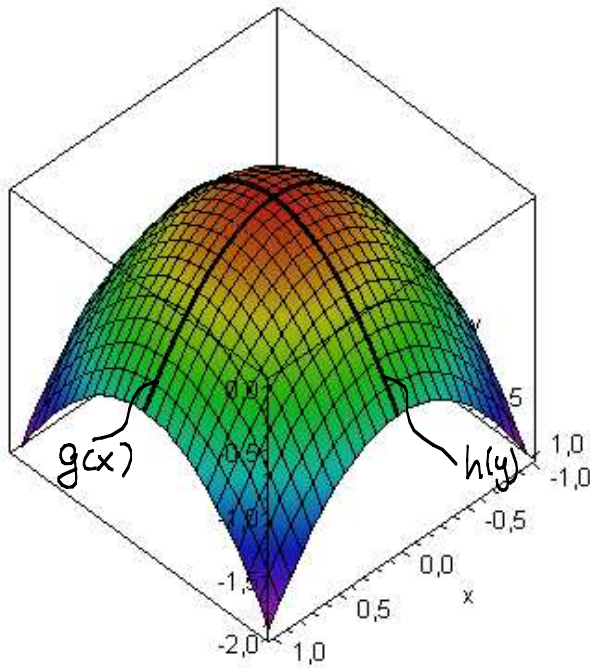




$g(x)$ bzw. $h(y)$ müssen in x^* bzw. y^* ein lokales Maximum besitzen, d.h. sie müssen konkav sein.

Anders ausgedrückt:

$$\frac{d^2}{dx^2} g(x^*) = f_{xx}(x^*, y^*) \leq 0 \text{ und}$$

$$\frac{d^2}{dy^2} h(y^*) = f_{yy}(x^*, y^*) \leq 0.$$

Gilt sogar $\frac{d^2}{dx^2} g(x^*) < 0$ und $\frac{d^2}{dy^2} h(y^*) < 0$, dann garantiert dies, dass g und h tatsächlich

in x^* bzw. y^* lokale Maxima (sogar strikte) besitzen. Anders ausgedrückt garantieren die Bedingungen $f_{xx}(x^*, y^*) < 0$ und $f_{yy}(x^*, y^*) < 0$, dass f ein lokales Maximum hat in den Richtungen durch (x^*, y^*) , die parallel zur x - bzw. y -Achse sind. Dies sagt aber noch nichts aus über das Verhalten von f , wenn wir uns von (x^*, y^*) aus in andere als diese beiden Richtungen bewegen.

Wir erinnern uns daran, dass man unter geeigneten Differenzierbarkeitsvoraussetzungen die Richtungsableitung von f in Richtung $\vec{v}$, $|\vec{v}|=1$, als

$\partial_{\vec{v}} f(x, y) = \langle \text{grad } f(x, y), \vec{v} \rangle = f_x(x, y) \cdot v_1 + f_y(x, y) \cdot v_2$ berechnen kann.

Differenziert man dieses noch einmal in Richtung $\vec{v}$, so erhält man

$$\begin{aligned} \partial_{\vec{v}}^2 f(x, y) &= \langle \text{grad}(f_x(x, y) \cdot v_1 + f_y(x, y) \cdot v_2), \vec{v} \rangle \\ &= \left\langle \begin{pmatrix} f_{xx}(x, y) \cdot v_1 + f_{xy}(x, y) \cdot v_2 \\ f_{xy}(x, y) \cdot v_1 + f_{yy}(x, y) \cdot v_2 \end{pmatrix}, \vec{v} \right\rangle \\ &= f_{xx}(x, y) v_1^2 + 2 f_{xy}(x, y) v_1 v_2 + f_{yy}(x, y) v_2^2 \\ &= \vec{v}^T H_f(x, y) \vec{v} \end{aligned}$$

Ist nun $\vec{v}^T H_f(x^*, y^*) \vec{v} < 0$, dann garantiert dies, dass f ein lokales Maximum in Richtung $\vec{v}$ besitzt. Wenn dies für alle $\vec{v}$ mit $|\vec{v}|=1$ oder äquivalent dazu für alle $\vec{v} \neq \vec{0}$ erfüllt ist, haben wir tatsächlich die Garantie für ein (striktes) lokales Maximum.

Aus der Linearen Algebra wissen wir, dass die Bedingung $\vec{v}^T H_f(x^*, y^*) \vec{v} < 0$ für alle $\vec{v} \neq \vec{0}$ gerade die Definition für die negative Definitheit der Hesse-Matrix $H_f(x^*, y^*)$ von f an der Stelle (x^*, y^*) ist. In der Linearen Algebra haben wir auch gesehen, wie man die negative Definitheit mit dem Hurwitz-Kriterium prüfen kann, das für den Fall einer 2×2 -Matrix in der unten stehenden Bedingung für lokale Extrema direkt eingesetzt wird.

Im Falle lokaler Minima führen analoge Überlegungen (konkav durch konvex ersetzen) auf die Vorzeichenbedingung $\vec{v}^T H_f(x^*, y^*) \vec{v} > 0$ für alle $\vec{v} \neq \vec{0}$, d.h. die positive Definitheit der Hesse-Matrix $H_f(x^*, y^*)$ an der Stelle (x^*, y^*) .

Wir fassen die Ergebnisse nun zusammen. Dazu bezeichnen wir noch mit

$$\Delta_f(x^*, y^*) = \det H_f(x^*, y^*) = f_{xx}(x^*, y^*) \cdot f_{yy}(x^*, y^*) - f_{xy}^2(x^*, y^*)$$

die Determinante der Hesse-Matrix.

Hinreichende Bedingung für lokale Extrema

Sei $f: \mathbb{D}_f \subseteq \mathbb{R}^2 \rightarrow \mathbb{R}$ zweimal stetig partiell differenzierbar und (x^*, y^*) aus dem Inneren von $\mathbb{D}_f$ ein stationärer Punkt von f . Dann gilt:

- 1) Ist $\Delta_f(x^*, y^*) > 0$, dann besitzt f an der Stelle (x^*, y^*) ein (striktes) lokales Extremum.
 - a) Ist zusätzlich $f_{xx}(x^*, y^*) < 0$ (oder $f_{yy}(x^*, y^*) < 0$), dann liegt ein relatives Maximum vor.
 - b) Ist zusätzlich $f_{xx}(x^*, y^*) > 0$ (oder $f_{yy}(x^*, y^*) > 0$), dann liegt ein relatives Minimum vor.

- 2) Ist $\Delta_f(x^*, y^*) < 0$, dann besitzt f an der Stelle (x^*, y^*) einen Sattelpunkt.
 3) Ist $\Delta_f(x^*, y^*) = 0$, so lässt sich mit diesem Kriterium keine Aussage machen.

Bemerkung: $f_{xx}(x^*, y^*) > 0$ und $\Delta_f(x^*, y^*) > 0$ bedeutet nach dem Hurwitz-Kriterium, dass $H_f(x^*, y^*)$ positiv definit ist; $f_{xx}(x^*, y^*) < 0$ und $\Delta_f(x^*, y^*) > 0$ dagegen, dass $H_f(x^*, y^*)$ negativ definit ist.

Wir wenden das hinreichende Kriterium nun auf die vorher bereits betrachteten Beispiele an.

Beispiel: $f(x, y) = 3xy - x^3 - y^3$.

Die stationären Punkte von f sind $P_1(0, 0)$ und $P_2(1, 1)$.

Es gilt:
$$H_f(x, y) = \begin{pmatrix} -6x & 3 \\ 3 & -6y \end{pmatrix}, \quad \Delta_f(x, y) = 36xy - 9$$

Da $\Delta_f(0, 0) = -9 < 0$ ist, besitzt f an der Stelle $(0, 0)$ einen Sattelpunkt.

Wegen $f_{xx}(1, 1) = -6 < 0$ und $\Delta_f(1, 1) = 27 > 0$ folgt, dass f an der Stelle $(1, 1)$ ein lokales Maximum $f(1, 1) = 1$ besitzt.

Beispiel: $f(x, y) = x \cdot \ln(x^2 + y^2)$, $D_f = \mathbb{R}^2 \setminus \{(0, 0)\}$.

Die stationären Punkte von f sind $P_1(0, 1)$, $P_2(0, -1)$, $P_3(e^{-1}, 0)$, $P_4(-e^{-1}, 0)$.

$$\text{grad } f(x, y) = \begin{pmatrix} \ln(x^2 + y^2) + \frac{2x^2}{x^2 + y^2} \\ \frac{2xy}{x^2 + y^2} \end{pmatrix}$$

Es gilt:
$$H_f(x, y) = \begin{pmatrix} \frac{2x(x^2 + 3y^2)}{(x^2 + y^2)^2} & \frac{2y(y^2 - x^2)}{(x^2 + y^2)^2} \\ \frac{2y(y^2 - x^2)}{(x^2 + y^2)^2} & \frac{2x(x^2 - y^2)}{(x^2 + y^2)^2} \end{pmatrix}$$

$$\begin{aligned} \Delta_f(x, y) &= \frac{1}{(x^2 + y^2)^4} (4x^2(x^2 + 3y^2)(x^2 - y^2) - 4y^2(y^2 - x^2)^2) \\ &= \frac{4(x^2 - y^2)}{(x^2 + y^2)^4} (x^4 + 3x^2y^2 - x^2y^2 + y^4) = \frac{4(x^2 - y^2)(x^2 + y^2)^2}{(x^2 + y^2)^4} = 4 \cdot \frac{x^2 - y^2}{(x^2 + y^2)^2} \end{aligned}$$

Da $\Delta_f(0,1) = -4 < 0$ und $\Delta_f(0,-1) = -4 < 0$, besitzt f an den Stellen $(0,1)$ und $(0,-1)$ Sattelpunkte.

Wegen $f_{xx}(e^{-1}, 0) = 2e > 0$ und $\Delta_f(e^{-1}, 0) = 4e^2 > 0$ folgt, dass f an der Stelle $(e^{-1}, 0)$ ein lokales Minimum $f(e^{-1}, 0) = -2e^{-1}$ besitzt.

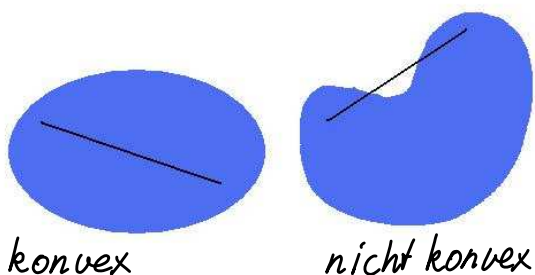
Wegen $f_{xx}(-e^{-1}, 0) = -2e < 0$ und $\Delta_f(-e^{-1}, 0) = 4e^2 > 0$ folgt, dass f an der Stelle $(-e^{-1}, 0)$ ein lokales Maximum $f(-e^{-1}, 0) = 2e^{-1}$ besitzt.

Wir haben bereits gesehen, dass es oft nicht einfach ist, globale Extrema zu bestimmen. Die Situation wird überschaubarer, wenn man geeignete zusätzliche Bedingungen an die Funktion bzw. den Bereich, für den globale Extremalstellen bestimmt werden sollen, stellt, was im Folgenden erläutert werden soll.

Konvexe Mengen und konvexe Funktionen

Der Vorteil konvexer Funktionen, die auf einer konvexen Menge definiert sind, besteht darin, dass sich die Frage nach globalen Extrema viel einfacher beantworten lässt.

Definition: Eine Menge $S \subseteq \mathbb{R}^2$ heißt konvex, wenn für beliebige Punkte $(a_1, a_2), (b_1, b_2) \in S$ gilt: $(\lambda a_1 + (1-\lambda)b_1, \lambda a_2 + (1-\lambda)b_2) \in S$ für jedes $\lambda \in [0, 1]$.



konvex

nicht konvex

Anschaulich bedeutet Konvexität, dass mit je zwei Punkten der Menge S auch die komplette Verbindungsstrecke der beiden Punkte in S liegt.

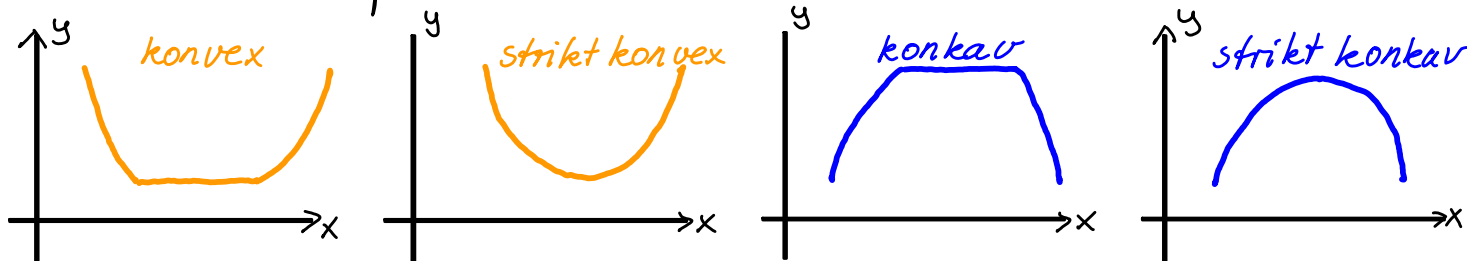
Diese im $\mathbb{R}^2$ sehr anschauliche Definition gilt sinngemäß auch im $\mathbb{R}^n$, aber

auch in $\mathbb{R}$. So sind die konvexen Teilmengen von $\mathbb{R}$ neben $\mathbb{R}$ selbst alle Intervalle, egal ob offen, halboffen, abgeschlossen, beschränkt oder unbeschränkt.

Bei den hinreichenden Bedingungen für lokale Extrema haben wir bereits gesehen, dass das Krümmungsverhalten (konvex, konkav) eine wichtige Rolle spielt. Anschaulich ist die Definition einer konvexen bzw. konkaven Funktion bei zwei Variablen analog zu der bei einer Variablen.

So nennt man eine Funktion auf einer konvexen Menge $\begin{matrix} \text{konvex} \\ \text{strikt konvex} \end{matrix}$, wenn mit je zwei Punkten die Verbindungsstrecke $\begin{matrix} \text{nicht unterhalb} \\ \text{oberhalb} \end{matrix}$ des Funktionsgraphen verläuft und $\begin{matrix} \text{konkav} \\ \text{strikt konkav} \end{matrix}$, wenn die Verbindungsstrecke $\begin{matrix} \text{nicht oberhalb} \\ \text{unterhalb} \end{matrix}$ verläuft.

Anschaulich für Funktionen einer Variablen



Wie auch bei Funktionen einer Variablen lässt sich unter geeigneten Differenzierbarkeitsvoraussetzungen das Krümmungsverhalten mit Hilfe der zweiten- hier partiellen- Ableitungen charakterisieren. Im Grunde haben wir dies schon bei unseren Überlegungen zu den hinreichenden Bedingungen für lokale Extrema gesehen. Genauer gilt:

Satz: Sei $f: D_f \subseteq \mathbb{R}^2 \rightarrow \mathbb{R}$, $D_f \neq \emptyset$ und D_f konvex. f sei zweimal stetig partiell differenzierbar. Dann gilt:

- 1) $H_f(x, y)$ positiv semidefinit für alle $(x, y) \in D_f \Leftrightarrow f$ konvex
- 2) $H_f(x, y)$ positiv definit für alle $(x, y) \in D_f \Rightarrow f$ strikt konvex
- 3) $H_f(x, y)$ negativ semidefinit für alle $(x, y) \in D_f \Leftrightarrow f$ konkav
- 4) $H_f(x, y)$ negativ definit für alle $(x, y) \in D_f \Rightarrow f$ strikt konkav

Bemerkung: Die Definitheitseigenschaften der symmetrischen Hesse-Matrix lassen sich auch hier wieder mit den in der Linearen Algebra behandelten Kriterien (z.B. Hurwitz-Kriterium) untersuchen.

Beispiel: $f(x,y) = x^4 + 2y^2$, $\mathbb{D}_f = \mathbb{R}^2$, $\mathbb{D}_f$ ist konvex.

Es gilt: $\text{grad } f(x,y) = \begin{pmatrix} 4x^3 \\ 4y \end{pmatrix}$, $H_f(x,y) = \begin{pmatrix} 12x^2 & 0 \\ 0 & 4 \end{pmatrix}$

Da $f_{xx}(x,y) = 12x^2 \geq 0$, $f_{yy}(x,y) = 4 \geq 0$ und $\Delta_f(x,y) = 48x^2 \geq 0$ für alle $(x,y) \in \mathbb{D}_f$, ist $H_f(x,y)$ nach dem Hurwitz-Kriterium positiv semidefinit für alle $(x,y) \in \mathbb{D}_f$. Somit ist f konvex auf $\mathbb{D}_f$.

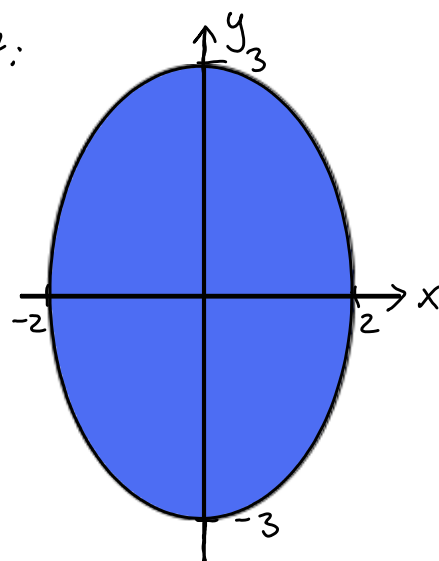
Beispiel: $f(x,y) = \sqrt{36 - 9x^2 - 4y^2}$

f ist definiert für alle $(x,y) \in \mathbb{R}^2$, für die gilt:

$$36 - 9x^2 - 4y^2 \geq 0 \Leftrightarrow \frac{x^2}{4} + \frac{y^2}{9} \leq 1$$

Somit ist $\mathbb{D}_f = \{(x,y) \in \mathbb{R}^2 : \frac{x^2}{4} + \frac{y^2}{9} \leq 1\}$.

Dabei handelt es sich um eine Ursprungs-ellipse, deren Rand die x -Achse in $(2,0)$ und $(-2,0)$ und die y -Achse in $(0,3)$ und $(0,-3)$ schneidet. Somit ist $\mathbb{D}_f$ konvex.



Da die partiellen Ableitungen von f auf dem Rand der Ellipse nicht existieren, betrachten wir den eingeschränkten Definitionsbereich

$$\tilde{\mathbb{D}}_f = \{(x,y) \in \mathbb{R}^2 : \frac{x^2}{4} + \frac{y^2}{9} < 1\}, \text{ d.h. die offene Ellipse.}$$

Es gilt:

$$\text{grad } f(x,y) = \begin{pmatrix} \frac{-9x}{\sqrt{36 - 9x^2 - 4y^2}} \\ \frac{-4y}{\sqrt{36 - 9x^2 - 4y^2}} \end{pmatrix}, \quad H_f(x,y) = \begin{pmatrix} \frac{36(y^2 - 9)}{(36 - 9x^2 - 4y^2)^{3/2}} & \frac{-36xy}{(36 - 9x^2 - 4y^2)^{3/2}} \\ \frac{-36xy}{(36 - 9x^2 - 4y^2)^{3/2}} & \frac{36(x^2 - 4)}{(36 - 9x^2 - 4y^2)^{3/2}} \end{pmatrix},$$

$$\begin{aligned} \Delta_f(x,y) &= \frac{36^2}{(36 - 9x^2 - 4y^2)^3} \{(y^2 - 9)(x^2 - 4) - x^2 y^2\} = \frac{36^2}{(36 - 9x^2 - 4y^2)^3} \{-9y^2 - 4x^2 + 36\} \\ &= \frac{36^2}{(36 - 9x^2 - 4y^2)^2} > 0 \text{ für alle } (x,y) \in \tilde{\mathbb{D}}_f. \end{aligned}$$

Da $36 - 9x^2 - 4y^2 > 0$ für alle $(x,y) \in \tilde{\mathbb{D}}_f$, ist auch insbesondere

$$36 - 4y^2 > 0 \Leftrightarrow y^2 - 9 < 0, \text{ d.h. } f_{xx}(x,y) < 0 \text{ für alle } (x,y) \in \tilde{\mathbb{D}}_f.$$

Insgesamt ist also $H_f(x,y)$ negativ definit auf $\tilde{\mathbb{D}}_f$ und damit f strikt konkav.

Der Vorteil konvexer (konkaver) Funktionen bei der Suche nach globalen Minima (Maxima) zeigt sich in dem folgenden Satz. Zuvor wollen wir noch bemerken, dass die Suche nach Maxima einer Funktion äquivalent ist zu der Suche nach Minima von $-f$. Es bedeutet daher keine Einschränkung, wenn wir uns im Folgenden auf die Untersuchung von Minima beschränken.

Satz: Sei $f: \mathbb{D}_f \subseteq \mathbb{R}^2 \rightarrow \mathbb{R}$, $\mathbb{D}_f \neq \emptyset$, $\mathbb{D}_f$ konvex. Dann gilt:

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

Im Folgenden betrachten wir nun einige Beispiele mit anwendungsorientiertem Hintergrund.

Beispiel: 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 $\mathbb{D}_f = \{(Q_1, Q_2) \in \mathbb{R}^2 : Q_1 > 0 \wedge Q_2 > 0\}$ eine konvexe Menge.

Als notwendige Bedingung für ein lokales Extremum ermitteln wir:

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

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 $\mathbb{D}_f$.

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

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

Der maximal mögliche Gewinn beträgt somit

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

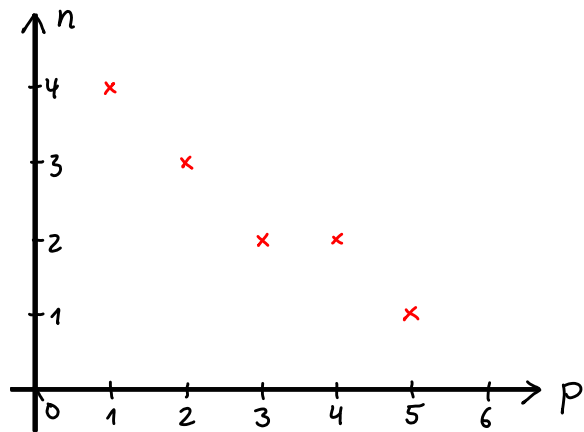
für die Mengen $Q_1^* = \frac{a_1 - \alpha}{2b_1}$, $Q_2^* = \frac{a_2 - \alpha}{2b_2}$ zu den Preisen $p_1^* = \frac{a_1 + \alpha}{2}$, $p_2^* = \frac{a_2 + \alpha}{2}$.

Lineare Regression

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: Für ein Produkt hat eine Marktanalyse folgende Daten (p_i, n_i) , $i = 1, 2, \dots, 5$ für die Nachfrage nach einem Produkt in Abhängigkeit vom Preis ergeben.

i	1	2	3	4	5
p_i	1	2	3	4	6
n_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 $N(p) = \alpha + \beta \cdot p$ 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. Dazu muss man zunächst festlegen, was unter "möglichst gering" verstanden werden soll, denn dafür gibt es prinzipiell sehr unterschiedliche Möglichkeiten. Eine Möglichkeit besteht darin, alle Abweichungen gleich zu gewichten, d.h. die Summe der Beträge der Abweichungen $\sum_{i=1}^5 |N(p_i) - n_i|$ zu minimieren. Eine andere Möglichkeit ist, die maximale Abweichung, d.h. den größten der Werte $|N(p_i) - n_i|$ möglichst klein zu bekommen. Geringere Abweichungen bleiben dabei unberücksichtigt. Die beiden bisher genannten Möglichkeiten sind mathematisch nicht so einfach zu lösen. Ein gängiges gut handhabbares Kriterium für Optimalität, d.h. hier eine "möglichst gute Gerade", ist die Minimierung der Summe der Fehlerquadrate. Man sucht also $\alpha, \beta \in \mathbb{R}$, so dass

$$f(\alpha, \beta) = \sum_{i=1}^5 (N(p_i) - n_i)^2 = \sum_{i=1}^5 (\alpha + \beta p_i - n_i)^2 \text{ minimal wird.}$$

Zunächst gilt: $f_\alpha(\alpha, \beta) = \sum_{i=1}^5 2(\alpha + \beta p_i - n_i)$, $f_\beta(\alpha, \beta) = \sum_{i=1}^5 2p_i(\alpha + \beta p_i - n_i)$,

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:

$$\text{grad } f(\alpha, \beta) = \vec{0} \Leftrightarrow \begin{cases} 5\alpha + 16\beta - 12 = 0 \\ 16\alpha + 66\beta - 30 = 0 \end{cases} \Leftrightarrow \begin{pmatrix} 5 & 16 \\ 16 & 66 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} 12 \\ 30 \end{pmatrix}$$

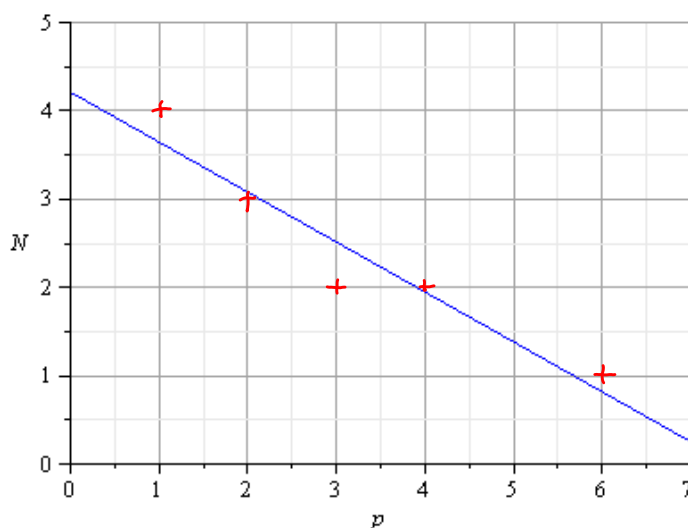
Dies ist ein lineares Gleichungssystem, das sich z.B. mit der Cramerschen Regel (vgl. Lineare Algebra) lösen lässt. Die eindeutig bestimmte Lösung lautet: $\alpha^* = \frac{156}{37}$, $\beta^* = -\frac{21}{37}$.

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

Da außerdem $f_{\alpha\alpha}(\alpha, \beta) = 10 > 0$, folgt, dass f strikt konvex ist auf der konvexen Menge $\mathbb{D}_f = \mathbb{R}^2$. 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

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



Den am Beispiel geschilderten 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 $(x_i, y_i), i=1, 2, \dots, m$. 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:

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

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

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:

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

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

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

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

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:

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

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) in $\mathbb{R}^2$ verteilt wären, bei denen alle x_i gleich sind!).

Da außerdem $f_{xx}(\alpha, \beta) = 2m > 0$ 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 $\mathcal{D}_f = \mathbb{R}^2$. f nimmt somit für die oben angegebenen α^*, β^* ihr globales Minimum an.

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

Wir bezeichnen mit

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

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

Zusammenfassung der Ergebnisse

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

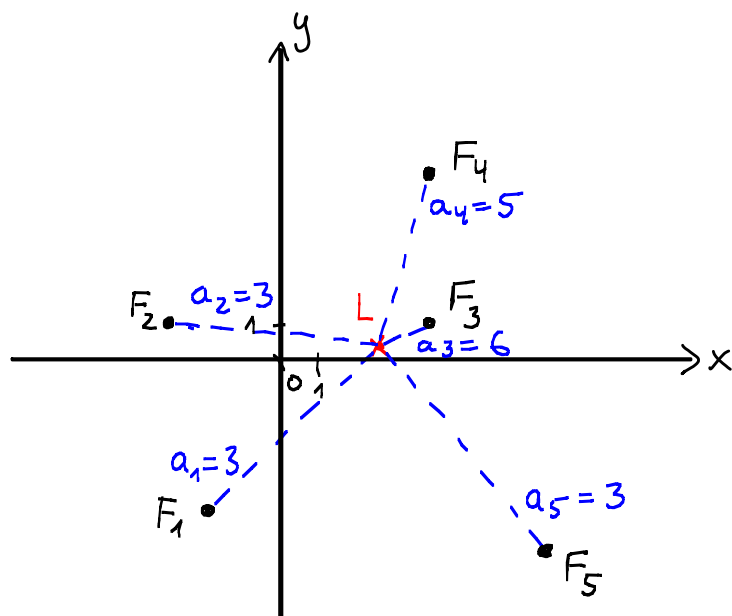
$$y(x) = \left(\bar{y} - r \cdot \frac{s_y}{s_x} \cdot \bar{x} \right) + r \cdot \frac{s_y}{s_x} \cdot x$$

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 zur 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

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

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:

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

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

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.