

Problem 2:

Sei $y = f(\mathbf{x}) = f(x_1, \dots, x_n)$ eine Funktion und $S \subset \mathbb{R}^n$ beschränkt und abgeschlossen.

Bestimme das globale Maximum und Minimum von f auf S .

Inhaltsverzeichnis

1	Motivation und Beispiele	2
2	Mathematisches Vorwissen	5
2.1	Teilmengen von \mathbb{R}^n	5
2.2	Wichtige Funktionsklassen	8
2.3	Gradient und Hessematrix	13
2.4	Die allgemeine Richtungsableitung	16
2.5	Die Taylor-Formel	18
2.6	Lokale Extremalstellen im Inneren	19
3	Lösungswege	22
3.1	Das allgemeine Maximierungsproblemen	22
3.2	f, \mathbf{g} stet. dfb, $\mathbf{g}(\mathbf{x}) = \mathbf{c}$, rang $D\mathbf{g}$ maximal \rightarrow Lagrange-Verfahren	25
3.2.1	n Variablen und eine Nebenbedingung	25
3.2.2	n Variablen und m Nebenbedingungen	29
3.3	f, \mathbf{g} stet. dfb, $\mathbf{g}(\mathbf{x}) \leq \mathbf{c}$, rang $D\mathbf{g}$ maximal \rightarrow Kuhn-Tucker	32
3.3.1	Zwei Variablen und eine Nebenbedingung	32
3.3.2	n Variablen und m Nebenbedingung	35
3.4	Numerische Verfahren zur Maximierung	37
4	Übungen	39

1 Motivation und Beispiele

Einführung

Zwei der wichtigsten und allgemeinsten ökonomische Probleme sind sicher:

1. Maximiere den Erfolg bei gegebenen Mitteln.
2. Erreiche einen gegebenen Erfolg mit minimalen Mitteln.

Ein ganz typisches Beispiel ist sicher das folgende:

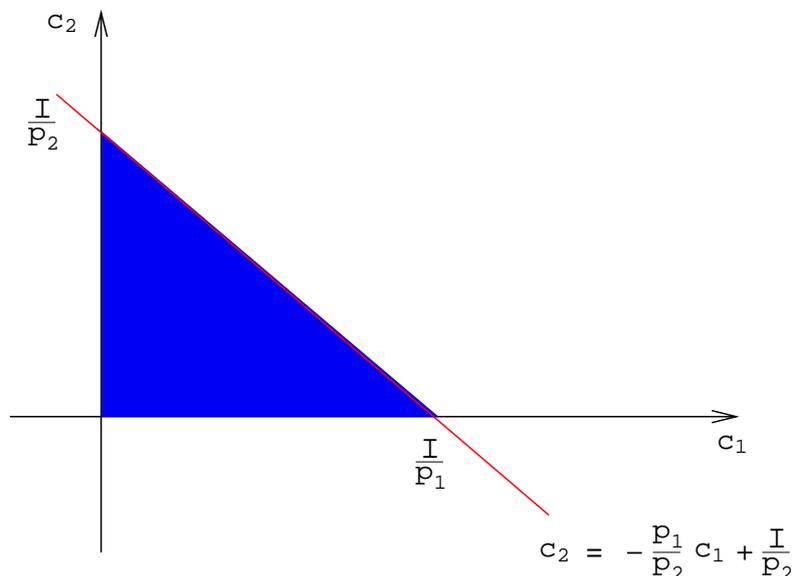
Beispiel 1.1 *Es bezeichne p_1 und p_2 die Preise von zwei Gütern und I das verfügbare Einkommen einer Person. Weiterhin sei $u(c_1, c_2)$ eine Nutzenfunktion, d.h. eine Funktion, die den Nutzen (Grad der Bedürfnisbefriedigung der Person) als Funktion der konsumierten Mengen c_1 und c_2 der beiden Güter darstellt. Wir suchen einen Konsumplan (c_1^*, c_2^*) , welcher die Nutzenfunktion $u(c_1, c_2)$ unter der Budgetrestriktion $p_1c_1 + p_2c_2 = I$ maximiert, also kurz*

$$\begin{aligned} \text{zu maximierende Zielfunktion} & : u(c_1, c_2) \\ \text{Nebenbedingung(en)} & : \phi(c_1, c_2) = p_1c_1 + p_2c_2 - I = 0 \\ & c_1 \geq 0 \\ & c_2 \geq 0 \end{aligned}$$

Hier kann die Nebenbedingung leicht nach c_2 aufgelöst werden und man erhält die Geradengleichung

$$c_2 = -\frac{p_1}{p_2}c_1 + \frac{I}{p_2}.$$

Da beide Preise positiv sind, ist die Gerade sicher fallend. Ausserdem erkennt man sofort, dass eine Änderung des Einkommens den Anstieg nicht beeinflusst. Alle Punkte der blauen Fläche haben die Eigenschaften: $c_1, c_2 \geq 0$ und $p_1c_1 + p_2c_2 \leq I$, d.h. die entsprechenden Güterbündel kann ich mir leisten.



Beispiel: Portfoliotheorie

Einführung In der Praxis besitzen Investoren wohl Tausende von verschiedenen Anlagemöglichkeiten wie z.B. Aktien, Anleihen, Fonds, Edelmetalle und Währungen. Normalerweise wird ein (vernünftiger) Investor sein Geld auf mehrere Anlagen verteilen (Portfolio). Doch auf welche Art sollte ein Investor sein Portfolio zusammenstellen? Die zwei wesentlichen Begriffe für jede Investition sind dabei die (erwartete) **Rendite**, die man natürlich maximieren möchte, und das (erwartete) **Risiko**, das man minimieren möchte.

Die erwartete Rendite könnten wir durch das arithmetische Mittel der relativen Kursänderungen der vergangenen Jahre schätzen. Als Mass für das Risiko eignet sich die Varianz (Stärke der Ausschläge) der erwarteten Rendite. Auch die erwartete Varianz einer Anlage kann aus den Daten der vergangenen Jahre geschätzt werden.

Im Jahr 1952 publizierte der 25-jährige Doktorand *Harry Markowitz* von der University of Chicago eine 15-seitige Arbeit mit dem Titel „Portfolio Selection“ im „Journal of Finance“. Für diesen Meilenstein der Finanzmathematik wurde dem Verfasser 1990 der Nobelpreis für Wirtschaftswissenschaften verliehen.

Mathematisches Modell Wir betrachten n verschiedene Geldanlagen $1, 2, \dots, n$ mit zugehörigen Gewichten $\mathbf{x} = (x_1, x_2, \dots, x_n)$, d.h. $\sum_{i=1}^n x_i = 1$ und $x_i \geq 0$ für alle i . Aus diesen Geldanlagen soll ein Portfolio so zusammengestellt werden, dass ein gegebenes Guthaben den Gewichten x_1, x_2, \dots, x_n entsprechend in die n Geldanlagen investiert wird. Die Aufteilung wollen wir nach gewissen sinnvollen Vorgaben optimal gestalten. Weiterhin seien

- $\hat{\mu}_i$ bzw. $\hat{\mu}_P$ die erwarteten Renditen der Geldanlage i bzw. des Gesamtportfolios P ;
- $\hat{\sigma}_i^2$ bzw. $\hat{\sigma}_P^2$ die erwarteten Varianzen der Renditen der Geldanlage i bzw. des Gesamtportfolios P ;
- $\hat{\sigma}_{ij}$ die erwartete Kovarianz der Renditen der Geldanlagen i und j .

Die Kovarianz ist ein Mass für den Gleichlauf zweier Zahlenreihen. Eine grosse positive Kovarianz würde bedeuten, dass sich beide Geldanlagen ähnlich entwickeln (beide fallen oder steigen gleichzeitig), z.B. bei zwei Aktien aus derselben Branche. Eine kleine (also negative) Kovarianz bedeutet dagegen, dass sich beide Geldanlagen eher gegenläufig entwickeln (steigt die eine, so fällt die andere im Kurs). Es scheint klar zu sein, dass diese Kovarianzen in die Berechnung der Varianz (Risiko) des Portfolios eingehen müssen.

Für die erwartete Rendite und die Varianz der Portfoliorendite gilt:

$$\hat{\mu}_P = \sum_{i=1}^n x_i \hat{\mu}_i = x_1 \hat{\mu}_1 + x_2 \hat{\mu}_2 + \dots + x_n \hat{\mu}_n$$

$$\hat{\sigma}_P^2 = \sum_{i=1}^n x_i^2 \hat{\sigma}_i^2 + \sum_{i \neq j} x_i x_j \hat{\sigma}_{ij} = \mathbf{x}^T \underbrace{\begin{pmatrix} \hat{\sigma}_1^2 & \hat{\sigma}_{12} & \dots & \hat{\sigma}_{1n} \\ \hat{\sigma}_{21} & \hat{\sigma}_2^2 & \dots & \hat{\sigma}_{2n} \\ \vdots & & \ddots & \vdots \\ \hat{\sigma}_{n1} & \hat{\sigma}_{n2} & \dots & \hat{\sigma}_n^2 \end{pmatrix}}_{=:C} \mathbf{x}$$

Wir könnten nun versuchen, ein besonders schönes Portfolio zusammenzustellen, z.B. eines, das eine vorgegebene erwartete Rendite $\hat{\mu}_P$ bei minimaler Varianz (interpretiert als Risiko) realisiert.

2 Mathematisches Vorwissen

2.1 Teilmengen von \mathbb{R}^n

ϵ -Umgebungen

Sei $\mathbf{x} \in \mathbb{R}^n$ und $\epsilon > 0$ eine reelle Zahl. Dann ist die ϵ -Umgebung von \mathbf{x} die Menge

$$\begin{aligned} U_\epsilon(\mathbf{x}) &= \{ \mathbf{y} \in \mathbb{R}^n \mid \text{Abstand von } \mathbf{y} \text{ zu } \mathbf{x} \text{ ist kleiner als } \epsilon \} \\ &= \{ \mathbf{y} \in \mathbb{R}^n \mid \|\mathbf{x} - \mathbf{y}\| < \epsilon \} \end{aligned}$$

Verbindungsstrecken

Seien $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Dann ist die Verbindungsstrecke $\overline{\mathbf{x}\mathbf{y}}$ dieser beiden Punkte die Menge

$$\overline{\mathbf{x}\mathbf{y}} = \{ t \cdot \mathbf{x} + (1 - t) \cdot \mathbf{y} \mid t \in [0, 1] \}$$

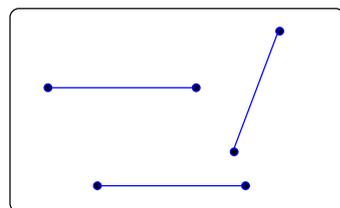
Dass die so beschriebene Menge tatsächlich die Verbindungsstrecke der beiden Punkte ist, kann man wie folgt einsehen:

$$t \cdot \mathbf{x} + (1 - t) \cdot \mathbf{y} = \mathbf{y} + t \cdot (\mathbf{x} - \mathbf{y}) = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} + t \cdot \begin{pmatrix} x_1 - y_1 \\ x_2 - y_2 \\ \vdots \\ x_n - y_n \end{pmatrix}$$

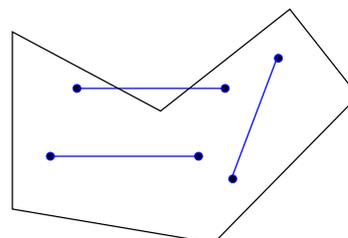
und das ist eine Geradengleichung mit Aufpunkt \mathbf{y} und Richtungsvektor $\mathbf{x} - \mathbf{y}$. Beachtet man nun noch die obige Einschränkung an den reellen Parameter t , erhält man als Anfangspunkt \mathbf{y} (falls $t = 0$) und als Endpunkt \mathbf{x} (falls $t = 1$).

Konvexe Mengen

Definition 2.1 Eine Teilmenge $M \subset \mathbb{R}^n$ heisst konvex, falls für alle Punkte $\mathbf{x}, \mathbf{y} \in M$ auch die gesamte Verbindungsstrecke $\overline{\mathbf{x}\mathbf{y}}$ in M liegt.



konvex



nicht konvex

Der Graph einer linearen Funktion (Gerade, Ebene, Hyperebene) teilt den umgebenen Raum in natürlicher Weise in zwei Teile, wobei wir jeden dieser Teile als Halbraum bezeichnen wollen.

Eine in vielen Anwendungen vorkommende Klasse von konvexen Mengen sind Mengen, die sich als Durchschnitt von endlich vielen abgeschlossenen Halbräumen schreiben lassen.

Satz 1 Sei A eine $m \times n$ -Matrix und $\mathbf{b} \in \mathbb{R}^m$ ein Vektor

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \quad \text{und} \quad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

Dann ist die Menge

$$M = \{ \mathbf{x} \in \mathbb{R}^n \mid A\mathbf{x} \leq \mathbf{b} \text{ und } \mathbf{x} \geq \mathbf{0} \}$$

konvex.

Beweisidee:

Seien $\mathbf{x}_1, \mathbf{x}_2 \in M$, dann gilt natürlich

$$A\mathbf{x}_1 \leq \mathbf{b} \tag{1}$$

$$A\mathbf{x}_2 \leq \mathbf{b} \tag{2}$$

$$\mathbf{x}_1 \geq \mathbf{0} \tag{3}$$

$$\mathbf{x}_2 \geq \mathbf{0}. \tag{4}$$

Für $t \in [0, 1]$ folgt dann sofort $tA\mathbf{x}_1 \leq t\mathbf{b}$, $(1-t)A\mathbf{x}_2 \leq (1-t)\mathbf{b}$, $t\mathbf{x}_1 \geq \mathbf{0}$ und $(1-t)\mathbf{x}_2 \geq \mathbf{0}$.
Aus der ersten und zweiten Ungleichung folgt dann

$$tA\mathbf{x}_1 + (1-t)A\mathbf{x}_2 = A(t\mathbf{x}_1 + (1-t)\mathbf{x}_2) \leq t\mathbf{b} + (1-t)\mathbf{b} = \mathbf{b}.$$

Aus der dritten und vierten Ungleichung folgt

$$t\mathbf{x}_1 + (1-t)\mathbf{x}_2 \geq \mathbf{0}.$$

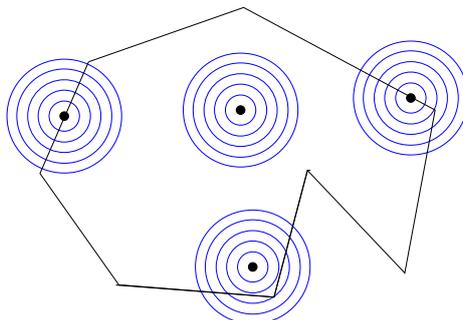
Somit gilt auch $t\mathbf{x}_1 + (1-t)\mathbf{x}_2 \in M$ und die Menge M ist konvex. □

Beschränkte Mengen

Definition 2.2 Eine Menge $S \subset \mathbb{R}^n$ heisst beschränkt, wenn es eine positive reelle Zahl ϵ gibt, so dass $S \subset U_\epsilon(\mathbf{0})$ gilt. (Kurz: S ist ganz in einer Kugel mit endlichem Radius enthalten).

Randpunkte und innere Punkte

Definition 2.3 Sei $M \subset \mathbb{R}^n$. Ein Punkt $\mathbf{x} \in M$ heisst Randpunkt von M , falls **jede** ϵ -Umgebung von \mathbf{x} sowohl Punkte aus M als auch aus dem Komplement $M^c = \mathbb{R}^n - M$ von M enthält. Ein Punkt von M , der kein Randpunkt ist, heisst innerer Punkt von M .



Ein Randpunkt von M kann ein Element von M sein oder auch nicht. Als Beispiel mag hier das halboffene Intervall $M = (1, 2] \subset \mathbb{R}$ dienen. Sicher hat M die beiden Randpunkte 1 und 2, denn jede Umgebung der beiden Punkte schneidet sowohl M als auch das Komplement von M , aber es gilt $2 \in M$ und $1 \notin M$.

Definition 2.4 Der Rand ∂M von M ist gegeben durch

$$\partial M = \{ \mathbf{x} \in \mathbb{R}^n \mid \text{für alle } \epsilon > 0 \text{ gilt } U_\epsilon(\mathbf{x}) \cap M \neq \emptyset \text{ und } U_\epsilon(\mathbf{x}) \cap M^c \neq \emptyset \}$$

Wenn man sich also auf dem Rand einer Menge M befindet, dann enthält jede Umgebung sowohl Punkte aus M als auch aus dem Komplement M^c . Für $M = (1, 2] \subset \mathbb{R}$ gilt $\partial M = \{1, 2\}$ (die Menge, die aus den beiden Punkten 1 und 2 besteht).

Offene und abgeschlossene Mengen

Definition 2.5 Eine Teilmenge $M \subset \mathbb{R}^n$ heisst offen, wenn es für jedes $\mathbf{x} \in M$ ein $\epsilon > 0$ gibt, so dass $U_\epsilon(\mathbf{x}) \subset M$. Eine Teilmenge $A \subset \mathbb{R}^n$ heisst abgeschlossen, wenn ihr Komplement $A^c = \mathbb{R}^n - A$ offen ist.

2.2 Wichtige Funktionsklassen

Lineare Funktionen

Definition 2.6 Eine lineare Funktion ist eine Funktion der Gestalt

$$\begin{aligned} f : \mathbb{R}^n &\longrightarrow \mathbb{R} \\ \mathbf{x} = (x_1, x_2, \dots, x_n)^T &\longmapsto a_1x_1 + a_2x_2 + \dots + a_nx_n = \mathbf{a} \bullet \mathbf{x} = f(\mathbf{x}) \end{aligned}$$

für beliebige reelle Zahlen a_1, a_2, \dots, a_n bzw. $\mathbf{a} = (a_1, a_2, \dots, a_n)^T$.

Im Allgemeinen definiert man eine lineare Funktion durch die folgende Eigenschaft: Für alle \mathbf{x}, \mathbf{y} im Definitionsbereich von f und alle reelle Zahlen a und b gilt:

$$f(a\mathbf{x} + b\mathbf{y}) = af(\mathbf{x}) + bf(\mathbf{y}).$$

Quadratische Formen

Definition 2.7 Sei A eine symmetrische $(n \times n)$ -Matrix. Dann heisst die Funktion

$$\begin{aligned} Q_A : \mathbb{R}^n &\longrightarrow \mathbb{R} \\ \mathbf{x} = (x_1, x_2, \dots, x_n)^T &\longmapsto \mathbf{x} \bullet A\mathbf{x} = \mathbf{x}^T A\mathbf{x} = \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j \end{aligned}$$

die zu A gehörige quadratische Form.

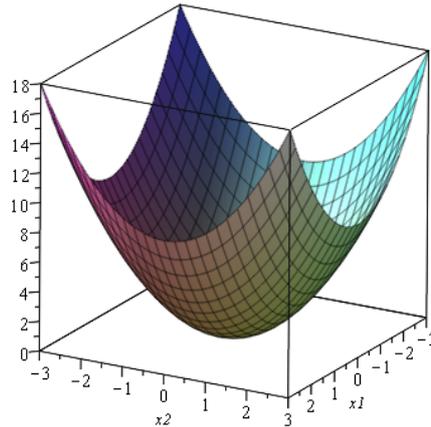
Bemerkungen:

- $Q_A(\mathbf{x})$ ist das Skalarprodukt der Vektoren \mathbf{x} und $A\mathbf{x}$.
- Q_A ist die Funktion, die jedem Vektor \mathbf{x} das Skalarprodukt der Vektoren \mathbf{x} und $A\mathbf{x}$ zuordnet.

Beispiel 2.1

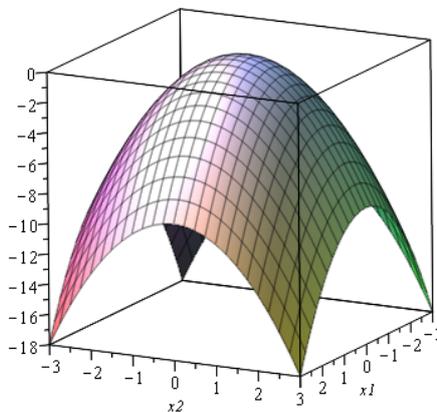
$$A = I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$Q_A(\mathbf{x}) = \mathbf{x}^T \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \mathbf{x} = (x_1, x_2) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = (x_1, x_2) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = x_1^2 + x_2^2.$$

**Beispiel 2.2**

$$B = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$$

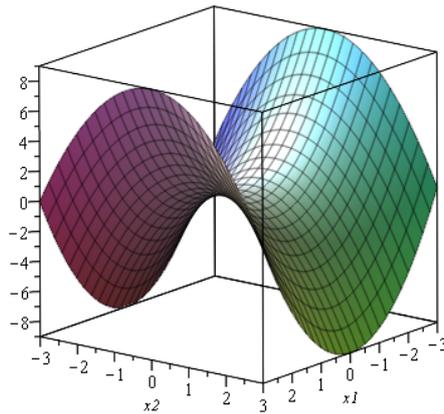
$$Q_B(\mathbf{x}) = (x_1, x_2) \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = (x_1, x_2) \begin{pmatrix} -x_1 \\ -x_2 \end{pmatrix} = -x_1^2 - x_2^2.$$



Beispiel 2.3

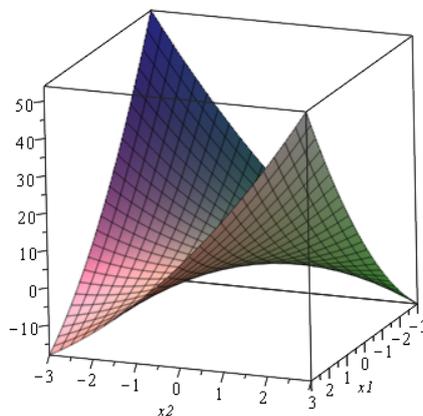
$$C = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

$$Q_C(\mathbf{x}) = (x_1, x_2) \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = (x_1, x_2) \begin{pmatrix} x_1 \\ -x_2 \end{pmatrix} = x_1^2 - x_2^2.$$

**Beispiel 2.4**

$$D = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$$

$$Q_D(\mathbf{x}) = (x_1, x_2) \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = (x_1, x_2) \begin{pmatrix} x_1 + 2x_2 \\ 2x_1 + x_2 \end{pmatrix} = x_1^2 + 4x_1x_2 + x_2^2.$$



Definition 2.8 Sei A eine (symmetrische) $(n \times n)$ -Matrix. Dann heisst A

- positiv definit, falls $Q_A(\mathbf{x}) > 0$
- positiv semidefinit, falls $Q_A(\mathbf{x}) \geq 0$
- negativ definit, falls $Q_A(\mathbf{x}) < 0$ (oder falls $-A$ positiv definit ist)
- negativ semidefinit, falls $Q_A(\mathbf{x}) \leq 0$ (oder falls $-A$ positiv semidefinit ist)

für alle Vektoren $\mathbf{x} \neq \mathbf{0}$ gilt. Die Matrix A heisst indefinit, wenn es sowohl Vektoren \mathbf{x} mit $Q_A(\mathbf{x}) > 0$ als auch Vektoren \mathbf{y} mit $Q_A(\mathbf{y}) < 0$ gibt.

Konkave und konvexe Funktionen

Zunächst beachten wir wieder, dass für alle $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ durch

$$G(t) = \mathbf{a} + t(\mathbf{b} - \mathbf{a}) = \mathbf{a} + t\mathbf{b} - t\mathbf{a} = (1-t)\mathbf{a} + t\mathbf{b}$$

die Gerade im \mathbb{R}^n durch die beiden Punkte \mathbf{a} und \mathbf{b} gegeben ist. Schränken wir den Parameter t noch auf das Intervall $[0, 1]$ ein, haben wir damit eine Beschreibung der Verbindungsstrecke der beiden Punkte.

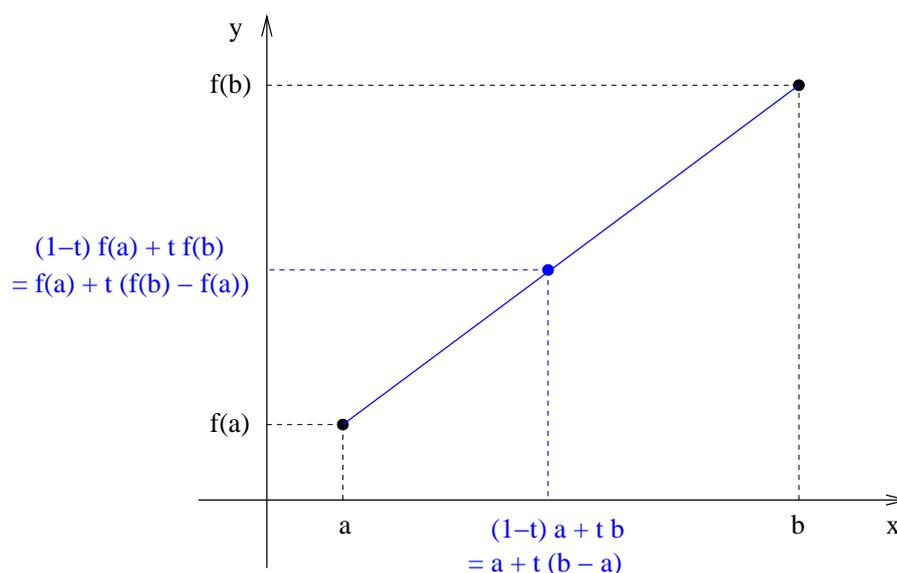
Für eine lineare Funktion

$$f(\mathbf{x}) = c_1x_1 + c_2x_2 + \cdots + c_nx_n = \mathbf{c} \bullet \mathbf{x}$$

mit $c_i \in \mathbb{R}$ gilt für $t \in (0, 1)$:

$$f((1-t)\mathbf{a} + t\mathbf{b}) = \mathbf{c} \bullet [(1-t)\mathbf{a} + t\mathbf{b}] = (1-t)\mathbf{c} \bullet \mathbf{a} + t\mathbf{c} \bullet \mathbf{b} = (1-t)f(\mathbf{a}) + tf(\mathbf{b})$$

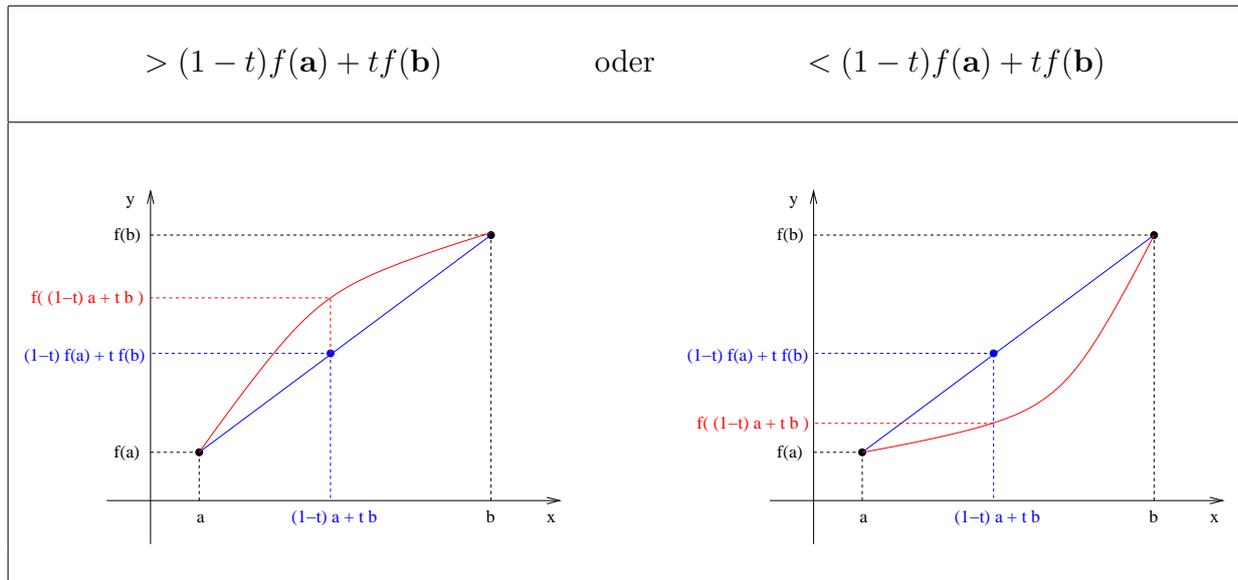
Diese Gleichung stimmt für lineare Funktionen und der Punkt $(1-t)f(\mathbf{a}) + tf(\mathbf{b}) = f(\mathbf{a}) + t(f(\mathbf{b}) - f(\mathbf{a}))$ liegt auf der Strecke, die $(\mathbf{a}, f(\mathbf{a}))^T$ mit $(\mathbf{b}, f(\mathbf{b}))^T$ im \mathbb{R}^{n+1} verbindet.



Für nichtlineare Funktionen f gilt im allgemeinen natürlich

$$f((1-t)\mathbf{a} + t\mathbf{b}) \neq (1-t)f(\mathbf{a}) + tf(\mathbf{b}).$$

d.h. die Zahl $f((1-t)\mathbf{a} + t\mathbf{b})$ ist



Definition 2.9 Sei $D \subset \mathbb{R}^n$ eine konvexe Menge. Dann heisst eine Funktion $f : D \rightarrow \mathbb{R}$

- (streng) konkav auf D , falls für alle $\mathbf{a}, \mathbf{b} \in D$ und alle $t \in (0, 1)$ gilt:

$$f((1-t)\mathbf{a} + t\mathbf{b}) \quad (>) \geq (1-t)f(\mathbf{a}) + tf(\mathbf{b})$$

- (streng) konvex auf D , falls für alle $\mathbf{a}, \mathbf{b} \in D$ und alle $t \in (0, 1)$ gilt:

$$f((1-t)\mathbf{a} + t\mathbf{b}) \quad (<) \leq (1-t)f(\mathbf{a}) + tf(\mathbf{b})$$

2.3 Gradient und Hessematrix

Definition 2.10 Sei $\mathbf{a} = (a_1, a_2, \dots, a_n) \in D \subset \mathbb{R}^n$. Der Vektor

$$\mathbf{grad} f(\mathbf{a}) = \begin{pmatrix} f_{x_1}(\mathbf{a}) \\ f_{x_2}(\mathbf{a}) \\ \vdots \\ f_{x_n}(\mathbf{a}) \end{pmatrix}$$

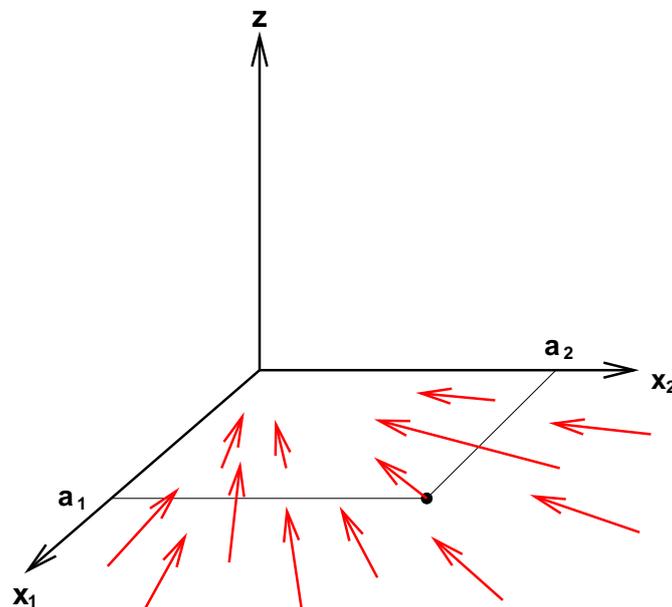
heißt der Gradient der Funktion f im Punkt \mathbf{a} .

Die Matrix

$$H_f(\mathbf{a}) = \begin{pmatrix} f_{x_1x_1}(\mathbf{a}) & f_{x_1x_2}(\mathbf{a}) & \dots & f_{x_1x_n}(\mathbf{a}) \\ f_{x_2x_1}(\mathbf{a}) & f_{x_2x_2}(\mathbf{a}) & \dots & f_{x_2x_n}(\mathbf{a}) \\ \vdots & \vdots & \vdots & \vdots \\ f_{x_nx_1}(\mathbf{a}) & f_{x_nx_2}(\mathbf{a}) & \dots & f_{x_nx_n}(\mathbf{a}) \end{pmatrix}$$

heißt die Hessematrix von f im Punkt \mathbf{a} .

Die Funktion $\mathbf{a} \mapsto \mathbf{grad} f(\mathbf{a})$ ist eine vektorwertige Funktion (Vektorfeld): Jedem $\mathbf{a} \in D$ wird ein Vektor $\mathbf{grad} f(\mathbf{a}) \in \mathbb{R}^n$ zugeordnet. Man kann das recht gut visualisieren, wenn man in jedem Punkt $\mathbf{a} \in D$ den Vektor $\mathbf{grad} f(\mathbf{a})$ anhängt.



Satz 2 (Eigenschaften des Gradienten)

- Der Gradient von f steht im Punkt \mathbf{a} orthogonal zur Niveaumenge $f(\mathbf{x}) = f(\mathbf{a})$.
- Der Gradient von f im Punkt \mathbf{a} zeigt in die Richtung des (lokal) maximalen Anstieges von f .

Eigenschaften des Gradienten am Beispiel einer Funktion $f(x_1, x_2)$

Sei $\mathbf{a} = (a_1, a_2)^T \in D$ ein Punkt. Wir betrachten eine sehr kleine Umgebung dieses Punktes und fragen, in welcher Richtung wir \mathbf{a} verlassen müssen, damit der Funktionswert von f maximal wächst. Das totale Differential von f in einem Punkt \mathbf{a} ist gegeben durch

$$df = f_{x_1}(\mathbf{a}) \underbrace{dx_1}_{x_1 - a_1} + f_{x_2}(\mathbf{a}) \underbrace{dx_2}_{x_2 - a_2} = (\mathbf{x} - \mathbf{a}) \bullet \mathbf{grad}f(\mathbf{a})$$

• gegeben:

– Niveaulinie $\{\mathbf{x} : f(\mathbf{x}) = f(\mathbf{a}) = \text{konstant}\}$

Für die Steigung dieser Niveaulinie in \mathbf{a} gilt dann

$$\frac{dx_2}{dx_1} = -\frac{f_{x_1}(\mathbf{a})}{f_{x_2}(\mathbf{a})} \quad \text{oder} \quad f_{x_2}(\mathbf{a}) dx_2 = -f_{x_1}(\mathbf{a}) dx_1$$

– Tangente T an diese Niveaulinie

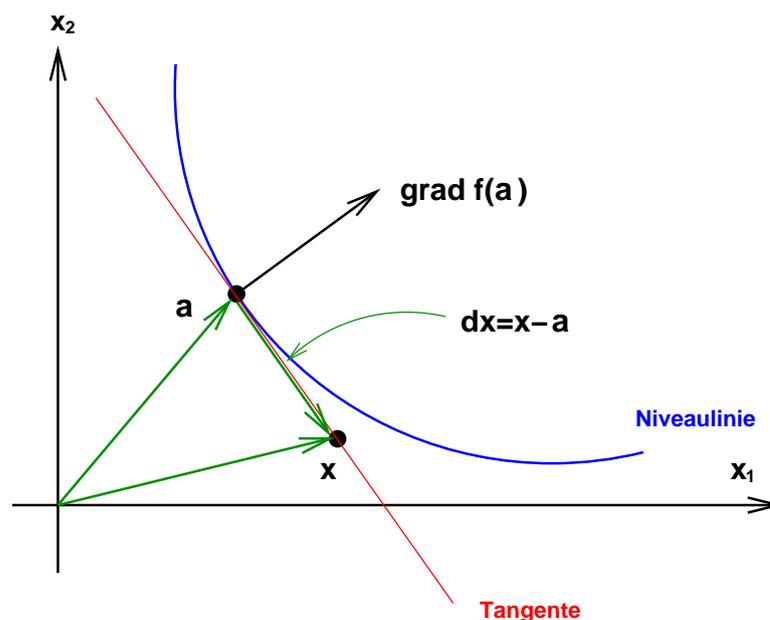
– \mathbf{x} ein beliebiger Punkt **auf** T

– $d\mathbf{x} = \mathbf{x} - \mathbf{a}$

Dann gilt:

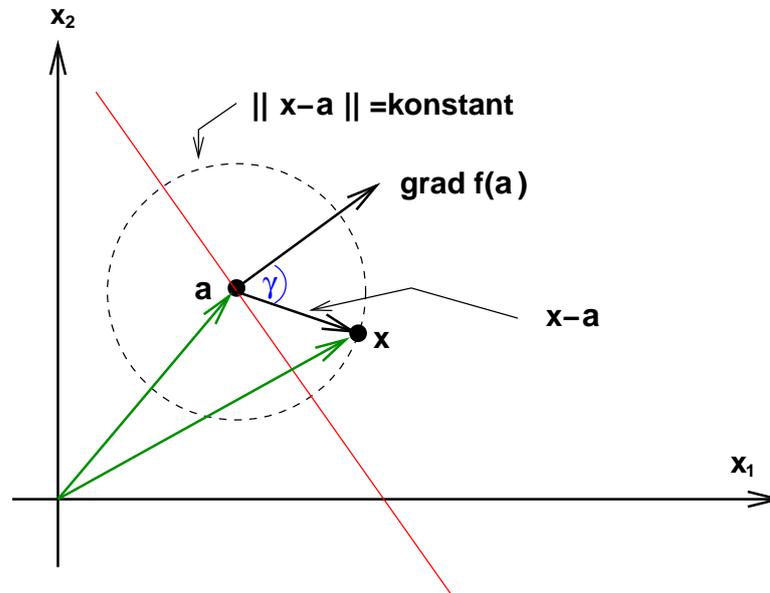
$$(\mathbf{x} - \mathbf{a}) \bullet \mathbf{grad}f(\mathbf{a}) = f_{x_1}(\mathbf{a}) dx_1 + f_{x_2}(\mathbf{a}) dx_2 = f_{x_1}(\mathbf{a}) dx_1 - f_{x_1}(\mathbf{a}) dx_1 = 0$$

$$T \perp \mathbf{grad}f(\mathbf{a})$$



- Für einen **beliebigen** Punkt \mathbf{x} sehr nahe bei \mathbf{a} gilt:

$$(\Delta f \approx) df = (\mathbf{x} - \mathbf{a}) \bullet \mathbf{grad}f(\mathbf{a}) = \|\mathbf{x} - \mathbf{a}\| \|\mathbf{grad}f(\mathbf{a})\| \cos(\gamma)$$



Wir betrachten nun

$$\frac{df}{\|\mathbf{x} - \mathbf{a}\|} = \|\mathbf{grad}f(\mathbf{a})\| \cos(\gamma)$$

und dieser Ausdruck ist

maximal	$\gamma = 0$	$\mathbf{x} - \mathbf{a} = \underbrace{\lambda}_{>0} \mathbf{grad}f(\mathbf{a})$ Stärkste Zunahme
minimal	$\gamma = \pi$	$\mathbf{x} - \mathbf{a} = \underbrace{\mu}_{<0} \mathbf{grad}f(\mathbf{a})$ Stärkste Abnahme
0	$\gamma = \pm \frac{\pi}{2}$	Tangente an Niveaulinie

Beispiel 2.5 Für die Funktion $f(x_1, x_2) = x_1^2 - x_1x_2 + x_2^2$ gilt zunächst

$$\mathbf{grad}f(x_1, x_2) = \begin{pmatrix} 2x_1 - x_2 \\ -x_1 + 2x_2 \end{pmatrix}$$

An der Stelle $(1, 2)^T$ ist somit $\begin{pmatrix} 0 \\ 3 \end{pmatrix}$ die Richtung des lokal stärksten Anstieges von f .

2.4 Die allgemeine Richtungsableitung

Die partiellen Ableitungen $\frac{\partial f(\mathbf{a})}{\partial x_i}$ geben die momentane Änderung der Funktionswerte von f wieder, wenn wir uns vom Punkt \mathbf{a} in Richtung \mathbf{e}_i fort bewegen. Es scheint unvernünftig, sich nur auf diese n Richtungen einzuschränken.

Definition 2.11 Sei $\mathbf{v} \in \mathbb{R}^n$ nicht der Nullvektor. Dann heisst der Grenzwert (falls existent)

$$\partial_{\mathbf{v}} f(\mathbf{a}) = \left. \frac{d}{dt} f(\mathbf{a} + t\mathbf{v}) \right|_{t=0} = \lim_{t \rightarrow 0} \frac{f(\mathbf{a} + t\mathbf{v}) - f(\mathbf{a})}{t}$$

die Ableitung von f an der Stelle \mathbf{a} längs \mathbf{v} .

Ist \mathbf{v} ein Einheitsvektor, dann heisst $\partial_{\mathbf{v}} f(\mathbf{a})$ die Richtungsableitung von f an der Stelle \mathbf{a} in Richtung \mathbf{v} .

Alternative Bezeichnungen:

$$\partial_{\mathbf{v}} f(\mathbf{a}) = \frac{\partial f}{\partial \mathbf{v}}(\mathbf{a}) = D_{\mathbf{v}} f(\mathbf{a})$$

Satz 3 Für jede auf einer offenen Menge D total differenzierbare Funktion f und jeden Vektor $\mathbf{v} \neq \mathbf{0}$ gilt:

$$\partial_{\mathbf{v}} f(\mathbf{a}) = \mathbf{grad} f(\mathbf{a}) \bullet \mathbf{v} = \sum_{i=1}^n f_{x_i}(\mathbf{a}) v_i$$

Beweisskizze:

f ist total differenzierbar in \mathbf{a} , also gilt

$$f(\mathbf{x}) = f(\mathbf{a}) + \mathbf{grad} f(\mathbf{a}) \bullet (\mathbf{x} - \mathbf{a}) + R(\mathbf{x}, \mathbf{a}) \quad \text{und} \quad \lim_{\mathbf{x} \rightarrow \mathbf{a}} \frac{R(\mathbf{x}, \mathbf{a})}{\|\mathbf{x} - \mathbf{a}\|} = 0$$

Setze $\mathbf{x} = \mathbf{a} + t\mathbf{v}$ und wir erhalten:

$$f(\mathbf{x}) - f(\mathbf{a}) = f(\mathbf{a} + t\mathbf{v}) - f(\mathbf{a}) = \mathbf{grad} f(\mathbf{a}) \bullet t\mathbf{v} + R(\mathbf{x}, \mathbf{a}).$$

Somit

$$\begin{aligned} \partial_{\mathbf{v}} f(\mathbf{a}) &= \lim_{t \rightarrow 0} \frac{f(\mathbf{a} + t\mathbf{v}) - f(\mathbf{a})}{t} \\ &= \lim_{t \rightarrow 0} \frac{\mathbf{grad} f(\mathbf{a}) \bullet t\mathbf{v} + R(\mathbf{x}, \mathbf{a})}{t} \\ &= \mathbf{grad} f(\mathbf{a}) \bullet \mathbf{v} + \lim_{t \rightarrow 0} \frac{R(\mathbf{x}, \mathbf{a})}{t} \\ &= \mathbf{grad} f(\mathbf{a}) \bullet \mathbf{v}. \end{aligned}$$

□

Für einige wichtige numerische Verfahren der Optimierung benötigen wir noch den Begriff der Aufstiegsrichtung einer Funktion in einem speziellen Punkt.

Definition 2.12 *Ein Vektor $\mathbf{v} \in \mathbb{R}^n$ mit $\mathbf{v} \neq \mathbf{0}$ und meist $\|\mathbf{v}\| = 1$ heisst Aufstiegsrichtung von f im Punkt \mathbf{a} , falls folgendes gilt:*

$$\partial_{\mathbf{v}} f(\mathbf{a}) = \mathbf{grad} f(\mathbf{a}) \bullet \mathbf{v} > 0.$$

Wir wissen bereits, dass der Gradient $\mathbf{grad} f(\mathbf{a})$ stets in die Richtung des stärksten Anstieges von f im Punkt \mathbf{a} zeigt. Das könnte man auch wie folgt ausdrücken:

Sei $f : \mathbb{R}^n \rightarrow \mathbb{R}$ stetig differenzierbar mit $\mathbf{grad} f(\mathbf{a}) \neq \mathbf{0}$. Dann ist der Einheitsvektor

$$\mathbf{v}^* = \frac{\mathbf{grad} f(\mathbf{a})}{\|\mathbf{grad} f(\mathbf{a})\|}$$

die eindeutige Lösung des Maximierungsproblems

$$\max_{\|\mathbf{v}\|=1} \mathbf{grad} f(\mathbf{a}) \bullet \mathbf{v}.$$

2.5 Die Taylor-Formel

Taylor-Polynome sind wichtige Hilfsmittel, um komplizierte Funktionen in der Nähe von interessanten Punkten durch einfache Funktionen zu approximieren. Wir werden uns in diesem Abschnitt auf quadratische Taylor-Polynome einschränken, aber das Verfahren lässt sich bei Bedarf verallgemeinern.

Satz 4 Sei $D \subset \mathbb{R}^n$ konvex, $\mathbf{a}, \mathbf{x} \in D$ und $f : D \rightarrow \mathbb{R}$ eine mindestens dreimal stetig differenzierbare Funktion. Das 2-te Taylor-Polynom der Funktion f in \mathbf{a} ist dann:

$$t_2(\mathbf{x}) = f(\mathbf{a}) + \mathbf{grad}f(\mathbf{a}) \bullet (\mathbf{x} - \mathbf{a}) + \frac{1}{2} (\mathbf{x} - \mathbf{a})^T H_f(\mathbf{a}) (\mathbf{x} - \mathbf{a})$$

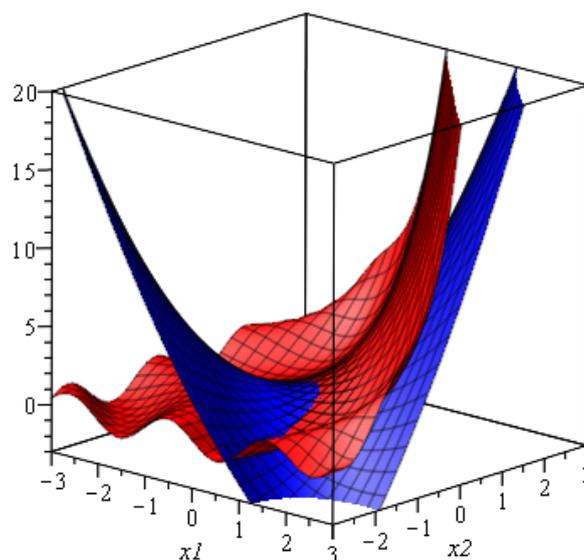
Für den Fehler, den man bei dieser Approximation macht, gilt:

$$\lim_{\mathbf{x} \rightarrow \mathbf{a}} \frac{f(\mathbf{x}) - t_2(\mathbf{x})}{\|\mathbf{x} - \mathbf{a}\|^2} = 0$$

Beispiel 2.6 Bestimmen Sie die quadratische Approximation der Funktion $f(x_1, x_2) = e^{x_1+x_2} + \sin(x_1x_2)$ an der Stelle $\mathbf{a} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$.

$$t_2(\mathbf{x}) = 1 + (1, 1) \mathbf{x} + \mathbf{x}^T \frac{1}{2} \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \mathbf{x} = 1 + x_1 + x_2 + \frac{1}{2}x_1^2 + 2x_1x_2 + \frac{1}{2}x_2^2$$

In der Skizze sehen Sie den Graph von f (rot) und Graph von P (blau):



2.6 Lokale Extremalstellen im Inneren

Definition 2.13 Sei $f : D \rightarrow \mathbb{R}$ eine Funktion in n Variablen und $\mathbf{a} \in D$. Dann heisst \mathbf{a} lokale Maximalstelle bzw. lokale Minimalstelle von f , wenn es eine ϵ -Umgebung $U_\epsilon(\mathbf{a})$ gibt, so dass für alle $x \in U_\epsilon(\mathbf{a}) \cap D$ gilt

$$f(\mathbf{x}) \leq f(\mathbf{a}) \quad \text{bzw.} \quad f(\mathbf{x}) \geq f(\mathbf{a})$$

Gelten diese Ungleichungen jeweils für **alle** $\mathbf{x} \in D$, dann heisst \mathbf{a} eine globale Maximalstelle bzw. globale Minimalstelle von f .

Die Bezeichnungen strikte Maximalstelle bzw. Minimalstelle wird verwendet, wenn \leq bzw. \geq durch $<$ bzw. $>$ ersetzt werden kann.

Die Zahl $f(\mathbf{a})$ heisst dann (fallweise) lokales bzw. globales Maximum bzw. Minimum.

Satz 5 (Lokale Extrema im Inneren von D) Sei $f : D \rightarrow \mathbb{R}$ eine stetig differenzierbare Funktion in n Variablen und $\mathbf{a} \in D$ ein Punkt im Inneren von D (d.h. es existiert eine ϵ -Umgebung $U_\epsilon(\mathbf{a})$ von \mathbf{a} mit $U_\epsilon(\mathbf{a}) \subset D$). Dann gilt:

$$\mathbf{a} \text{ ist lokale Extremalstelle von } f \quad \implies \quad \mathbf{grad}f(\mathbf{a}) = \mathbf{0}$$

Beweisidee:

Nach Voraussetzung besitzen die reellen partiellen Funktionen $h_i(t) := f(\mathbf{a} + t\mathbf{e}_i)$ für $1 \leq i \leq n$ in $t = 0$ eine (innere) lokale Extremalstelle. Also gilt:

$$0 = \left. \frac{d}{dt} h_i(t) \right|_{t=0} = \frac{\partial f}{\partial x_i}(\mathbf{a})$$

□

Definition 2.14 Ein Punkt $\mathbf{a} \in \mathbb{R}^n$ heisst stationärer Punkt von f , wenn $\mathbf{grad}f(\mathbf{a}) = \mathbf{0}$ gilt. Ein stationärer Punkt, der keine Extremalstelle ist, heisst Sattelpunkt.

Der Satz besagt also, dass man die inneren lokalen Extremalstellen unter den stationären Punkten zu suchen hat, dass aber nicht jeder stationäre Punkt eine Extremalstelle sein muss. Zur Identifizierung ist oft das folgende Resultat geeignet.

Satz 6 Sei

- $D \subset \mathbb{R}^n$,
- $f : D \rightarrow \mathbb{R}$ eine zweimal stetig differenzierbare Funktion,
- $\mathbf{a} \in D$ ein innerer Punkt.

Dann gilt:

$$\begin{aligned} \mathbf{grad}f(\mathbf{a}) = \mathbf{0} \text{ und } H_f(\mathbf{a}) \text{ positiv definit} &\implies \mathbf{a} \text{ ist (strikte) lokale Minimalstelle} \\ \mathbf{grad}f(\mathbf{a}) = \mathbf{0} \text{ und } H_f(\mathbf{a}) \text{ negativ definit} &\implies \mathbf{a} \text{ ist (strikte) lokale Maximalstelle} \\ \mathbf{grad}f(\mathbf{a}) = \mathbf{0} \text{ und } H_f(\mathbf{a}) \text{ indefinit} &\implies \mathbf{a} \text{ ist Sattelpunkt} \end{aligned}$$

Beweisidee:

Wir benutzen das 2-te Taylor-Polynom von f in \mathbf{a} und beachten, dass \mathbf{a} ein stationärer Punkt ist, also $\mathbf{grad}f(\mathbf{a}) = \mathbf{0}$ gilt. In einer kleinen Umgebung U von \mathbf{a} gilt

$$\begin{aligned} f(\mathbf{x}) - f(\mathbf{a}) &\approx t_2(\mathbf{x}) - f(\mathbf{a}) \\ &= \mathbf{grad}f(\mathbf{a}) \bullet (\mathbf{x} - \mathbf{a}) + \frac{1}{2} (\mathbf{x} - \mathbf{a})^T H_f(\mathbf{a}) (\mathbf{x} - \mathbf{a}) \\ &= \frac{1}{2} (\mathbf{x} - \mathbf{a})^T H_f(\mathbf{a}) (\mathbf{x} - \mathbf{a}) \end{aligned}$$

Da alle zweiten partiellen Ableitungen stetig sind, ändert sich auch $H_f(\mathbf{a})$ stetig in allen Variablen. Insbesondere können wir die Umgebung U so wählen, dass die Matrix $H_f(\mathbf{a})$ für alle $\mathbf{x} \in U$ ihre Definitheit

- positiv definit,
- negativ definit bzw.
- indefinit

nicht ändert. In dieser Umgebung U von \mathbf{a} ist dann die rechte Seite der obigen Gleichung stets

- positiv,
- negativ bzw.
- positiv oder negativ.

Das bedeutet für alle $\mathbf{x} \in U$ somit

- $f(\mathbf{x}) - f(\mathbf{a}) \geq 0$ oder $f(\mathbf{x}) \geq f(\mathbf{a})$ und \mathbf{a} ist lokale Minimalstelle
- $f(\mathbf{x}) - f(\mathbf{a}) \leq 0$ oder $f(\mathbf{x}) \leq f(\mathbf{a})$ und \mathbf{a} ist lokale Maximalstelle bzw.
- $f(\mathbf{x}) - f(\mathbf{a})$ ist sowohl positiv als auch negativ und \mathbf{a} ist Sattelpunkt.

□

Spezialfall $n = 2$ Sei $\mathbf{a} = (a_1, a_2)^T$ ein stationärer Punkt der zweimal stetig differenzierbaren Funktion f . Für die Hessematrix gilt dann

$$H_f(\mathbf{a}) = \begin{pmatrix} f_{x_1x_1}(\mathbf{a}) & f_{x_1x_2}(\mathbf{a}) \\ f_{x_2x_1}(\mathbf{a}) & f_{x_2x_2}(\mathbf{a}) \end{pmatrix} \longrightarrow \det H_f(\mathbf{a}) = f_{x_1x_1}(\mathbf{a})f_{x_2x_2}(\mathbf{a}) - f_{x_1x_2}(\mathbf{a})^2$$

und der obige lokale Extremalstellentest lautet:

$$\det H_f(\mathbf{a}) > 0 \text{ und } f_{x_1x_1}(\mathbf{a}) > 0 \implies \mathbf{a} \text{ ist lokale Minimalstelle}$$

$$\det H_f(\mathbf{a}) > 0 \text{ und } f_{x_1x_1}(\mathbf{a}) < 0 \implies \mathbf{a} \text{ ist lokale Maximalstelle}$$

$$\det H_f(\mathbf{a}) < 0 \implies \mathbf{a} \text{ ist Sattelpunkt}$$

3 Lösungswege

3.1 Das allgemeine Maximierungsproblemen

Zunächst schränken wir uns ab sofort (und ohne Einschränkung der Allgemeinheit) auf **Maximierungsprobleme** ein. Offensichtlich hängt der zu wählende (effizienteste) Lösungsweg eines solchen Problems sowohl von der Zielfunktion f als auch von der Struktur der Menge S ab. Hier einige wesentliche Eigenschaften, die f bzw. S haben könnten:

f könnte

- stetig (auf ganz S),
- differenzierbar (auf ganz S) und sogar
- konkav bzw. konvex (auf ganz S),
- rein quadratisch oder
- linear sein.

S könnte

- zusammenhängend,
- beschränkt,
- abgeschlossen,
- beschränkt und abgeschlossen (d.h. kompakt) und
- konvex sein.

Ab sofort soll die Menge S stets als Lösungsmenge eines Systems von m Ungleichungen beschrieben werden können. Somit lässt sich das allgemeine Problem wie folgt darstellen:

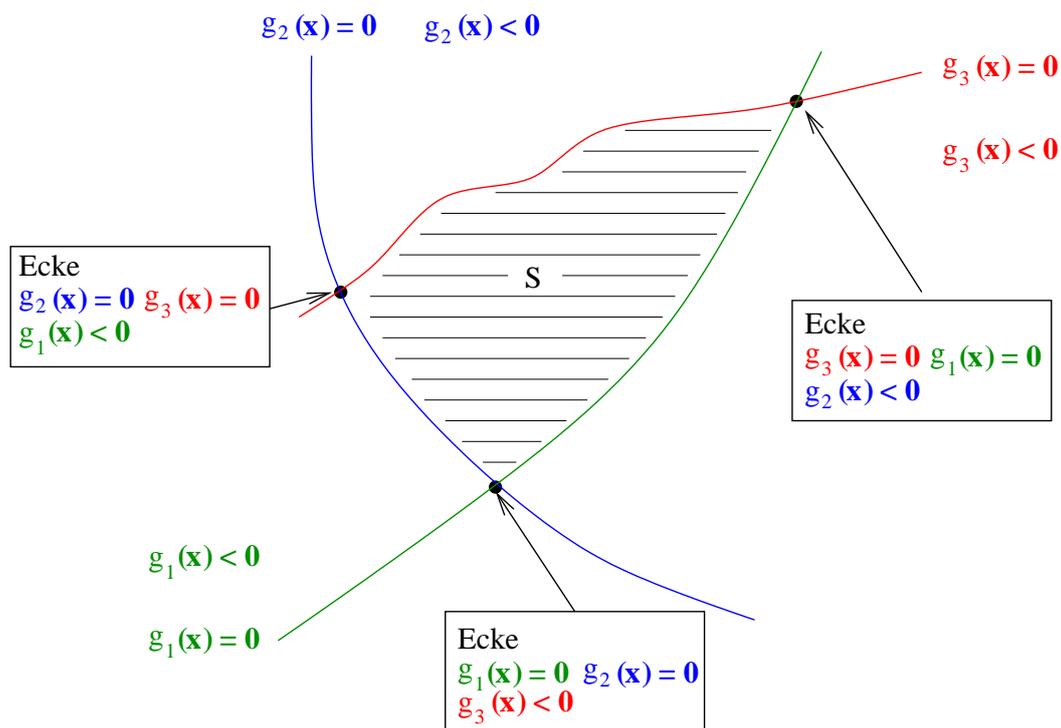
zu maximierende Zielfunktion : $y = f(x_1, x_2, \dots, x_n) = f(\mathbf{x})$

$$\text{Nebenbedingungen} : \begin{cases} g_1(x_1, x_2, \dots, x_n) = g_1(\mathbf{x}) \leq c_1 \\ g_2(x_1, x_2, \dots, x_n) = g_2(\mathbf{x}) \leq c_2 \\ \dots \\ g_m(x_1, x_2, \dots, x_n) = g_m(\mathbf{x}) \leq c_m \end{cases}$$

Die durch die Nebenbedingungen definierte Menge

$$S = \{ \mathbf{x} \in \mathbb{R}^n \mid g_1(\mathbf{x}) \leq c_1, g_2(\mathbf{x}) \leq c_2, \dots, g_m(\mathbf{x}) \leq c_m \}$$

wird auch zulässige Menge genannt. Für $n = 2$ kann man diese Mengen meist recht schnell skizzieren, aber schon in diesem einfachen Fall können diese Gebilde recht kompliziert aussehen.



Ganz allgemein kann man sagen, dass als Kandidaten für mögliche globale Extremalstellen folgende Punkte in Frage kommen:

1. die Ecken von S (falls vorhanden). Das sind die Lösungen von möglichen Kombinationen

$$\begin{cases} g_i(\mathbf{x}) = g_j(\mathbf{x}) = \dots = g_k(\mathbf{x}) = 0 \\ g_l(\mathbf{x}) < c_l \quad \text{sonst} \end{cases} .$$

2. Die Extremalstellen auf den berandenden Kurven-, Flächen- bzw. Hyperflächenstücken von S .
3. Die stationären Stellen von f im Inneren von S

$$\text{Int}(S) = \{\mathbf{x} \in \mathbb{R}^n \mid g_1(\mathbf{x}) < c_1, g_2(\mathbf{x}) < c_2, \dots, g_m(\mathbf{x}) < c_m\}.$$

4. Die Punkte in S , in denen f nicht differenzierbar ist.

Der grösste bzw. kleinste Funktionswert an allen diesen Punkten liefert das globale Maximum bzw. Minimum von f auf S .

Wir werden uns deshalb ab sofort auch auf das Auffinden der (möglichst aller) lokalen Extremalstellen konzentrieren. Dabei stellen wir uns auf den Standpunkt, dass das Auffinden der globalen Extremalstelle dann nur noch auf dem simplen Vergleich der zugehörigen Funktionswerte beruht.

Die Bestimmung der Ecken und die Untersuchung der Randstücke von S ist für $n > 2$ im allgemeinen aber sehr schwierig oder unmöglich.

Beispiel 3.1 Das Optimierungsproblem hat die folgende allgemeine Gestalt:

$$\text{Zielfunktion} \quad : y = f(x_1, x_2) = x_1^2 x_2 (4 - x_1 - x_2)$$

$$\text{Nebenbedingung} \quad : \begin{cases} g_1(x_1, x_2) = -x_2 \leq 0 \\ g_2(x_1, x_2) = -x_1 \leq 0 \\ g_3(x_1, x_2) = x_1 + x_2 \leq 6 \end{cases}$$

1. Ecken: $(0, 0)$, $(6, 0)$ und $(0, 6)$

2. Extremalstellen auf den Randkurven

- Randstück $x_2 = 0$

$$\text{Zielfunktion: } f(x_1, 0) = 0$$

- Randstück $x_1 = 0$

$$\text{Zielfunktion: } f(0, x_2) = 0$$

- Randstück $x_1 + x_2 = 6$

$$\text{Zielfunktion: } h(x_1) = f(x_1, 6 - x_1) = 2x_1^3 - 12x_1^2$$

$$h'(x_1) = 0 \text{ lösen: } x_1 = 0 \text{ oder } x_1 = 4$$

$$\text{Kandidaten: } (0, 6) \text{ und } (4, 2)$$

3. Stationäre Stellen im Inneren

$$0 = f_{x_1} = x_1 x_2 (8 - 3x_1 - 2x_2)$$

$$0 = f_{x_2} = x_1^2 (4 - x_1 - 2x_2)$$

$$\text{Kandidaten: } (2, 1)$$

Kandidat \mathbf{x}	$f(\mathbf{x})$	
$(0, 6)$	0	
$(6, 0)$	0	
$(0, 0)$	0	
$(4, 2)$	-64	globales Minimum
$(2, 1)$	4	globales Maximum

Die hier vorgestellte Methode kann man wohl als „klassisch“ bezeichnen. Seit den 50er Jahren des vergangenen Jahrhunderts haben Ökonomen allerdings solche Probleme gelöst, indem sie die Methode der Lagrange-Multiplikatoren verallgemeinert haben. Insbesondere H. W. **Kuhn** und A. W. **Tucker** haben diese Methode entwickelt.

Definition 3.1 Die Funktion (in $n + m$ Variablen)

$$L(x_1, \dots, x_n, \lambda_1, \dots, \lambda_m) = f(x_1, x_2, \dots, x_n) - \sum_{j=1}^m \lambda_j (g_j(x_1, x_2, \dots, x_n) - c_j)$$

oder kurz

$$L(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) - \sum_{j=1}^m \lambda_j (g_j(\mathbf{x}) - c_j)$$

heißt Lagrange-Funktion des Optimierungsproblems.

3.2 f, g stet. dfb, $g(\mathbf{x}) = c$, rang Dg maximal \longrightarrow Lagrange-Verfahren

3.2.1 n Variablen und eine Nebenbedingung

Das Optimierungsproblem hat die folgende allgemeine Gestalt:

$$\begin{aligned} \text{Zielfunktion} & : y = f(x_1, x_2, \dots, x_n) \\ \text{Nebenbedingung} & : g(x_1, x_2, \dots, x_n) = c \end{aligned}$$

Satz 7 Zu jeder Lösung \mathbf{x}^* des obigen Extremalwertproblems mit stetig differenzierbaren Funktionen f und ϕ und $\mathbf{grad} \phi(\mathbf{x}^*) \neq \mathbf{0}$ gibt es eine Zahl $\lambda^* \in \mathbb{R}$, den so genannten Lagrange-Multiplikator, so dass

$$\mathbf{grad} f(\mathbf{x}^*) = \lambda^* \mathbf{grad} \phi(\mathbf{x}^*)$$

Beweis:

Sei $\mathbf{x}^* = (x_1^*, x_2^*, \dots, x_n^*)$. Wegen $\mathbf{grad} \phi(\mathbf{x}^*) \neq \mathbf{0}$ können wir (nach eventueller Ummummerierung) $\phi_{x_n}(\mathbf{x}^*) \neq 0$ annehmen. Nach dem Satz über implizite Funktionen kann man deshalb die Nebenbedingung $\phi(x_1, x_2, \dots, x_n) = 0$ lokal um \mathbf{x}^* nach x_n auflösen: $x_n = h(x_1, x_2, \dots, x_{n-1})$.

Die Voraussetzung besagt nun, dass die Funktion

$$f(x_1, x_2, \dots, x_{n-1}, h(x_1, x_2, \dots, x_{n-1}))$$

in $(x_1^*, x_2^*, \dots, x_{n-1}^*)$ ein lokales Extremum besitzt. Folglich gilt für alle $1 \leq i \leq n-1$ (Kettenregel und implizites differenzieren):

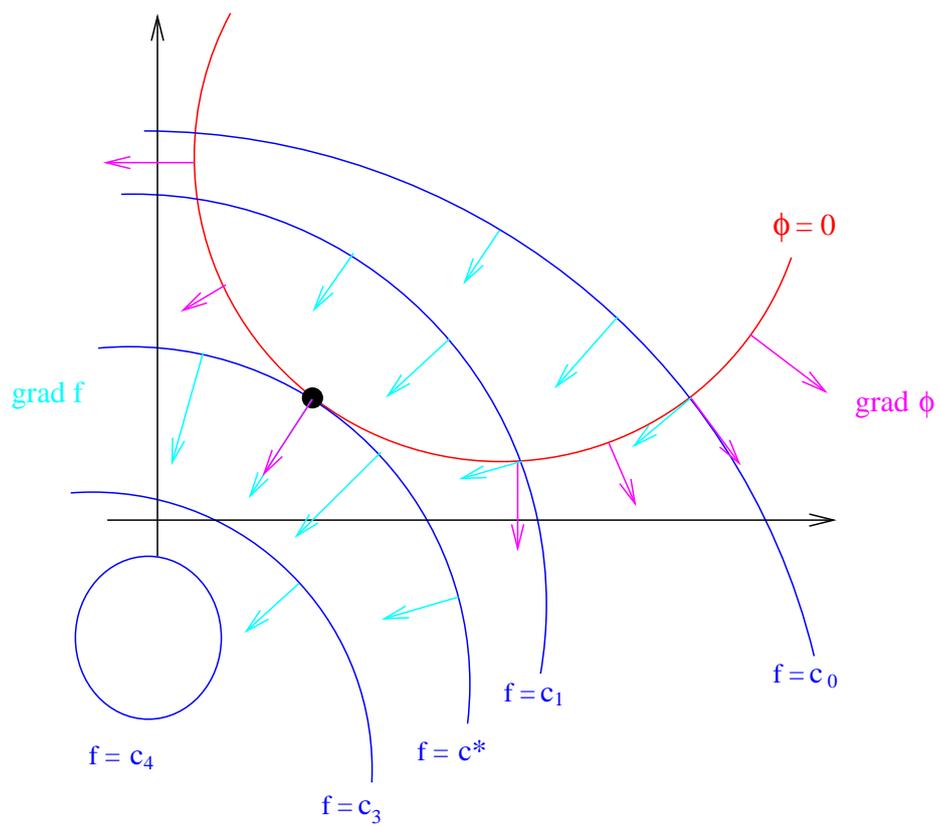
$$\begin{aligned} 0 &= \left. \frac{\partial}{\partial x_i} f(x_1, x_2, \dots, x_{n-1}, h(x_1, x_2, \dots, x_{n-1})) \right|_{\mathbf{x}=\mathbf{x}^*} \\ &= \sum_{j=1}^{n-1} f_{x_j}(\mathbf{x}^*) \frac{\partial x_j}{\partial x_i} + f_{x_n}(\mathbf{x}^*) \frac{\partial h}{\partial x_i} \\ &= f_{x_i}(\mathbf{x}^*) + f_{x_n}(\mathbf{x}^*) \cdot \left(-\frac{\phi_{x_i}(\mathbf{x}^*)}{\phi_{x_n}(\mathbf{x}^*)} \right) \\ &= f_{x_i}(\mathbf{x}^*) + \phi_{x_i}(\mathbf{x}^*) \cdot \left(-\frac{f_{x_n}(\mathbf{x}^*)}{\phi_{x_n}(\mathbf{x}^*)} \right) \end{aligned}$$

Setzen wir nun $\lambda^* = \frac{f_{x_n}(\mathbf{x}^*)}{\phi_{x_n}(\mathbf{x}^*)}$ so folgt die Behauptung.

□

Das bedeutet, dass die beiden Gradienten von f und ϕ im Punkt \mathbf{x}^* parallel sein müssen. Im Fall $n = 2$ bedeutet dies, dass sich die Niveaulinie $f(x_1, x_2) = f(x_1^*, x_2^*)$ und die Kurve $\phi(x_1, x_2) = 0$ im Extrempunkt berühren.

Wir suchen also Punkte auf der Kurve $\phi = 0$ in denen die beiden Gradienten zu ϕ und f gleiche oder entgegengesetzte Richtung haben.



Satz 8 (Lagrange, n Variablen, eine Nebenbedingung in Gleichungsform) *Seien f und g stetig differenzierbare Funktionen und sei \mathbf{x}^* die Lösung des Optimierungsproblems*

$$\begin{aligned} \text{Zielfunktion} & : y = f(x_1, x_2, \dots, x_n) \\ \text{Nebenbedingung} & : \phi(x_1, x_2, \dots, x_n) = 0 \quad \text{oder} \quad g(x_1, x_2, \dots, x_n) = c \end{aligned}$$

Weiterhin gelte $\mathbf{grad} \phi(\mathbf{x}^*) \neq \mathbf{0}$. Dann existiert eine reelle Zahl λ^* , so dass $(\mathbf{x}^*, \lambda^*) = (x_1^*, x_2^*, \dots, x_n^*, \lambda^*)$ ein stationärer Punkt der Lagrange-Funktion ist, d.h.

$$\begin{aligned} L_{x_1}(\mathbf{x}^*, \lambda^*) &= 0 \\ &\vdots \\ L_{x_n}(\mathbf{x}^*, \lambda^*) &= 0 \\ L_\lambda(\mathbf{x}^*, \lambda^*) &= 0 \end{aligned} \quad \text{oder kurz} \quad \boxed{\mathbf{grad} L(\mathbf{x}^*, \lambda^*) = \mathbf{0}}$$

Beweisidee:

Wir wissen bereits, dass an einer möglichen Extremalstelle \mathbf{x}^* unter der gegebenen Nebenbedingung die vektorielle Gleichung $\mathbf{grad} f(\mathbf{x}^*) = \lambda^* \mathbf{grad} \phi(\mathbf{x}^*)$ oder die dazu äquivalente Gleichung $\mathbf{grad} f(\mathbf{x}^*) - \lambda^* \mathbf{grad} \phi(\mathbf{x}^*) = \mathbf{0}$ erfüllt ist. Das entspricht genau den ersten n Gleichungen $\mathbf{grad}_x L(\mathbf{x}^*, \lambda^*) = \mathbf{0}$ (wir betrachten hier nur die partiellen Ableitungen nach den Koordinaten x_i). Die letzte Gleichung $L_\lambda(\mathbf{x}^*, \lambda^*) = 0$ ist einfach die Nebenbedingung.

□

Beispiel 3.2 *Wir untersuchen das Problem*

$$\begin{aligned} \text{Zielfunktion} & f(x_1, x_2) = 8 \ln(x_1) + 2 \ln(x_2) \\ \text{Nebenbedingung} & \phi(x_1, x_2) = 2x_1 + x_2 - 5 = 0 \end{aligned}$$

Für die Lagrange-Funktion $L(x_1, x_2, \lambda)$ und deren partielle Ableitungen gilt:

$$\begin{aligned} L(x_1, x_2, \lambda) &= 8 \ln(x_1) + 2 \ln(x_2) - \lambda(2x_1 + x_2 - 5) \\ 0 = L_{x_1}(x_1, x_2, \lambda) &= 8/x_1 - 2\lambda \\ 0 = L_{x_2}(x_1, x_2, \lambda) &= 2/x_2 - \lambda \\ 0 = L_\lambda(x_1, x_2, \lambda) &= 2x_1 + x_2 - 5 \end{aligned}$$

Die mögliche Extremalstelle hat also die Koordinaten $(x_1^*, x_2^*) = (2, 1)$ mit $\lambda^* = 2$. Für die Gradienten der Zielfunktion und der Nebenbedingung gilt nun weiterhin

$$\begin{aligned} \mathbf{grad} f(x_1, x_2) &= \begin{pmatrix} f_{x_1}(x_1, x_2) \\ f_{x_2}(x_1, x_2) \end{pmatrix} = \begin{pmatrix} 8/x_1 \\ 2/x_2 \end{pmatrix} \\ \mathbf{grad} f(2, 1) &= \begin{pmatrix} 4 \\ 2 \end{pmatrix} \\ \mathbf{grad} \phi(x_1, x_2) &= \begin{pmatrix} 2 \\ 1 \end{pmatrix} \end{aligned}$$

Man erkennt, dass die Gradienten von f und ϕ im Punkt $(2, 1)$ auf einer Geraden liegen.

$$\mathbf{grad} f(2, 1) = \lambda^* \mathbf{grad} \phi(2, 1) \quad \text{oder} \quad \begin{pmatrix} 4 \\ 2 \end{pmatrix} = 2 \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

Deutung des Lagrange-Multiplikators λ^*

Wir betrachten wieder das Problem

$$\begin{aligned} \text{Zielfunktion} & : y = f(x_1, x_2, \dots, x_n) \\ \text{Nebenbedingung} & : g(x_1, x_2, \dots, x_n) = c \end{aligned}$$

wobei wir eventuell vorkommende Konstanten der Nebenbedingung auf die rechte Seite bringen. Die Lösung dieses Problems \mathbf{x}^* wird im allgemeinen von c abhängen. Es gilt also

$$\mathbf{x}^* = \mathbf{x}^*(c) = (x_1^*(c), x_2^*(c), \dots, x_n^*(c))$$

und die Optimalwertfunktion ist

$$f^*(c) = f(\mathbf{x}^*(c)) = f(x_1^*(c), x_2^*(c), \dots, x_n^*(c))$$

Satz 9 *Der Lagrange-Multiplikator ist die Rate, mit der sich der optimale Wert der Zielfunktion bezüglich der Änderung von c in der Nebenbedingung ändert.*

$$\boxed{\frac{df^*(c)}{dc} = \lambda^*(c)}$$

Beweis:

Mit der Kettenregel und dem Satz 7 folgt sofort:

$$\begin{aligned} \frac{df^*(c)}{dc} &= \mathbf{grad} f(\mathbf{x}^*(c)) \bullet \frac{d}{dc} \mathbf{x}^*(c) \\ &= \lambda^* \mathbf{grad} g(\mathbf{x}^*(c)) \bullet \frac{d}{dc} \mathbf{x}^*(c) \\ &= \lambda^* \frac{dg(\mathbf{x}^*(c))}{dc} \\ &= \lambda^* \frac{dc}{dc} = \lambda^* \end{aligned}$$

□

Ist nun insbesondere dc eine kleine Änderung von c , dann gilt:

$$\boxed{f^*(c + dc) - f^*(c) \approx df^*(c) = \lambda^*(c) dc}$$

3.2.2 n Variablen und m Nebenbedingungen

Wir betrachten ein Optimierungsproblem mit m Nebenbedingungen (in Gleichungsform)

$$\begin{aligned} \text{Zielfunktion} & : y = f(x_1, x_2, \dots, x_n) = f(\mathbf{x}) \\ \text{Nebenbedingungen} & : \begin{cases} g_1(\mathbf{x}) = g_1(x_1, x_2, \dots, x_n) = c_1 \\ g_2(\mathbf{x}) = g_2(x_1, x_2, \dots, x_n) = c_2 \\ \dots \\ g_m(\mathbf{x}) = g_m(x_1, x_2, \dots, x_n) = c_m \end{cases} \end{aligned}$$

Für die Existenz eines Lagrange-Multiplikators im letzten Kapitel mussten wir verlangen, dass der Gradient der (einzigen) Nebenbedingung nicht der Nullvektor ist. Diese Forderung müssen wir auf die m Nebenbedingungen verallgemeinern. Dazu betrachten wir die Matrix, die alle m Gradienten der Nebenbedingungen sammelt.

Definition 3.2 Seien $\mathbf{g} = (g_1, g_2, \dots, g_m)$ stetig differenzierbare Funktionen in n Variablen. Dann heißt die $m \times n$ Matrix

$$D\mathbf{g}(\mathbf{x}) = \begin{pmatrix} \frac{\partial g_1}{\partial x_1}(\mathbf{x}) & \frac{\partial g_1}{\partial x_2}(\mathbf{x}) & \dots & \frac{\partial g_1}{\partial x_n}(\mathbf{x}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial x_1}(\mathbf{x}) & \frac{\partial g_m}{\partial x_2}(\mathbf{x}) & \dots & \frac{\partial g_m}{\partial x_n}(\mathbf{x}) \end{pmatrix}$$

Jacobi-Matrix von \mathbf{g} im Punkt \mathbf{x} . Ein Punkt \mathbf{x} heißt kritischer Punkt von \mathbf{g} , falls der Rang der Matrix $D\mathbf{g}(\mathbf{x})$ nicht maximal ist.

Satz 10 (Lagrange, n Variablen, m Nebenbedingungen in Gleichungsform) Seien f und $\mathbf{g} = (g_1, g_2, \dots, g_m)$ stetig differenzierbare Funktionen und sei \mathbf{x}^* die Lösung des Optimierungsproblems

$$\begin{aligned} \text{Zielfunktion} & : y = f(\mathbf{x}) \\ \text{Nebenbedingungen} & : \begin{cases} g_1(\mathbf{x}) = c_1 \\ \dots \\ g_m(\mathbf{x}) = c_m \end{cases} \quad \text{oder kurz} \quad \mathbf{g}(\mathbf{x}) = \mathbf{c} \end{aligned}$$

Weiterhin sei \mathbf{x}^* kein kritischer Punkt von $D\mathbf{g}$, d.h. im allgemeinen $\text{rang } D\mathbf{g}(\mathbf{x}^*) = m$.

Dann existieren reelle Zahlen $\boldsymbol{\lambda}^* = (\lambda_1^*, \dots, \lambda_m^*)$, so dass $(\mathbf{x}^*, \boldsymbol{\lambda}^*) = (x_1^*, x_2^*, \dots, x_n^*, \lambda_1^*, \dots, \lambda_m^*)$ ein stationärer Punkt der Lagrange-Funktion ist, d.h.

$$L_{x_1}(\mathbf{x}^*, \boldsymbol{\lambda}^*) = 0, \dots, L_{x_n}(\mathbf{x}^*, \boldsymbol{\lambda}^*) = 0$$

und

oder kurz

$$\boxed{\text{grad } L(\mathbf{x}^*, \boldsymbol{\lambda}^*) = \mathbf{0}}$$

$$L_{\lambda_1}(\mathbf{x}^*, \boldsymbol{\lambda}^*) = 0, \dots, L_{\lambda_m}(\mathbf{x}^*, \boldsymbol{\lambda}^*) = 0$$

Beweisidee:

Die Forderungen $L_{\lambda_1}(\mathbf{x}^*, \boldsymbol{\lambda}^*) = 0, \dots, L_{\lambda_m}(\mathbf{x}^*, \boldsymbol{\lambda}^*) = 0$ sind die Nebenbedingungen, die jeder interessante Punkt erfüllen muss. Die neuen und viel interessanteren Gleichungen

$$\begin{aligned} 0 = L_{x_1}(\mathbf{x}^*, \boldsymbol{\lambda}^*) &= \frac{\partial f}{\partial x_1}(\mathbf{x}^*) - \sum_{j=1}^m \lambda_j^* \frac{\partial g_j}{\partial x_1}(\mathbf{x}^*) \\ &\vdots \\ 0 = L_{x_n}(\mathbf{x}^*, \boldsymbol{\lambda}^*) &= \frac{\partial f}{\partial x_n}(\mathbf{x}^*) - \sum_{j=1}^m \lambda_j^* \frac{\partial g_j}{\partial x_n}(\mathbf{x}^*) \end{aligned}$$

lassen sich auch einfacher in vektorieller Form darstellen:

$$\mathbf{grad} f(\mathbf{x}^*) - \sum_{j=1}^m \lambda_j^* \mathbf{grad} g_j(\mathbf{x}^*) = \mathbf{0} \quad (\star)$$

Um diese Gleichung zu beweisen, gehen wir wie folgt vor:

1. Weg:

Wir wissen, dass der Punkt \mathbf{x}^* eine Lösung des Systems

$$\begin{aligned} f(\mathbf{x}) &= c_0 \\ g_1(\mathbf{x}) &= c_1 \\ &\vdots \\ g_m(\mathbf{x}) &= c_m \end{aligned}$$

ist. Mit Hilfe des Satzes über implizite Funktionen (unter der Voraussetzung des Satzes, dass $\text{rang } D\mathbf{g}(\mathbf{x}^*) = m$) kann gezeigt werden, dass der Rang der Matrix

$$\begin{pmatrix} \frac{\partial f}{\partial x_1}(\mathbf{x}^*) & \frac{\partial f}{\partial x_2}(\mathbf{x}^*) & \dots & \frac{\partial f}{\partial x_n}(\mathbf{x}^*) \\ \frac{\partial g_1}{\partial x_1}(\mathbf{x}^*) & \frac{\partial g_1}{\partial x_2}(\mathbf{x}^*) & \dots & \frac{\partial g_1}{\partial x_n}(\mathbf{x}^*) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial x_1}(\mathbf{x}^*) & \frac{\partial g_m}{\partial x_2}(\mathbf{x}^*) & \dots & \frac{\partial g_m}{\partial x_n}(\mathbf{x}^*) \end{pmatrix}$$

nicht $m+1$, sondern m ist. Das heisst, dass die Zeilen (und die Spalten) linear abhängig sind. Also existieren Zahlen $\alpha_0, \alpha_1, \dots, \alpha_n$ so dass

$$\alpha_0 \begin{pmatrix} \frac{\partial f}{\partial x_1}(\mathbf{x}^*) \\ \vdots \\ \frac{\partial f}{\partial x_n}(\mathbf{x}^*) \end{pmatrix} + \alpha_1 \begin{pmatrix} \frac{\partial g_1}{\partial x_1}(\mathbf{x}^*) \\ \vdots \\ \frac{\partial g_1}{\partial x_n}(\mathbf{x}^*) \end{pmatrix} + \dots + \alpha_m \begin{pmatrix} \frac{\partial g_m}{\partial x_1}(\mathbf{x}^*) \\ \vdots \\ \frac{\partial g_m}{\partial x_n}(\mathbf{x}^*) \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$$

Für den Koeffizienten α_0 gilt $\alpha_0 \neq 0$, denn andernfalls würde ein Widerspruch zu $D\mathbf{g}(\mathbf{x}^*) = m$ entstehen. Letztendlich dividiert man die gesamte Gleichung durch α_0 , bedenkt dass die Vektoren gerade die Gradienten aller vorkommenden Funktionen sind und setzt dann stets $\lambda_i = -\alpha_i/\alpha_0$. \square

2. Weg:

Dieser Weg zum Beweis der Gleichung (\star) lässt sogar eine ökonomische Deutung der Lagrange-Multiplikatoren zu. Allerdings müssen wir an die Differenzierbarkeit der **Optimalwertfunktion**

$$f^*(\mathbf{c}) = \max\{f(\mathbf{x}) \mid \mathbf{g}(\mathbf{x}) = \mathbf{c}\}$$

glauben. Wäre f eine Gewinnfunktion und $\mathbf{c} = (c_1, \dots, c_m)$ der Ressourcenvektor, dann wäre $f^*(\mathbf{c})$ der maximale Gewinn für die gegebenen Ressourcen.

Für einen gegebenen Vektor \mathbf{c} sei $\mathbf{x}^* = \mathbf{x}^*(\mathbf{c})$ der zugehörige optimale Wert, d.h. $f(\mathbf{x}^*) = f^*(\mathbf{c})$ und sicher gilt für alle \mathbf{x} auch $f(\mathbf{x}) \leq f^*(\mathbf{g}(\mathbf{x}))$. Somit hat die reelle Funktion

$$\phi(\mathbf{x}) = f(\mathbf{x}) - f^*(\mathbf{g}(\mathbf{x}))$$

ein Maximum für $\mathbf{x} = \mathbf{x}^*$. Falls die Optimalwertfunktion differenzierbar ist folgt somit für alle $i = 1, \dots, n$:

$$0 = \frac{\partial \phi}{\partial x_i}(\mathbf{x}^*) = \frac{\partial f}{\partial x_i}(\mathbf{x}^*) - \sum_{j=1}^m \left[\frac{\partial f^*}{\partial c_j}(\mathbf{c}) \right]_{\mathbf{c}=\mathbf{g}(\mathbf{x}^*)} \frac{\partial g_j}{\partial x_i}(\mathbf{x}^*)$$

Setzen wir nun noch

$$\lambda_j := \frac{\partial f^*}{\partial c_j}(\mathbf{c}) \approx f^*(\mathbf{c} + \mathbf{e}_j) - f^*(\mathbf{c})$$

folgt die Gleichung (*). □

Beispiel 3.3

$$\text{Zielfunktion} \quad : y = x_1^2 + x_2^2 + x_3^2$$

$$\text{Nebenbedingungen} \quad : \begin{cases} g_1(\mathbf{x}) = x_1 + 2x_2 + x_3 = 30 \\ g_2(\mathbf{x}) = 2x_1 - x_2 - 3x_3 = 10 \end{cases}$$

Die Jacobi-Matrix

$$D\mathbf{g}(\mathbf{x}) = \begin{pmatrix} 1 & 2 & 1 \\ 2 & -1 & -3 \end{pmatrix}$$

hat für alle \mathbf{x} den maximalen Rang 2.

Lagrange-Funktion

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

Partielle Ableitungen

$$\begin{aligned} 0 &= L_{x_1} = 2x_1 - \lambda_1 - 2\lambda_2 \\ 0 &= L_{x_2} = 2x_2 - 2\lambda_1 + \lambda_2 \\ 0 &= L_{x_3} = 2x_3 - \lambda_1 + 3\lambda_2 \\ 0 &= L_{\lambda_1} = x_1 + 2x_2 + x_3 - 30 \\ 0 &= L_{\lambda_2} = 2x_1 - x_2 - 3x_3 - 10 \end{aligned}$$

Lösung (10, 10, 0, 12, 4)

3.3 f, g stet. dfb, $g(\mathbf{x}) \leq c$, rang Dg maximal \longrightarrow Kuhn-Tucker

3.3.1 Zwei Variablen und eine Nebenbedingung

Das Optimierungsproblem hat die folgende allgemeine Gestalt, wobei wir ohne Einschränkung der Allgemeinheit annehmen wollen, dass wir die Zielfunktion maximieren wollen.

$$\begin{aligned} \text{maximiere Zielfunktion} & : y = f(x_1, x_2) = f(\mathbf{x}) \\ \text{Nebenbedingung} & : g(x_1, x_2) = g(\mathbf{x}) \leq c \end{aligned}$$

Nehmen wir an, dass $\mathbf{x}^* = (x_1^*, x_2^*)$ dieses Problem löst und sei

$$L(x_1, x_2) = f(x_1, x_2) - \lambda (g(x_1, x_2) - c)$$

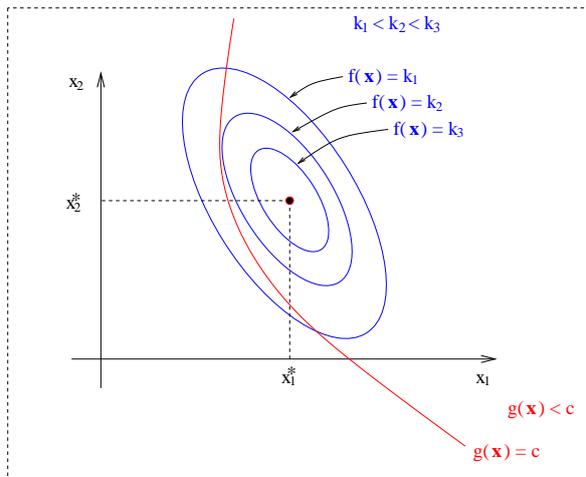
die Lagrange-Funktion des Optimierungsproblems. Dann trifft genau einer der folgenden beiden Fälle zu.

$$g(\mathbf{x}^*) < c$$

Die Bedingung $g(\mathbf{x}) \leq c$ heisst dann inaktiv oder schlaff.

$$g(\mathbf{x}^*) = c$$

Die Bedingung $g(\mathbf{x}) \leq c$ heisst dann aktiv oder bindend.



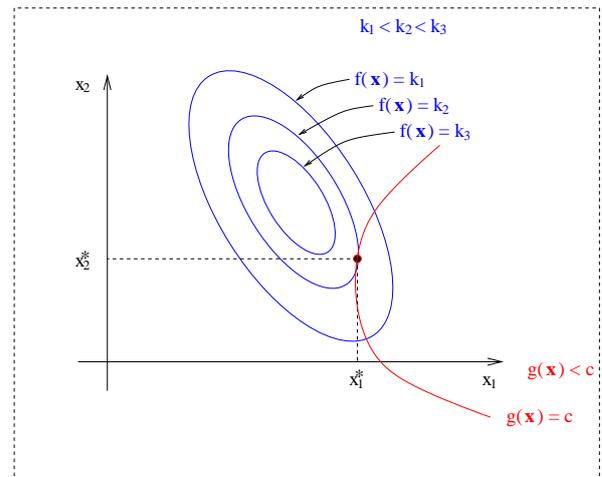
\mathbf{x}^* ist gewöhnliche innere Extremalstelle von f , d.h. hier muss gelten:
 $f_{x_1}(\mathbf{x}^*) = 0$ und $f_{x_2}(\mathbf{x}^*) = 0$

Also hier:

Setze $\lambda = 0$ und löse die Gleichungen

$$L_{x_1}(\mathbf{x}^*) = f_{x_1}(\mathbf{x}^*) = 0$$

$$L_{x_2}(\mathbf{x}^*) = f_{x_2}(\mathbf{x}^*) = 0$$



\mathbf{x}^* löst das Lagrange-Problem mit Zielfunktion f und Nebenbedingung $g(\mathbf{x}) = c$ und beachte, dass $\lambda \geq 0$ ist, denn **grad** $g(\mathbf{x}^*)$ und **grad** $f(\mathbf{x}^*)$ zeigen in die selbe Richtung.

Also hier:

Setze $\lambda \geq 0$ und löse die Gleichungen

$$L_{x_1}(\mathbf{x}^*) = 0$$

$$L_{x_2}(\mathbf{x}^*) = 0$$

Überprüfe, ob alle gefundenen Kandidaten $g(\mathbf{x}) \leq c$ erfüllen.

Beide Fälle zusammenfassend, können wir den folgenden Algorithmus zur Lösung des Problems

$$\begin{aligned} \text{maximiere Zielfunktion} & : y = f(x_1, x_2) = f(\mathbf{x}) \\ \text{Nebenbedingung} & : g(x_1, x_2) = g(\mathbf{x}) \leq c \end{aligned}$$

formulieren.

Rezept zur Lösung nach Kuhn-Tucker

1. Ordne der Nebenbedingung einen konstanten Lagrange-Multiplikator zu und definiere die Lagrange-Funktion:

$$L(x_1, x_2) = f(x_1, x_2) - \lambda (g(x_1, x_2) - c)$$

2. Setze

$$\begin{aligned} L_{x_1}(x_1, x_2) &= 0 \\ L_{x_2}(x_1, x_2) &= 0 \end{aligned}$$

3. Führe die so genannte komplementäre Schlaffheitsbedingung ein.

Erste Formulierung:

$$\lambda \geq 0 \quad \text{und} \quad \lambda = 0 \quad \text{falls} \quad g(\mathbf{x}) < c$$

Zweite äquivalente Formulierung:

$$\lambda \geq 0 \quad \text{und} \quad \lambda [g(\mathbf{x}) - c] = 0$$

4. Verlange, dass alle Kandidaten die Bedingung $g(\mathbf{x}) \leq c$ erfüllen.

Die Bedingungen 2. und 3. werden auch als Kuhn-Tucker-Bedingungen bezeichnet. Diese Bedingungen sind notwendige Bedingungen für eine Lösung des Optimierungsproblems und im Allgemeinen weit davon entfernt, hinreichend zu sein.

Beispiel 3.4 Wir wollen das folgende Optimierungsproblem nach Kuhn-Tucker lösen:

$$\begin{aligned} \text{maximiere Zielfunktion} & : f(x_1, x_2) = x_1^2 + x_2^2 + x_2 - 1 \\ \text{Nebenbedingung} & : g(x_1, x_2) = x_1^2 + x_2^2 \leq 1 \end{aligned}$$

1. Die Lagrange-Funktion ist:

$$L(x_1, x_2) = x_1^2 + x_2^2 + x_2 - 1 - \lambda (x_1^2 + x_2^2 - 1)$$

2. Setze

$$\begin{aligned} L_{x_1}(x_1, x_2) = 2x_1 - 2\lambda x_1 & = 0 & (i) \\ L_{x_2}(x_1, x_2) = 2x_2 + 1 - 2\lambda x_2 & = 0 & (ii) \end{aligned}$$

3. Die komplementäre Schlaffheitsbedingung ist:

$$\lambda \geq 0 \quad \text{und} \quad \lambda = 0 \quad \text{falls} \quad x_1^2 + x_2^2 < 1 \quad (iii)$$

4. Am Ende müssen wir noch verlangen, dass alle Kandidaten die Bedingung $x_1^2 + x_2^2 \leq 1$ erfüllen.

Wir suchen nun zunächst Punkte, die die Bedingungen (i), (ii) und (iii) erfüllen.

Aus (i) folgt $\lambda = 1$ oder $x_1 = 0$, wobei der Fall $\lambda = 1$ mit (ii) ein Widerspruch ergibt. Es bleibt also $x_1 = 0$.

- Nehmen wir an, dass $x_1^2 + x_2^2 = 1$. Dann folgt mit $x_1 = 0$ sofort $x_2 = \pm 1$.
 - Wählen wir zuerst $x_2 = 1$, so folgt mit (ii) $\lambda = 3/2$ und Bedingung (iii) ist erfüllt. Somit ist $(0, 1)$ mit $\lambda = 3/2$ ein Optimalitätskandidat.
 - Wählen wir $x_2 = -1$, so folgt mit (ii) $\lambda = 1/2$ und Bedingung (iii) ist erfüllt. Somit ist $(0, -1)$ mit $\lambda = 1/2$ ein Optimalitätskandidat.
- Betrachten wir schliesslich den Fall, dass $x_1 = 0$ und auch $x_1^2 + x_2^2 < 1$. Das heisst, dass $-1 < x_2 < 1$ sein muss. Dann impliziert (iii) natürlich $\lambda = 0$ und aus (ii) folgt $x_2 = -1/2$. Somit ist auch $(0, -1/2)$ mit $\lambda = 0$ ein Optimalitätskandidat.

Die stetige Funktion f muss auf jeder beschränkten und abgeschlossenen Menge ihr Maximum annehmen und die drei gefundenen Kandidaten sind die einzig möglichen Lösungen. Anhand der Funktionswerte

$$f(0, 1) = 1, \quad f(0, -1) = -1 \quad \text{und} \quad f(0, -1/2) = -5/4$$

erkennen wir, dass der Punkt $(0, 1)$ das obige Maximierungsproblem löst.

3.3.2 n Variablen und m Nebenbedingung

Wir betrachten wieder das allgemeine Problem einer zu maximierenden Zielfunktion und m Nebenbedingungen in Ungleichungsform.

$$\begin{array}{ll} \text{maximiere Zielfunktion} & : y = f(x_1, x_2, \dots, x_n) = f(\mathbf{x}) \\ \text{Nebenbedingungen} & : \begin{cases} g_1(x_1, x_2, \dots, x_n) = g_1(\mathbf{x}) \leq c_1 \\ g_2(x_1, x_2, \dots, x_n) = g_2(\mathbf{x}) \leq c_2 \\ \dots \\ g_m(x_1, x_2, \dots, x_n) = g_m(\mathbf{x}) \leq c_m \end{cases} \end{array}$$

Definition 3.3 Sei \mathbf{x}^* eine lokale Maximalstelle des obigen Maximierungsproblems. Eine Nebenbedingung $g_i(\mathbf{x}) \leq c_i$ heisst

- bindend (oder aktiv) für \mathbf{x}^* , falls $g_i(\mathbf{x}^*) = c_i$ gilt und
- nicht bindend (oder inaktiv) für \mathbf{x}^* , falls $g_i(\mathbf{x}^*) < c_i$ gilt.

Satz 11 Seien f und $\mathbf{g} = (g_1, g_2, \dots, g_m)$ stetig differenzierbare Funktionen und sei \mathbf{x}^* eine lokale Lösung des Optimierungsproblems

$$\begin{array}{ll} \text{maximiere Zielfunktion} & : y = f(\mathbf{x}) \\ \text{Nebenbedingungen} & : \begin{cases} g_1(\mathbf{x}) \leq c_1 \\ \dots \\ g_m(\mathbf{x}) \leq c_m \end{cases} \quad \text{oder kurz} \quad \mathbf{g}(\mathbf{x}) \leq \mathbf{c} \end{array}$$

Die Nebenbedingungen sollen so geordnet sein, dass die ersten m_0 Nebenbedingungen bindend und die letzten $m - m_0$ Nebenbedingungen nicht bindend für \mathbf{x}^* sind. Weiterhin sei der Rang der Jacobi-Matrix (der bindenden Nebenbedingungen)

$$\begin{pmatrix} \frac{\partial g_1}{\partial x_1}(\mathbf{x}^*) & \dots & \frac{\partial g_1}{\partial x_n}(\mathbf{x}^*) \\ \vdots & \ddots & \vdots \\ \frac{\partial g_{m_0}}{\partial x_1}(\mathbf{x}^*) & \dots & \frac{\partial g_{m_0}}{\partial x_n}(\mathbf{x}^*) \end{pmatrix}$$

gleich m_0 , also maximal.

Dann existieren reelle Zahlen $\boldsymbol{\lambda}^* = (\lambda_1^*, \dots, \lambda_m^*)$, so dass

1. $L_{x_1}(\mathbf{x}^*, \boldsymbol{\lambda}^*) = 0, \dots, L_{x_n}(\mathbf{x}^*, \boldsymbol{\lambda}^*) = 0$,
2. $\lambda_1^* \geq 0, \dots, \lambda_m^* \geq 0$,
3. $\lambda_1^* [g_1(\mathbf{x}^*) - c_1] = 0, \dots, \lambda_m^* [g_m(\mathbf{x}^*) - c_m] = 0$ und
4. $g_1(\mathbf{x}^*) \leq c_1, \dots, g_m(\mathbf{x}^*) \leq c_m$.

Rezept zur Lösung nach Kuhn-Tucker

1. Ordne jeder Nebenbedingung Lagrange-Multiplikator $(\lambda_1, \lambda_2, \dots, \lambda_m)$ zu und definiere die Lagrange-Funktion:

$$L(\mathbf{x}) = f(\mathbf{x}) - \sum_{j=1}^m \lambda_j (g_j(\mathbf{x}) - c_j)$$

2. Setze alle partiellen Ableitungen von L nach den x_i für $i = 1, 2, \dots, n$ gleich 0

$$0 = L_{x_i}(\mathbf{x}) = f_{x_i}(\mathbf{x}) - \sum_{j=1}^m \lambda_j g_{x_i}(\mathbf{x})$$

3. Führe die komplementären Schlaffheitsbedingungen für alle $j = 1, 2, \dots, m$ ein.
Erste Formulierung:

$$\lambda_j \geq 0 \quad \text{und} \quad \lambda_j = 0 \quad \text{falls} \quad g_j(\mathbf{x}) < c_j$$

Zweite äquivalente Formulierung:

$$\lambda_j \geq 0 \quad \text{und} \quad \lambda_j [g_j(\mathbf{x}) - c_j] = 0$$

4. Verlange, dass alle Kandidaten die Bedingungen $g_j(\mathbf{x}) \leq c_j$ für alle $j = 1, 2, \dots, m$ erfüllen.

Es werden alle \mathbf{x} mit den zugehörigen Werten $\lambda_1, \lambda_2, \dots, \lambda_m$ bestimmt, die die Bedingungen 2., 3. und 4. erfüllen. Das sind die Lösungskandidaten, von denen mindestens einer das Problem löst, falls es überhaupt lösbar ist.

3.4 Numerische Verfahren zur Maximierung

Sei $f : S(\subset \mathbb{R}^n) \rightarrow \mathbb{R}$ stetig differenzierbar. Unser Ziel ist es wieder, eine (lokale) Maximalstelle von f zu finden. Ein typisches numerisches Verfahren zur Lösung dieses Problems wird durch den folgenden recht allgemeinen Modellalgorithmus beschrieben.

Wähle einen Startwert $\mathbf{x}_0 \in S$ und führe folgende vier Schritte für $k = 0, 1, 2, \dots$ genügend oft aus:

1. Prüfe auf Abbruch (meistens STOP falls $\mathbf{grad} f(\mathbf{x}_k) = \mathbf{0}$ ist).
2. Berechne eine Aufstiegsrichtung $\mathbf{v}_k \in \mathbb{R}^n$, d.h. eine Richtung mit $\partial_{\mathbf{v}_k} f(\mathbf{x}_k) > 0$.
3. Bestimme eine Schrittweite $\sigma_k > 0$ s.d.

$$f(\mathbf{x}_k + \sigma_k \mathbf{v}_k) > f(\mathbf{x}_k)$$

gilt und die Zunahme der Zielfunktion, d.h. der Ausdruck $f(\mathbf{x}_k + \sigma_k \mathbf{v}_k) - f(\mathbf{x}_k)$ hinreichend gross ist.

4. Setze $\mathbf{x}_{k+1} := \mathbf{x}_k + \sigma_k \mathbf{v}_k$.

Aufstiegsrichtungen Typische Aufstiegsrichtungen in \mathbf{x}_k sind

-

$$\mathbf{v}_k = \frac{\mathbf{grad} f(\mathbf{x}_k)}{\|\mathbf{grad} f(\mathbf{x}_k)\|},$$

die die lineare Approximation (Tangentialebene) von f maximiert und

- die so genannte Newton-Richtung

$$\mathbf{v}_k = (H_f(x_k))^{-1} \mathbf{grad} f(\mathbf{x}_k),$$

die die quadratische Approximation (Taylor-Polynom 2. Ordnung) von f maximiert.

Schrittweiten Um die Schrittweiten σ_k zu steuern, gibt es verschiedene Möglichkeiten. Wir müssen natürlich stets sicherstellen, dass die neu konstruierten Punkte in der Menge S verbleiben.

- Bestimme die globalen Maximalstellen der Funktion (in einer reellen Variablen)

$$F(t) = f(\mathbf{x}_k + t\mathbf{v}_k)$$

für $t \in [0, \infty)$, d.h. wir suchen eine Lösung $t^* = \sigma_k$ der Gleichung

$$0 = F'(t^*) = \mathbf{grad} f(\mathbf{x}_k + t^* \mathbf{v}_k) \bullet \mathbf{v}_k.$$

Wir dürfen aber im Allgemeinen nicht darauf hoffen, dieses Problem exakt lösen zu können.

- Iterative Methoden Wir geben eine maximale Schrittweite, z.B. $t = 1$, vor und halbieren diese Schrittweite $t \rightarrow \frac{1}{2}t$ bis man für einen Wert $\frac{1}{2^i} = \sigma_k$ mit

$$f(\mathbf{x}_k + \frac{1}{2^i} \mathbf{v}_k) > f(\mathbf{x}_k)$$

gefunden hat.

Beispiel 3.5 Sei A eine negativ definite, symmetrische $n \times n$ -Matrix, also $A^T = A$ und $\mathbf{x}^T A \mathbf{x} < 0$ für alle $\mathbf{x} \neq \mathbf{0}$. Wir definieren die Funktion

$$f(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T A \mathbf{x} - \mathbf{b}^T \mathbf{x}$$

für ein $\mathbf{b} \in \mathbb{R}^n$, deren Graph man sich als Verallgemeinerung einer Parabel (Paraboloid) vorstellen könnte. Daher hat f eine eindeutig bestimmte lokale (und auch globale) Maximalstelle \mathbf{x}^* . Es gilt

$$\mathbf{grad} f(\mathbf{x}) = A \mathbf{x} - \mathbf{b}$$

und die Maximalstelle \mathbf{x}^* ist die eindeutig bestimmte Lösung ($\text{rang} A = n$) des linearen Gleichungssystems

$$\mathbf{grad} f(\mathbf{x}^*) = A \mathbf{x}^* - \mathbf{b} = \mathbf{0}.$$

Die beiden Probleme

1. Löse $A \mathbf{x} = \mathbf{b}$ und
2. Maximiere $f(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T A \mathbf{x} - \mathbf{b}^T \mathbf{x}$

sind also gleichwertig.

Wir wollen nun das Aufstiegsverfahren auf das Maximierungsproblem anwenden. Wir wählen die übliche Aufstiegsrichtung

$$\mathbf{v}_k = \frac{\mathbf{grad} f(\mathbf{x}_k)}{\|\mathbf{grad} f(\mathbf{x}_k)\|} = \frac{A \mathbf{x}_k - \mathbf{b}}{\|A \mathbf{x}_k - \mathbf{b}\|}.$$

Zur Schrittweitenbestimmung versuchen wir nun, die reelle Funktion

$$F(t) = f\left(\mathbf{x}_k + t \frac{A \mathbf{x}_k - \mathbf{b}}{\|A \mathbf{x}_k - \mathbf{b}\|}\right)$$

für $t \in [0, \infty)$ zu maximieren. Das gesuchte t^* muss wieder die folgende notwendige und hinreichende Bedingung erfüllen:

$$\begin{aligned} 0 = F'(t^*) &= \mathbf{grad} f\left(\mathbf{x}_k + t^* \frac{A \mathbf{x}_k - \mathbf{b}}{\|A \mathbf{x}_k - \mathbf{b}\|}\right) \cdot \frac{A \mathbf{x}_k - \mathbf{b}}{\|A \mathbf{x}_k - \mathbf{b}\|} \\ &= \|A \mathbf{x}_k - \mathbf{b}\| + \frac{t^*}{\|A \mathbf{x}_k - \mathbf{b}\|^2} (A \mathbf{x}_k - \mathbf{b})^T A (A \mathbf{x}_k - \mathbf{b}), \end{aligned}$$

also

$$t^* = \frac{-\|A \mathbf{x}_k - \mathbf{b}\|^3}{(A \mathbf{x}_k - \mathbf{b})^T A (A \mathbf{x}_k - \mathbf{b})}$$

und somit erhalten wir die Iterationsvorschrift

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \frac{\|A \mathbf{x}_k - \mathbf{b}\|^2}{(A \mathbf{x}_k - \mathbf{b})^T A (A \mathbf{x}_k - \mathbf{b})} (A \mathbf{x}_k - \mathbf{b}).$$

4 Übungen

Zur Lösung jeder Aufgabe gehört der vollständige und nachvollziehbare Rechenweg. Jede nicht triviale Behauptung sollte (kurz) begründet werden.

1. Ein Unternehmen verwendet Kapital K , Arbeit A und Land L , um Q Einheiten eines Gutes zu produzieren, wobei

$$Q(K, A, L) = K^{2/3} + A^{1/2} + L^{1/3}.$$

Weiterhin seien p der Verkaufspreis pro Einheit, k der Preis pro Einheit Kapital, a der Preis pro Einheit Arbeit und l der Preis pro Einheit Land.

- (a) Wie lautet die Gewinnfunktion? Bestimmen Sie die Funktionen $K^*(p, k, a, l)$, $A^*(p, k, a, l)$ und $L^*(p, k, a, l)$, die den Gewinn maximieren. 5
- (b) Zeigen Sie: $\frac{\partial Q^*}{\partial k} = -\frac{\partial K^*}{\partial p}$. 5

2. Betrachten Sie das folgende allgemeine Optimierungsproblem mit $c, \alpha, \beta, p, q, m > 0$:

$$\begin{array}{ll} \text{Zielfunktion} & Q(K, A) = c \cdot K^\alpha \cdot A^\beta \\ \text{Nebenbedingung} & g(K, A) = p \cdot K + q \cdot A = m \end{array}$$

- (a) Bestimmen Sie die Optimalwertfunktionen $Q^* = Q^*(p, q, m)$, $K^* = K^*(p, q, m)$, $A^* = A^*(p, q, m)$ und λ^* . 7
- (b) Sind die Funktionen $K^* = K^*(p, q, m)$ und $A^* = A^*(p, q, m)$ homogen? Bestimmen Sie gegebenenfalls den Homogenitätsgrad. Hätte man diese Eigenschaft der beiden Funktionen voraussehen können? 7
- (c) Sei $\alpha = \beta = 1$, $c = 120$, $p = 2$ und $q = 5$. Bestimmen Sie $\frac{dQ^*(m)}{dm}$ und $Q^*(101) - Q^*(100)$. 6

3. Betrachten Sie das folgende Optimierungsproblem:

$$\begin{array}{ll} \text{maximiere Zielfunktion} & f(x_1, x_2) = x_1^2 + 2x_2^2 - x_1 \\ \text{Nebenbedingung} & g(x_1, x_2) = x_1^2 + x_2^2 \leq 1 \end{array}$$

- (a) Schreiben Sie die Lagrange-Funktion und die Kuhn-Tucker-Bedingungen für dieses Problem auf. 3
- (b) Bestimmen Sie alle Paare (x_1, x_2) , die die Kuhn-Tucker-Bedingungen erfüllen. Es gibt fünf Kandidaten. Bestimmen Sie die Lösung des Problems. 7

4. Wenden Sie das Kuhn-Tucker-Verfahren zur Lösung des folgenden Optimierungsproblems an. Jeder Schritt muss logisch nachvollziehbar sein. 15

$$\text{maximiere Zielfunktion} \quad : f(m, x) = m + \ln(x)$$

$$\text{Nebenbedingungen} \quad : \begin{cases} g_1(m, x) = m + x \leq 5 \\ g_2(m, x) = -m \leq 0 \\ g_3(m, x) = -x \leq 0 \end{cases}$$

5. Führen Sie alle Rechnungen in Beispiel 3.5 im Detail aus und finden Sie eventuelle Fehler. 5