichungssystems geschieht völlig analog zum letzten Beispiel. Es ergibt sich die von m abhängige Lösung:

$$x_1^*(m) = \frac{1}{3}m \quad , x_2^*(m) = \frac{1}{2}m \; , x_3^*(m) = \frac{1}{6}m \; , \lambda^*(m) = \frac{1}{72}m^5$$

Setzt man dies nun in die Nutzenfunktion ein, so erhält man die von m abhängige Funktion

$$\tilde{u}(m) = \frac{1}{432} \cdot m^6 = u(x_1^*(m), x_2^*(m), x_3^*(m))$$

Die Funktion $\tilde{u}$ heißt Optimalwertfunktion des Problems. Es gilt:

$$\frac{d}{dm}\,\tilde{u}(m) = \frac{1}{72}\,m^5 = \lambda^*(m)$$

Damit gibt der Lagrange-Multiplikator $\lambda^*(m)$ die momentane Rate an, mit der sich der optimale Wert der Nutzenfunktion ändert, wenn sich das Einkommen m ändert.

Die lineare Approximation von $\tilde{u}$ an einer Stelle $\tilde{m}$ ist (vgl. Analysis I):

$$t_1(m) = \tilde{u}(\tilde{m}) + \frac{d}{dm}\,\tilde{u}(\tilde{m}) \cdot (m - \tilde{m})$$

$$= \tilde{u}(\tilde{m}) + \lambda^*(\tilde{m}) \cdot (m - \tilde{m}) \approx \tilde{u}(m),$$

wenn m nahe bei $\tilde{m}$ ist. Anders ausgedrückt ist dann

$$\tilde{u}(m) - \tilde{u}(\tilde{m}) \approx \lambda^*(\tilde{m}) \cdot (m - \tilde{m}).$$

Für $m = \tilde{m} + 1$ bedeutet dies speziell:

$$\tilde{u}(\tilde{m}+1) - \tilde{u}(\tilde{m}) \approx \lambda^*(\tilde{m}).$$

$\lambda^*(\tilde{m})$ gibt also näherungsweise an, um wie viel sich der Nutzen ändert, wenn $\tilde{m}$ um 1 erhöht wird.

In der Ökonomie heißt λ^* auch Schattenpreis.

Wir betrachten die Zusammenhänge nun an einem allgemeinen Optimierungsproblem der Form:

$$\min\,(\max)\; f(x_1, x_2, \dots, x_n)$$

$$\text{so dass} \quad g_i(x_1, x_2, \dots, x_n) = 0 \;, \; i = 1, 2, \dots, m$$

Für die Restriktionen g_i schreiben wir:

$$g_i(x_1, x_2, \ldots, x_n) = c_i - h_i(x_1, x_2, \ldots, x_m) = 0, \quad i = 1, 2, \ldots, m$$

mit Konstanten c_i.

Außerdem setzen wir voraus, dass alle Funktionen hinreichend oft differenzierbar sind.

Lösungen der notwendigen Bedingungen, die sich aus dem Nullsetzen der partiellen Ableitungen der Lagrangefunktion ergeben, werden nun im Allgemeinen von $c_1, c_2, \ldots, c_m$ abhängen.

Ist $x_j^*(c_1, c_2, \ldots, c_m)$, $j = 1, \ldots, n$, $\lambda_i^*(c_1, c_2, \ldots, c_m)$, $i = 1, \ldots, m$, eine solche Lösung, so erhält man durch Einsetzen in die Zielfunktion f eine Funktion $\tilde{f}$, die von $c_1, c_2, \ldots, c_m$ abhängt:

$$\tilde{f}(c_1, c_2, \ldots, c_m) = f(x_1^*(c_1, c_2, \ldots, c_m), \ldots, x_n^*(c_1, c_2, \ldots, c_m))$$

Die so entstehende Funktion heißt Optimalwertfunktion.

Der optimale Wert der Zielfunktion ändert sich in Abhängigkeit von der Wahl der Konstanten $c_1, c_2, \ldots, c_m$ in den Restriktionen.

Es gilt nun allgemein folgender Zusammenhang:

$$\frac{\partial}{\partial c_\ell} \tilde{f}(c_1, c_2, \ldots, c_m) = \lambda_\ell^*(c_1, c_2, \ldots, c_m)$$

Für interessierte Studierende wird dies im Folgenden nachgewiesen.
Für den Beweis benötigt man insbesondere die allgemeine Kettenregel!

Aus dem Nullsetzen der partiellen Ableitungen der Lagrangefunktion

$$L(x_1, x_2, \ldots, x_n; \lambda_1, \lambda_2, \ldots, \lambda_m) = f(x_1, x_2, \ldots, x_n) + \sum_{i=1}^{m} \lambda_i [c_i - h_i(x_1, x_2, \ldots, x_n)]$$

nach der Variablen x_j, $j = 1, 2, \ldots, n$, folgt zunächst

$$\frac{\partial}{\partial x_j} f(x_1^*, x_2^*, \ldots, x_n^*) = \sum_{i=1}^{m} \lambda_i^* \cdot \frac{\partial h_i}{\partial x_j}(x_1^*, x_2^*, \ldots, x_n^*) \qquad (*_1)$$

Für die partiellen Ableitungen von $f(x_1^*(c_1, \ldots, c_m), \ldots, x_n^*(c_1, \ldots, c_m))$ nach c_ℓ gilt mit Hilfe der Kettenregel:

$$\frac{\partial}{\partial c_\ell}\tilde{f}(c_1,\dots,c_m)=\frac{\partial}{\partial c_\ell}f(x_1^*(c_1,\dots,c_m),\dots,x_n^*(c_1,\dots,c_m))$$

$$=\sum_{j=1}^{n}\frac{\partial}{\partial x_j}f(x_1^*,\dots,x_n^*)\cdot\frac{\partial}{\partial c_\ell}x_j^*(c_1,c_2,\dots,c_m)$$

Einsetzen von $(*_1)$

$$=\sum_{j=1}^{n}\sum_{i=1}^{m}\lambda_i^*\frac{\partial}{\partial x_j}h(x_1^*,\dots,x_n^*)\cdot\frac{\partial}{\partial c_\ell}x_j^*(c_1,\dots,c_m)$$

$$=\sum_{i=1}^{m}\lambda_i^*\sum_{j=1}^{n}\frac{\partial}{\partial x_j}h(x_1^*,\dots,x_n^*)\cdot\frac{\partial}{\partial c_\ell}x_j^*(c_1,\dots,c_m)\quad(*_2)$$

Differenziert man die durch die Restriktionen gegebenen Gleichungen $h_i(x_1^*,\dots,x_n^*)=c_i$ auf beiden Seiten partiell nach c_ℓ, so erhält man

$$\frac{\partial}{\partial c_\ell}h_i(x_1^*,\dots,x_n^*)=\sum_{j=1}^{n}\frac{\partial}{\partial x_j}h_i(x_1^*,\dots,x_n^*)\cdot\frac{\partial}{\partial c_\ell}x_j^*(c_1,\dots,c_m)$$

$$=\begin{cases}1, & \text{falls } i=\ell\\ 0, & \text{falls } i\neq\ell\end{cases}$$

Setzt man dies nun in die Gleichung $(*_2)$ ein, so erhält man

$$\frac{\partial}{\partial c_\ell}\tilde{f}(c_1,\dots,c_m)=\lambda_\ell^*(c_1,\dots,c_m)\text{ , d.h. die Behauptung.}$$

Die Formel $\frac{\partial}{\partial c_\ell}\tilde{f}(c_1,\dots,c_m)=\lambda_\ell^*(c_1,\dots,c_m)$ zeigt:

Der Lagrange-Multiplikator $\lambda_i^*(c_1,\dots,c_m)$ für die i-te Nebenbedingung ist die momentane Rate, mit der sich der Optimalwert der Zielfunktion ändert, wenn sich die Konstante c_i in der i-ten Restriktion ändert.

Wir betrachten wieder die lineare Approximation von $\tilde{f}$ um $(\tilde{c}_1,\dots,\tilde{c}_m)$, d.h.

$$t(c_1,\dots,c_m)=\tilde{f}(\tilde{c}_1,\dots,\tilde{c}_m)+\left\langle \operatorname{grad}\tilde{f}(\tilde{c}_1,\dots,\tilde{c}_m),\begin{pmatrix}c_1-\tilde{c}_1\\ \vdots\\ c_m-\tilde{c}_m\end{pmatrix}\right\rangle$$

$$=\tilde{f}(\tilde{c}_1,\dots,\tilde{c}_m)+\sum_{i=1}^{m}\frac{\partial}{\partial c_i}\tilde{f}(\tilde{c}_1,\dots,\tilde{c}_m)\cdot(c_i-\tilde{c}_i)$$

$$=\tilde{f}(\tilde{c}_1,\dots,\tilde{c}_m)+\sum_{i=1}^{m}\lambda_i^*(\tilde{c}_1,\dots,\tilde{c}_m)\cdot(c_i-\tilde{c}_i)$$

$$\approx\tilde{f}(c_1,\dots,c_m),$$

wenn $(c_1,\dots,c_m)$ nahe bei $(\tilde{c}_1,\dots,\tilde{c}_m)$ ist. Dann ist also

$$\tilde{f}(c_1,\dots,c_m)-\tilde{f}(\tilde{c}_1,\dots,\tilde{c}_m)\approx\sum_{i=1}^{m}\lambda_i^*(\tilde{c}_1,\dots,\tilde{c}_m)\cdot(c_i-\tilde{c}_i)$$

Für $c_\ell = \tilde{c}_\ell + 1$, $c_i = \tilde{c}_i$ für $i \neq \ell$ ergibt sich daraus speziell:

$$\tilde{f}(\tilde{c}_1, \dots, \tilde{c}_{\ell-1}, \tilde{c}_\ell + 1, \tilde{c}_{\ell+1}, \dots, \tilde{c}_m) - \tilde{f}(\tilde{c}_1, \dots, \tilde{c}_m) \approx \lambda_\ell^*(\tilde{c}_1, \dots, \tilde{c}_m) \quad (*)$$

$\lambda_\ell^*(\tilde{c}_1, \dots, \tilde{c}_m)$ gibt also an, um wie viel sich der Optimalwert ungefähr ändert, wenn $\tilde{c}_\ell$ um 1 Einheit erhöht wird.

In der Ökonomie heißt λ_ℓ^* auch Schattenpreis, der einer Einheit der Ressource ℓ zugeschrieben wird.

Beispiel 3.4.10: Wir betrachten das Optimierungsproblem

$$\min f(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2$$

so dass

$$g_1(x_1, x_2, x_3) = c_1 - x_1 - 2x_2 - x_3 = 0$$

$$g_2(x_1, x_2, x_3) = c_2 - 2x_1 + x_2 + 3x_3 = 0$$

Lagrangefunktion:

$$L(x_1, x_2, x_3; \lambda_1, \lambda_2) = x_1^2 + x_2^2 + x_3^2 + \lambda_1(c_1 - x_1 - 2x_2 - x_3) + \lambda_2(c_2 - 2x_1 + x_2 + 3x_3)$$

Partielle Ableitungen:

$$L_{x_1}(x_1, x_2, x_3; \lambda_1, \lambda_2) = 2x_1 - \lambda_1 - 2\lambda_2$$

$$L_{x_2}(x_1, x_2, x_3; \lambda_1, \lambda_2) = 2x_2 - 2\lambda_1 + \lambda_2$$

$$L_{x_3}(x_1, x_2, x_3; \lambda_1, \lambda_2) = 2x_3 - \lambda_1 + 3\lambda_2$$

$$L_{\lambda_1}(x_1, x_2, x_3; \lambda_1, \lambda_2) = c_1 - x_1 - 2x_2 - x_3$$

$$L_{\lambda_2}(x_1, x_2, x_3; \lambda_1, \lambda_2) = c_2 - 2x_1 + x_2 + 3x_3$$

Nullsetzen der partiellen Ableitungen liefert:

1) $2x_1 - \lambda_1 - 2\lambda_2 = 0$

2) $2x_2 - 2\lambda_1 + \lambda_2 = 0$

3) $2x_3 - \lambda_1 + 3\lambda_2 = 0$

4) $x_1 + 2x_2 + x_3 = c_1$

5) $2x_1 - x_2 - 3x_3 = c_2$

Dies ist ein lineares Gleichungssystem, das sich z.B. mit dem Gauß-Algorithmus lösen lässt.

$$\begin{array}{rrrrr|l}
2 & 0 & 0 & -1 & -2 & 0 \\
0 & 2 & 0 & -2 & 1 & 0 \\
0 & 0 & 2 & -1 & 3 & 0 \\
1 & 2 & 1 & 0 & 0 & c_1 \\
2 & -1 & -3 & 0 & 0 & c_2
\end{array} \quad (-\tfrac{1}{2}) \quad (-1)$$

$$\begin{array}{rrrr|l}
2 & 0 & -2 & 1 & 0 \\
0 & 2 & -1 & 3 & 0 \\
2 & 1 & \tfrac{1}{2} & 1 & c_1 \\
-1 & -3 & 1 & 2 & c_2
\end{array} \quad (-1) \quad (\tfrac{1}{2})$$

$$\begin{array}{rrr|l}
2 & -1 & 3 & 0 \\
1 & \tfrac{5}{2} & 0 & c_1 \\
-3 & 0 & \tfrac{5}{2} & c_2
\end{array} \quad (-\tfrac{1}{2}) \quad (\tfrac{3}{2})$$

$$\begin{array}{rr|l}
3 & -\tfrac{3}{2} & c_1 \\
-\tfrac{3}{2} & 7 & c_2
\end{array} \quad (\tfrac{1}{2})$$

$$\begin{array}{r|l}
\tfrac{25}{4} & \tfrac{1}{2} c_1 + c_2
\end{array}$$

Rückwärtsauflösen liefert:

$$\lambda_2^* = \frac{2}{25} c_1 + \frac{4}{25} c_2$$

$$\lambda_1^* = \frac{28}{75} c_1 + \frac{2}{25} c_2$$

$$x_3^* = \frac{1}{15} c_1 - \frac{1}{5} c_2$$

$$x_2^* = \frac{1}{3} c_1$$

$$x_1^* = \frac{4}{15} c_1 + \frac{1}{5} c_2$$

Die Optimalwertfunktion ergibt sich durch Einsetzen von x_1^*, x_2^*, x_3^* in f zu:

$$\tilde{f}(c_1, c_2) = \frac{14}{75} c_1^2 + \frac{2}{25} c_1 c_2 + \frac{2}{25} c_2^2 .$$

Man bestätigt leicht den oben allgemein aufgestellten Zusammenhang:

$$\frac{\partial}{\partial c_1} \tilde{f}(c_1, c_2) = \frac{28}{75} c_1 + \frac{2}{25} c_2 = \lambda_1^*(c_1, c_2)$$

$$\frac{\partial}{\partial c_2} \tilde{f}(c_1, c_2) = \frac{2}{25} c_1 + \frac{4}{25} c_2 = \lambda_2^*(c_1, c_2)$$

Wählen wir z.B. $c_1 = 30$, $c_2 = 10$, so erhalten wir

$$\tilde{f}(30, 10) = 200, \quad \lambda_1^*(30, 10) = 12, \quad \lambda_2^*(30, 10) = 4.$$

Mit (*) auf S. 18 erhält man z.B.:

$$\tilde{f}(31, 9) - \tilde{f}(30, 10) \approx \lambda_1^*(30, 10) \cdot (31 - 30) + \lambda_2^*(30, 10) \cdot (9 - 10)$$
$$= 8$$

Zum Vergleich die exakten Werte:

$$\tilde{f}(31, 9) - \tilde{f}(30, 10) = \frac{15614}{75} - 200 = \frac{614}{75} = 8.18\overline{6}$$

III.5 Optimierung mit Ungleichungsrestriktionen

In diesem Abschnitt betrachten wir folgendes Optimierungsproblem:

$$\min f(\vec{x})$$
$$\text{so dass } h_j(\vec{x}) \leq 0 \quad , j = 1, \dots, m$$

Hier haben wir nun zu berücksichtigen, dass die Restriktionen in Form von Ungleichungen gegeben sind. Wie im Abschnitt vorher definieren wir wieder die zum Problem gehörende Lagrangefunktion durch

$$L(\vec{x}; \mu_1, \dots, \mu_m) = f(\vec{x}) + \sum_{i=1}^{m} \mu_i h_i(\vec{x}).$$

Um an Kandidaten für die Lösung des Optimierungsproblems zu gelangen, fordern wir hier wieder

$$(1) \quad \frac{\partial L}{\partial x_i}(\vec{x}) = 0 \quad , i = 1, \dots, n$$

und führen die sogenannte komplementäre Schlupfbedingung ein, d.h.

$$(2) \quad \mu_j \geq 0 \text{ mit } \mu_j = 0 \text{ falls } h_j(\vec{x}) < 0, \; j = 1, \dots, m.$$

Schließlich müssen auch die Ungleichungsrestriktionen erfüllt sein.

Die Bedingungen (1) und (2) heißen Karush-Kuhn-Tucker Bedingungen, kurz KKT-Bedingungen.

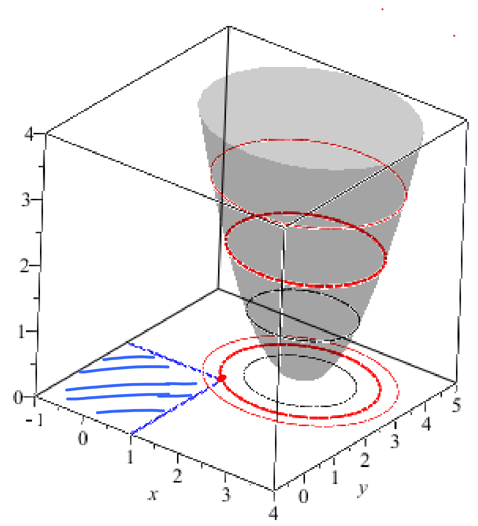
Die Bedingung (2) bedeutet, dass in den beiden Ungleichungen $\mu_j \geq 0$, $h_j(\vec{x}) \leq 0$ höchstens einmal ein striktes "<" bzw. ">" auftreten darf.

Ist $h_j(\tilde{\vec{x}}) = 0$ an einer Stelle $\tilde{\vec{x}}$, so sagt man, dass die Restriktion aktiv ist. Bevor wir zu einer präziseren Formulierung notwendiger Bedingungen kommen, erläutern wir die Begriffe an einem anschaulichen Beispiel.

Beispiel 3.5.1: Wir betrachten das Problem

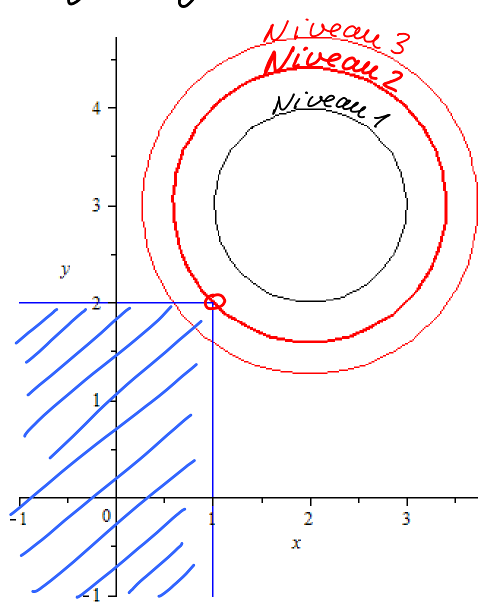
$$\min f(x,y) = (x-2)^2 + (y-3)^2$$
$$\text{so dass } \quad h_1(x,y) = x - 1 \leq 0$$
$$h_2(x,y) = y - 2 \leq 0$$

Anschaulich stellt sich die Situation folgendermaßen dar.



Die Zielfunktion stellt ein Paraboloid mit Minimum im Punkt $(2,3,0)$ dar. Der durch die Restriktionen festgelegte zulässige Bereich in der xy-Ebene ist blau schraffiert. Die zu der schwarzen Niveaulinie gehörenden (x,y)-Werte liegen alle außerhalb des zulässigen Bereichs. Erst bei der dicker gezeichneten Niveaulinie gibt es einen Wert (x,y) (nämlich $(1,2)$), der im zulässigen Bereich liegt. Niveaulinien mit höherem Niveau haben dann mehr Punkte im zulässigen Bereich, aber eben auch mit größerem zugehörigen Funktionswert.

Die nebenstehende Graphik verdeutlich die Situation noch einmal in der xy-Ebene. Die zu den Niveaulinien gehörenden Funktionswerte nehmen mit größer werdenden Kreisen zu. Somit ist anschaulich klar, dass $(x^*,y^*)=(1,2)$ das Optimierungsproblem löst.



Im Folgenden überlegen wir nun, was uns in diesem Beispiel die KKT-Bedingungen liefern.

Die Lagrangefunktion ist gegeben durch

$L(x,y;\mu_1,\mu_2)=(x-2)^2+(y-3)^2+\mu_1(x-1)+\mu_2(y-2)$, d.h. die KKT-Bedingungen sind

(1) a) $L_x(x,y;\mu_1,\mu_2)=2(x-2)+\mu_1=0$ b) $L_y(x,y;\mu_1,\mu_2)=2(y-3)+\mu_2=0$

(2) a) $\mu_1 \geq 0$ mit $\mu_1=0$ falls $x-1<0$ b) $\mu_2 \geq 0$ mit $\mu_2=0$ falls $y-2<0$

1. Fall: Beide Restriktionen aktiv, d.h. $x=1$, $y=2$.
Dann liefert (1) a) $\mu_1 = 2$ und (1) b) $\mu_2 = 2$.
Damit ist $(x_1^*, y_1^*) = (1,2)$ möglicher Kandidat.

2. Fall: 1. Restriktion aktiv, 2. Restriktion inaktiv, d.h. $x = 1$ und $y < 2$ mit $\mu_2 = 0$.
Dann liefert (1) b) $y = 3$ im Wiederspruch zu $y < 2$.
Kein möglicher Kandidat.

3. Fall: 1. Restriktion inaktiv, 2. Restriktion aktiv, d.h. $x < 1$ mit $\mu_1 = 0$ und $y = 2$.
Dann liefert (1) a) $x = 2$ im Wiederspruch zu $x < 1$.
Kein möglicher Kandidat.

4. Fall: Beide Restriktionen inaktiv, d.h. $x < 1$ mit $\mu_1 = 0$ und $y < 2$ mit $\mu_2 = 0$.
Dann liefern (1) a) und b) $x = 2$ und $y = 3$ im Wiederspruch zu $x < 1$ und $y < 2$.
Kein möglicher Kandidat.

Insgesamt erfüllt also nur $(x_1^*, y_1^*) = (1,2)$ die KKT-Bedingungen.

Wir formulieren nun eine entsprechende notwendige Bedingung. An Stelle der Rangbedingung für die Jacobi-Matrix der Restriktionsfunktionen tritt nun eine entsprechende Bedingung für die aktiven Restriktionen.

Notwendige Bedingung 3.5.2: Es seien $f, h_1, \dots, h_m : \mathbb{R}^n \longrightarrow \mathbb{R}$, C^1-Funktionen und f besitze an einem inneren Punkt $\vec{x}^*$ von $\mathbb{D}_f \cap \mathbb{D}_{h_1} \cap \dots \cap \mathbb{D}_{h_m}$ ein lokales Minimum unter den Restriktionen $h_j(\vec{x}^*) \leq 0$, $j = 1, \dots, m$.
Der Rang der Matrix, die aus den Gradeënten der in $\vec{x}^*$ aktiven Restriktionen an der Stelle $\vec{x}^*$ gebildet wird, sei maximal, d.h. die Gradeënten der an der Stelle $\vec{x}^*$ aktiven Restriktionen seien linear unabhängig.
Dann existeren Zahlen $\mu_1, \dots, \mu_m$, so dass für die Lagrangefunktion $L(\vec{x}; \mu_1, \dots, \mu_m) = f(\vec{x}) + \sum_{i=1}^{m} \mu_i h_i(\vec{x})$ gilt:

(1) $\frac{\partial L}{\partial x_i}(\vec{x}^*; \mu_1, \dots, \mu_m) = 0$, $i = 1, \dots, n$

(2) $\mu_j \geq 0$ mit $\mu_j = 0$, falls $h_j(\vec{x}^*) < 0$, $j = 1, \dots, m$.

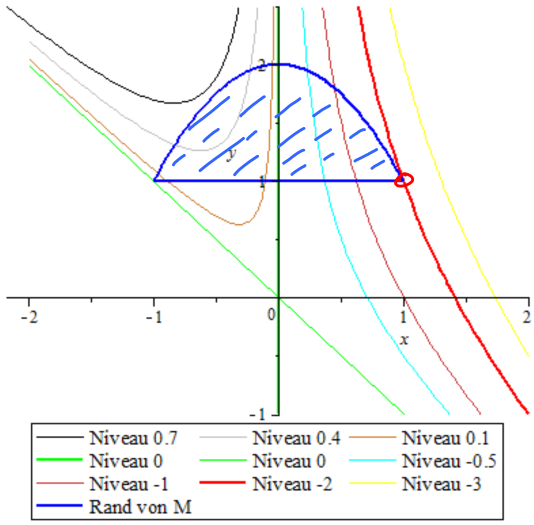
Die in 3.5.2 genannte Rangbedingung wird als constraint qualification (CF) bezeichnet. Wenn die CF-Bedingung an einem optimalen Punkt nicht erfüllt ist, kann es sein, dass dort die KKT-Bedingungen nicht erfüllt sind. Um alle möglichen Kandidaten zu ermitteln, müssen daher genau genommen zwei Dinge getan werden:

- Bestimmung aller zulässigen Punkte, die die KKT-Bedingungen erfüllen,
- Bestimmung aller zulässigen Punkte, an denen die CF-Bedingung nicht erfüllt ist.

Beispiel 3.5.3: $\min f(x,y) = -xy - x^2$

so dass $h_1(x,y) = x^2 + y - 2 \leq 0$

$h_2(x,y) = -y + 1 \leq 0$



In der nebenstehenden Graphik ist der zulässige Bereich blau umrandet und schraffiert. Zusätzlich sind einige Niveaulinien von f eingezeichnet. Man erkennt, dass das Minimum angenommen wird, wo die rote Niveaulinie den zulässigen Bereich berührt.

Die Lagrangefunktion des Problems ist

$L(x,y;\mu_1,\mu_2) = -xy - x^2 + \mu_1(x^2+y-2) + \mu_2(-y+1)$, d.h. die KKT-Bedingungen lauten:

(1) a) $L_x(x,y;\mu_1,\mu_2) = -y - 2x + 2\mu_1 x = 0$ b) $L_y(x,y;\mu_1,\mu_2) = -x + \mu_1 - \mu_2 = 0$

(2) a) $\mu_1 \geq 0$ mit $\mu_1 = 0$ falls $x^2+y-2 < 0$ b) $\mu_2 \geq 0$ mit $\mu_2 = 0$ falls $-y+1 < 0$

Die Gradienten der Restriktionen sind $\nabla h_1(x,y) = (2x, 1)$, $\nabla h_2(x,y) = (0,-1)$

1. Fall: Beide Restriktionen sind aktiv, d.h. $y = 1$ und $x^2 + y = 2$, also $x = \pm 1$.

Für $x_1 = y_1 = 1$ liefern (1)a) und b) $\mu_1 = \frac{3}{2}$, $\mu_2 = \frac{1}{2}$. Weiter gilt, dass $\nabla h_1(x_1, y_1) = (2,1)$, $\nabla h_2(x_1,y_1) = (0,-1)$ linear unabhängig sind, d.h. die

CF-Bedingung ist erfüllt.

Für $x_2=-1, y_2=1$ liefern (1)a) und b) $\mu_1=\frac{1}{2}, \mu_2=\frac{3}{2}$. Weiter gilt, dass $\nabla h_1(x_2,y_2)=(-2,1)$, $\nabla h_2(x_2,y_2)=(0,-1)$ linear unabhängig sind, d.h. die CF-Bedingung ist erfüllt.

Mögliche Kandidaten: $(x_1,y_1)=(1,1)$, $(x_2,y_2)=(-1,1)$

2. Fall: 1. Restriktion aktiv, 2. Restriktion inaktiv, d.h. $x^2+y=2$ und $y>1$ mit $\mu_2=0$. Mit (1) b) erhält man $\mu_1=x$, eingesetzt in (1) a) $-y-2x+2x^2=0$, d.h. wegen $x^2+y=2$ die Gleichung:

$3x^2-2x-2=0 \iff x=\frac{1}{3}\pm\frac{1}{3}\sqrt{7}$. Wegen $\mu_1=x\geq 0$ ist nur $x=\frac{1}{3}+\frac{1}{3}\sqrt{7}$ zulässig. Dann ist aber $y=2-x^2=2-(\frac{1}{3}+\frac{1}{3}\sqrt{7})^2=\frac{10}{9}-\frac{2}{9}\sqrt{7}<1$ im Wiederspruch zu $y>1$.

Da $\nabla h_1(x,y)=(2x,1)$ linear unabhängig für jedes x, ist die CF-Bedingung für alle (x,y) in diesem Fall erfüllt.

Keine möglichen Kandidaten in diesem Fall.

3. Fall: 1. Restriktion inaktiv, 2. Restriktion aktiv, d.h. $x^2+y<2$ mit $\mu_1=0$ und $y=1$. Mit (1) a) folgt $x=-\frac{1}{2}$ und mit (1)b) $\mu_2=\frac{1}{2}$.

Da $\nabla h_2(x,y)=(0,-1)$ linear unabhängig für alle (x,y) in diesem Fall, gibt es einen möglichen Kandidaten $(x_3,y_3)=(-\frac{1}{2},1)$.

4. Fall: Beide Restriktionen inaktiv, d.h. $x^2+y<2$ und $y>1$ mit $\mu_1=\mu_2=0$. Aus (1) b) und a) erhält man $x=0$ und $y=0$ im Wiederspruch zu $y>1$. Da beide Restriktionen inaktiv sind, ist die CF-Bedingung automatisch erfüllt.

Keine Kandidaten in diesem Fall.

Für die drei Kandidaten gilt nun: $f(1,1)=-2$, $f(-1,1)=0$, $f(-\frac{1}{2},1)=\frac{1}{4}$.

Da die Funktion f stetig ist auf dem abgeschlossenen und beschränkten zulässigen Bereich $M=\{(x,y): x^2+y-2\leq 0, -y+1\leq 0\}$, nimmt f dort ihr Minimum an. Somit ist $(x^*,y^*)=(1,1)$ mit $f(1,1)=-2$ Lösung des Optimierungsproblems.

Auch im Falle von Ungleichungsrestriktionen spielen Konvexitätseigenschaften bei hinreichenden Kriterien eine wichtige Rolle.

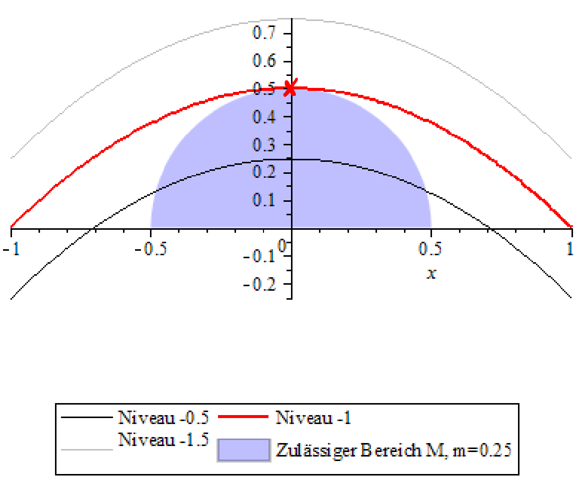
Satz 3.5.4: Wir betrachten $\min f(\vec{x})$, so dass $h_j(\vec{x}) \leq 0$, $j = 1, 2, \ldots, m$, mit der Lagrangefunktion $L(\vec{x}; \mu_1, \ldots, \mu_m)$. Sei $\vec{x}^*$ zulässig und erfülle die KKT-Bedingungen mit zugehörigen $\mu_1^*, \ldots, \mu_m^*$. Dann gilt: Ist $L^*(\vec{x}^*)$ konvex, dann ist $\vec{x}^*$ optimal.

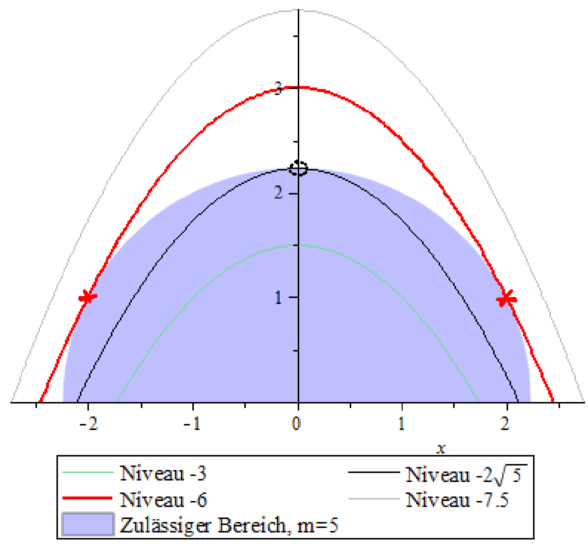
Beispiel 3.5.5: Sei m eine positive Konstante. Wir betrachten

$$\min \; -(x^2 + 2y),$$

$$\text{so dass} \quad \left.\begin{aligned} x^2 + y^2 &\leq m \\ y &\geq 0 \end{aligned}\right\} \text{ oberer Halbkreis um den Ursprung mit Radius } m$$

Wir analysieren die Situation zunächst mit Hilfe der unten stehenden Graphiken. Es sind (wie sich herausstellen wird) zwei typische Situationen gezeichnet: Links ist $m = \frac{1}{4}$, d.h. $0 < m < 1$ und rechts $m = 5$, d.h. $m \geq 1$. Mit Hilfe der eingezeichneten Niveaulinien sieht man, dass das Minimum im zulässigen Bereich in genau einem Punkt, rechts in zwei Punkten angenommen wird. Dies ist, wie wir mit Hilfe der unten stehenden Rechnung sehen werden, typisch für die genannten Fälle $0 < m < 1$ und $m \geq 1$.





Die Lagrangefunktion lautet: $L(x,y;\mu_1,\mu_2) = -x^2-2y+\mu_1(x^2+y^2-m)-\mu_2 y$

KKT-Bedingungen:

(1) a) $L_x(x,y;\mu_1,\mu_2) = -2x+2\mu_1 x = 0$ b) $L_y(x,y;\mu_1,\mu_2) = -2+2\mu_1 y-\mu_2 = 0$

(2) a) $\mu_1 \geq 0$ mit $\mu_1 = 0$ falls $x^2+y^2<m$ b) $\mu_2 \geq 0$ mit $\mu_2 = 0$ falls $y=0$

1. Fall: Beide Restriktionen sind aktiv, d.h. $x^2+y^2=m$ und $y=0$.

Aus (1) b) ergibt sich mit $y=0$: $\mu_2=-2$ im Wiederspruch zu (2) b).

Keine möglichen Kandidaten

2. Fall: 1. Restriktion aktiv, 2. Restriktion inaktiv, d.h. $x^2+y^2=m$ und $y>0$

mit $\mu_2=0$. (1) a) liefert $x(1-\mu_1)=0 \Leftrightarrow x=0 \vee \mu_1=1$

(i) $x=0$ in $x^2+y^2=m$ liefert $y=\pm\sqrt{m}$. Da $y>0$, kommt nur $y=\sqrt{m}$

in Frage. Mit (1) b) erhält man $\mu_1 = \frac{1}{\sqrt{m}}$

Kandidat $(x_1,y_1)=(0,\sqrt{m})$ mit $\mu_1=\frac{1}{\sqrt{m}}$, $\mu_2=0$

(ii) $\mu_1=1$. Mit (1) b) folgt $y=1$ und aus $x^2+y^2=m$ dann $x=\pm\sqrt{m-1}$,

falls $m\geq 1$.

Kandidaten falls $m\geq 1$: $(x_2,y_2)=(\sqrt{m-1},1)$, $(x_3,y_3)=(-\sqrt{m-1},1)$ mit

$\mu_1=1$, $\mu_2=0$

3. Fall: 1. Restriktion inaktiv, 2. Restriktion aktiv, d.h. $x^2+y^2<m$ mit

$\mu_1=0$ und $y=0$. Dann folgt mit 1) b) $\mu_2=-2$ im Wiederspruch zu $\mu_2\geq 0$.

Kein Kandidat.

4. Fall: Beide Restriktionen inaktiv, d.h. $\mu_1=\mu_2=0$ im Wiederspruch

zu (1) b).

Kein Kandidat.

Wir überprüfen nun die Konvexitätseigenschaften der hinreichenden Bedingung.

(i) $0<m<1$; $L^*(x,y;\mu_1^*,\mu_2^*) = -x^2-2y+\frac{1}{\sqrt{m}}(x^2+y^2-m)$

$$H_{L^*}(x,y) = \begin{pmatrix} -2+\frac{2}{\sqrt{m}} & 0 \\ 0 & \frac{2}{\sqrt{m}} \end{pmatrix} = H_{L^*}(0,\sqrt{m})$$

Da für $0<m<1$: $-2+\frac{2}{\sqrt{m}}>0$ und $\begin{vmatrix} -2+\frac{2}{\sqrt{m}} & 0 \\ 0 & \frac{2}{\sqrt{m}} \end{vmatrix} >0$, ist

$H_{L^*}(0,\sqrt{m})$ positiv definit, d.h. L^* dort strikt konvex.

Also ist im Fall $0<m<1$ $f(0,\sqrt{m})=-2\sqrt{m}$ minimal auf dem zulässigen Bereich. (Für $m=1$ nur konvex, für $m>1$ indefinit)

(ii) Für $m\geq 1$ haben wir noch $(\pm\sqrt{m-1},1)$ zu beachten.

Es gilt $f(\pm\sqrt{m-1},1) = -(m-1)-2=-(m+1) \leq -2\sqrt{m}$, da

$m+1\geq 2\sqrt{m} \Longleftrightarrow (m+1)^2\geq 4m \Longleftrightarrow m^2+2m+1\geq 4m$

$\Longleftrightarrow m^2-2m+1\geq 0 \Longleftrightarrow (m-1)^2\geq 0$ (offensichtlich wahr)

Also ist f für $m\geq 1$ minimal für $(\pm\sqrt{m-1},1)$ mit $f(\pm\sqrt{m-1},1)=-(m+1)$

Bemerkung 3.5.6: Es gibt auch schwächere hinreichende Bedingungen als die in Satz 3.5.4 formulierten. Diese setzen aber weitere Begriffe wie quasikonvexe Funktionen und geeignete Kriterien zur Überprüfung voraus. Im Rahmen dieser Vorlesung gehen wir darauf nicht weiter ein.

III.6 Nichtnegativitätsbedingungen

In vielen praktischen Optimierungsproblemen können oder müssen Variablen als nichtnegativ angenommen werden. Wir setzen im Folgenden voraus, dass die Variablen so geordnet sind, dass für die ersten k Variablen Vorzeichenbedingungen gestellt werden, d.h. wir betrachten

$$\min f(\vec{x})$$

so dass $h_j(\vec{x})\leq 0$, $j=1,2,\dots,m$

$x_\ell \geq 0$, $\ell=1,2,\dots,k$

Prinzipiell lassen sich die Vorzeichenbedingungen auch als Restriktionen der Form $h_{m+\ell}(\vec{x})=-x_\ell\leq 0$, $\ell=1,2,\dots,k$ schreiben. Die zugehörige Lagrangefunktion lautet dann

$$L(\vec{x};\mu_1,\dots,\mu_m;\varkappa_1,\dots,\varkappa_k)=f(\vec{x})+\sum_{j=1}^{m}\mu_j h_j(\vec{x})+\sum_{i=1}^{k}\varkappa_i(-x_i)$$

und die Methoden des letzten Kapitels können verwendet werden.

Damit erhält man folgende KKT-Bedingungen.

1) a) $\frac{\partial f}{\partial x_i}(\vec{x}) + \sum_{j=1}^{m} \mu_j \frac{\partial h_j}{\partial x_i}(\vec{x}) - \kappa_i = 0 \quad , i = 1, 2, \dots, k$

b) $\frac{\partial f}{\partial x_i}(\vec{x}) - \sum_{j=1}^{m} \mu_j \frac{\partial h_j}{\partial x_i}(\vec{x}) = 0 \quad , i = k+1, \dots, n$

2) a) $\mu_j \geq 0$ mit $\mu_j = 0$ falls $h_j(\vec{x}) < 0$, $j = 1, 2, \dots, m$

b) $\kappa_i \geq 0$ mit $\kappa_i = 0$ falls $-x_i < 0$, $i = 1, 2, \dots, k$

Die Bedingungen 1)a) und 2)b) lassen sich nun weiter zusammenfassen. Da $\kappa_i \geq 0$ und $\kappa_i = 0$ falls $x_i > 0$, sind 1)a) und 2)b) zusammen äquivalent zu der Bedingung: $\frac{\partial f}{\partial x_i}(\vec{x}) + \sum_{j=1}^{m} \mu_j \frac{\partial h_j}{\partial x_i}(\vec{x}) \geq 0$, $i = 1, \dots, k$, wobei das Gleichheitszeichen gelten muss, falls $x_i > 0$ ist.

Insgesamt haben wir also die KKT-Bedingungen:

1) a) $\frac{\partial f}{\partial x_i}(\vec{x}) + \sum_{j=1}^{m} \mu_j \frac{\partial h_j}{\partial x_i}(\vec{x}) \geq 0$ ($= 0$ falls $x_i > 0$), $i = 1, 2, \dots, k$

b) $\frac{\partial f}{\partial x_i}(\vec{x}) + \sum_{j=1}^{m} \mu_j \frac{\partial h_j}{\partial x_i}(\vec{x}) = 0$, $i = k+1, \dots, n$

2) $\mu_j \geq 0$ mit $\mu_j = 0$ falls $h_j(\vec{x}) < 0$, $j = 1, 2, \dots, m$.

Beispiel 3.6.1: $\min f(x,y) = \frac{1}{2}x^2 - \frac{3}{2}x - \frac{1}{12}y$

so dass $x \leq 5$ (d.h. $h_1(x,y) = x - 5 \leq 0$)

$y \leq 1 + x$ (d.h. $h_2(x,y) = y - x - 1 \leq 0$)

$x \geq 0$

Die Lagrangefunktion lautet $L(x, y; \mu_1, \mu_2) = \frac{1}{2}x^2 - \frac{2}{3}x - \frac{1}{12}y + \mu_1(x-5) + \mu_2(y-x-1)$.

Die KKT-Bedingungen sind:

1) a) $x - \frac{2}{3} + \mu_1 - \mu_2 \geq 0$ ($= 0$ falls $x > 0$)

b) $-\frac{1}{12} + \mu_2 = 0$

2) (i) $\mu_1 \geq 0$ mit $\mu_1 = 0$ falls $x < 5$

(ii) $\mu_2 \geq 0$ mit $\mu_2 = 0$ falls $y < 1 + x$

Aus 1)b) folgt unmittelbar $\mu_2 = \frac{1}{12} > 0$. Nach 2)(ii) muss damit $y = 1 + x$ gelten.

Wir betrachten nun den Fall $\mu_1 > 0$. Dann ist mit 2)(i) $x=5$. Einsetzen von $x=5$ und $\mu_2 = \frac{1}{12}$ in 1)a) liefert $\mu_1 = -5 + \frac{2}{3} + \frac{1}{12} < 0$ im Wiederspruch zu $\mu_1 > 0$.

Für $\mu_1 = 0$ gilt mit 1)a) $x \geq \frac{2}{3} + \frac{1}{12} > 0$. Also ist nach 1)a) $x = \frac{3}{4}$ und $y = 1 + x = \frac{7}{4}$.

Insgesamt erfüllt nur der Punkt $(x^*, y^*) = (\frac{3}{4}, \frac{7}{4})$ mit $\mu_1 = 0, \mu_2 = \frac{1}{12}$ die KKT-Bedingungen.

Beispiel 3.6.2: Wir betrachten das Problem

$$\max \; g(x,y) = \ln(1+x) + y$$

so dass $px + y \leq m$, $x \geq 0$, $y \geq 0$, wobei $p \in (0,1)$, $m > 0$ sein soll.

Das Problem ist äquivalent zu $\min f(x,y) = -\ln(1+x) - y$ mit denselben Restriktionen. Die KKT-Bedingungen lauten also:

1) (i) $-\frac{1}{1+x} + \mu p \geq 0$ $(=0$ falls $x > 0)$

(ii) $-1 + \mu \geq 0$ $(=0$ falls $y > 0)$

2) $\mu \geq 0$ mit $\mu = 0$ falls $px + y < m$

Wegen 1)(ii) kann $\mu = 0$ nicht gelten. Also muss $\mu > 0$ sein, d.h. mit 2) $px + y = m$ (*). Da $p, m > 0$, kann nicht gleichzeitig $x = 0, y = 0$ sein.

$x = 0$: Mit (*) ist dann $y = m > 0$ und mit 1)(ii) $\mu = 1 > 0$

$y = 0$: Mit (*) ist dann $x = \frac{m}{p} > 0$ und mit 1)(i) $\mu = \frac{1}{m+p} > 0$

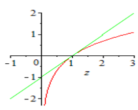
$x > 0, y > 0$: Mit 1)(ii) ist $\mu = 1 > 0$, mit 1)(i) $x = \frac{1}{p} - 1 > 0$, (*) $y = m + 1 - p > 0$

Es gibt also drei Punkte, die die KKT-Bedingungen erfüllen:

$(x_1^*, y_1^*) = (0, m)$ mit $\mu_1^* = 1$, $f_1^* = f(0,m) = -m$

$(x_2^*, y_2^*) = (\frac{m}{p}, 0)$ mit $\mu_2^* = \frac{1}{m+p}$, $f_2^* = f(\frac{m}{p}, 0) = \ln(p) - \ln(m+p)$

$(x_3^*, y_3^*) = (\frac{1}{p} - 1, m + 1 - p)$ mit $\mu_3^* = 1$, $f_3^* = f(\frac{1}{p} - 1, m+1-p) = \ln(p) - m + 1 - p$

Für jedes $z > 0$ gilt: $\ln(z) \leq z - 1$.  Damit:

$f_3^* \leq f_1^* \iff \ln(p) - m + 1 - p \leq -m \iff \ln p \leq p - 1$ ✓

$f_3^* \leq f_2^* \iff \ln(p) - m + 1 - p \leq \ln(p) - \ln(m+p) \iff \ln(m+p) \leq m - 1 + p$ ✓

III.7 Optimierung mit Gleichungs- und Ungleichungsrestriktionen

Wir betrachten noch kurz den Fall, dass Gleichungs- und Ungleichungsrestriktionen, sowie Vorzeichenbedingungen in einem Optimierungsproblem auftreten, d.h.

$$\min f(\vec{x})$$

$$\text{so dass } \quad g_i(\vec{x}) = 0 \quad , i = 1, \dots, m$$

$$h_j(\vec{x}) \leq 0 \quad , j = 1, \dots, s$$

$$x_\ell \geq 0 \quad , \ell = 1, \dots, k$$

Die Behandlung des Problems ergibt sich aus den vorhergehenden Kapiteln. Die Lagrangefunktion mit den Lagrangemultiplikatoren $\lambda_1, \dots, \lambda_m, \mu_1, \dots, \mu_s$ ist $L(\vec{x}; \lambda_1, \dots, \lambda_m; \mu_1, \dots, \mu_s) = f(\vec{x}) + \sum_{i=1}^{m} \lambda_i g_i(\vec{x}) + \sum_{j=1}^{s} \mu_j h_j(\vec{x})$.

Die KKT-Bedingungen sind gegeben durch:

1) a) $\frac{\partial f}{\partial x_\ell}(\vec{x}) + \sum_{i=1}^{m} \lambda_i \frac{\partial g_i}{\partial x_\ell}(\vec{x}) + \sum_{j=1}^{s} \mu_j \frac{\partial h_j}{\partial x_\ell}(\vec{x}) \geq 0$ ($= 0$ falls $x_\ell > 0$), $\ell = 1, \dots, k$

   b) $\frac{\partial f}{\partial x_\ell}(\vec{x}) + \sum_{i=1}^{m} \lambda_i \frac{\partial g_i}{\partial x_\ell}(\vec{x}) + \sum_{j=1}^{s} \mu_j \frac{\partial h_j}{\partial x_\ell}(\vec{x}) = 0$, $\ell = k+1, \dots, n$

2) $\mu_j \geq 0$ mit $\mu_j = 0$ falls $h_j(\vec{x}) < 0$, $j = 1, \dots, m$.