# Kapitel 4

# Nichtlineare Optimierung

Wir wollen nun etwas allgemeinere Optimierungsprobleme betrachten. Das allgemeine Problem lautete

$$\min_{x \in S} f(x).$$

Bei einer linearen Zielfunktion ist diese Fragestellung nur sinnvoll, wenn der Bereich $S$ beschränkt ist, bei nicht-linearen Zielfunktionen ist oft schon die unbeschränkte Optimierungsaufgabe mit $S = \mathbb{R}^n$ schwer.

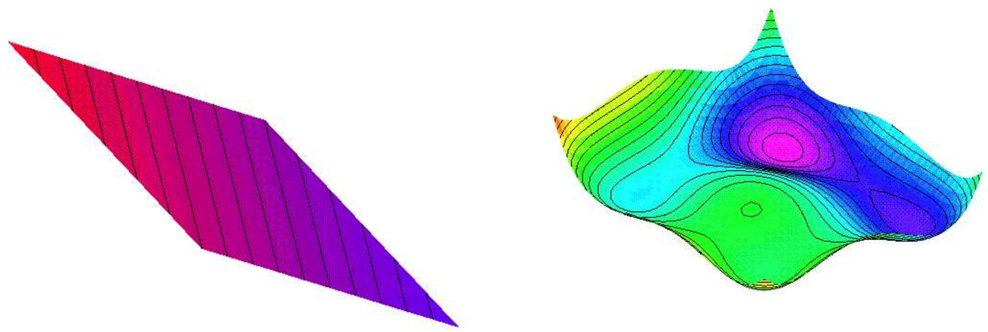


Abbildung 4.1: Lineare Funktion gegen $x^4 + y^4 - 5x^2 - 4y^2 + 5x + 2y - 1.5$

Bei der Minimierung nicht-linearen Funktionen spielt die folgende einfache Strategie eine zentrale Rolle. Ausgehend von einem Punkt $x_i$ suche eine Abstiegsrichtung und gehe in dieser Richtung bestmöglich zu $x_{i+1}$. Iteriere bis es keine Abstiegsrichtung mehr gibt. Im allgemeinen findet man so kein *globales* Minimum, sondern nur *lokale* (oder auch relative) Minima. Aber auch die Theorie macht oftmals nur Aussagen über *lokale* Minima.

**Definition 4.0.3** *Sei $S \subseteq \mathbb{R}^n, f : S \to \mathbb{R}$ und $x^* \in S$. Dann heißt $x^*$ ein* lokales *Minimum von $f$ über $S$ (oder auch* relatives *Minimum), wenn es ein $\varepsilon > 0$ gibt mit $\forall x \in S \cap U_\varepsilon(x^*) : f(x) \geq f(x^*)$. Gilt sogar $\forall x \in S \cap U_\varepsilon(x^*) : f(x) > f(x^*)$, so ist $x^*$ ein* striktes lokales *Minimum.*

*Falls $\forall x \in S : f(x) \geq f(x^*)$, so heißt $x^*$ ein* globales *Minimum und analog zum Vorigen sprechen wir von einem* strikten globalen *Minimum falls die letzte Ungleichung stets strikt ist.*

# 4.1 Wiederholung aus der mehrdimensionalen Differentialrechnung

## 4.1.1 Kurven

Wie bereits erwähnt haben wir im wesentlichen nur Mittel zur Bestimmung lokaler Minima zur Hand. Die Werkzeuge dafür liefert die Analysis. Eines der zentralen Anliegen der Analysis ist es, Funktionen lokal durch lineare Funktionen (und evtl. Terme höherer Ordnung) zu approximieren. Dafür muß die Funktion aber lokal „hinreichend dicht"definiert sein. Oftmals betreibt man deshalb Analysis nur auf *offenen Mengen $U$*, das sind Mengen, bei denen zu jedem $x \in U$ ein $\varepsilon > 0$ mit $U_\varepsilon(x) \subseteq U$ existiert. Unsere Mengen $S$ werden im allgemeinen nicht offen sein, aber wir werden stets annehmen, daß die Kostenfunktion auf einer offenen Menge $U$ definiert ist, die $S$ enthält.

Um die Optimierungsstrategie aus der Einleitung dieses Kapitels präzisieren zu können, wiederholen wir nun Wege und Richtungen von Wegen im $\mathbb{R}^n$.

**Definition 4.1.1** *Sei $I \subseteq \mathbb{R}$ ein Intervall. Eine* Kurve *oder ein* Weg *im $\mathbb{R}^n$ ist eine stetige Abbildung*

$$\begin{array}{rcll} c: & I & \to \mathbb{R}^n & \\ & t & \mapsto & c(t) = (c_1(t), \ldots, c_n(t))^\top, \end{array}$$

*d.h. alle Komponentenfunktionen sind stetig. Analog heißt eine Kurve $k-$fach* (stetig) differenzierbar*, wenn alle Komponentenfunktionen $k-$fach (stetig) differenzierbar sind, für $k \in \mathbb{N}$.*

*Ist $t_0 \in I$, so nennen wir $c'(t_0) := (c'_1(t_0), \ldots, c'_n(t_0))$ den* Tangentialvektor an *$c$ in $t_0$.*

**Beispiel 4.1.2** *Wir betrachten die übliche Parametrisierung des Einheitskreises*

$$\begin{array}{rccl} c: & [0,2\pi] & \to \mathbb{R}^2 & \\ & t & \mapsto & c(t) = (\sin(t), \cos(t)), \end{array}$$

*in* $t_0 = \pi$*. Dann ist* $c(t_0) = (0, -1)$ *und* $c'(t_0) = (\cos(\pi), -\sin(\pi)) = (-1, 0)$*.*

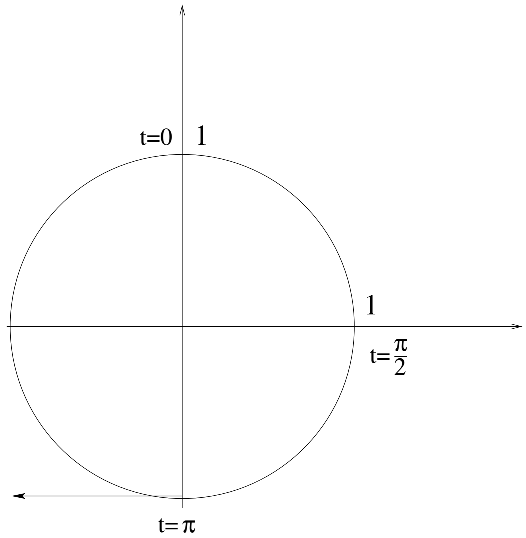



Abbildung 4.2: Eine Tangente an den Kreis

**Proposition 4.1.3** *Sei* $S \subseteq \mathbb{R}^n, f : S \to \mathbb{R}$ *und* $x^* \in S$*. Ist* $x^*$ *ein lokales Minimum (striktes lokales Minimum) von* $f$*, so gilt für jeden Weg* $c : [0, \varepsilon] \to S$ *mit* $c(0) = x^*$ *und* $c'(0) \neq 0$*:*

$$\exists \delta \leq \varepsilon \,\forall 0 < \alpha \leq \delta : (f \circ c)(\alpha) := f(c(\alpha)) \geq f(x^*) \text{ (bzw. } (f(c(\alpha)) > f(x^*))\text{).}$$

**Beweis.** Ist $x$ ein (striktes) lokales Minimum, so gibt es ein $\eta$ mit $f(x) \overset{>}{\geq} f(x^*)$ für alle $x \in U_\eta(x^*)$. Sei nun $c : [0, \varepsilon] \to S$ ein Weg mit $c(0) = x^*$ und $c'(0) \neq 0$. Da $c$ stetig ist, gibt es ein $\delta > 0$ mit $0 < \alpha \leq \delta$ impliziert $c(t) \in U_\varepsilon(x^*)$. □

### 4.1.2 Partielle Ableitungen

Im letzten Abschnitt haben wir Abbildungen $\mathbb{R} \to \mathbb{R}^n$ untersucht. In der Optimierung haben wir es bei der Zielfunktion üblicherweise mit Abbildungen $\mathbb{R}^n \to \mathbb{R}$ zu tun (vgl. Abbildung 4.1). Dafür definieren wir Richtungsableitungen entlang eines Weges. Eine besondere Rolle spielen hierbei die Richtungen der Koordinatenachsen.

**Definition 4.1.4** *Sei $I \subseteq \mathbb{R}$ ein Intervall, $t_0 \in I$, $S \subseteq \mathbb{R}^n$, $c : I \to \mathbb{R}^n$ eine stetig differenzierbare Kurve und $f : S \to \mathbb{R}$ mit $p = c(t_0) \in S$ und $c'(t_0) = d$. Existiert dann $(f \circ c)'(t_0)$, so heißt diese Zahl die* Richtungsableitung $\frac{\partial f}{\partial d}$ von $f$ in Richtung $d$ in $p$. *Ist $c'(t_0) = e_i$, so heißt*

$$\frac{\partial f}{\partial x_i}(p) := (f \circ c)'(t_0)$$

*die $i$*-te partielle Ableitung von $f$. *Der* Gradient $\nabla f(p)$ von $f$ in $p$ *ist der Zeilenvektor der partiellen Ableitungen $(\frac{\partial f}{\partial x_1}(p), \ldots, \frac{\partial f}{\partial x_n}(p))$.*

*Ist $h : U \to \mathbb{R}^k$ eine vektorwertige Ableitung, so bezeichnen wir mit $Jh$ die* Jacobische *d.i. die Matrix der partiellen Ableitungen*

$$Jh(x) := \begin{pmatrix} \frac{\partial h_1}{\partial x_1} & \cdots & \frac{\partial h_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_n}{\partial x_1} & \cdots & \frac{\partial h_n}{\partial x_n} \end{pmatrix}.$$

*Existieren alle partiellen Ableitungen einer reellwertigen Funktion $f$ und sind stetig, so sagen wir $f$ ist* stetig differenzierbar. *Wir können nun die Definition der Jacobischen auch auf den Gradienten anwenden. Zunächst definieren wir dafür die* zweiten partiellen Ableitungen

$$\frac{\partial^2 f}{\partial x_i \partial x_j} := \frac{\partial \left(\frac{\partial f}{\partial x_i}\right)}{\partial x_j} =: \frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}\right).$$

*Für $\frac{\partial^2 f}{\partial x_i \partial x_i}$ schreiben wir auch kürzer $\frac{\partial^2 f}{\partial x_i^2}$. Existieren alle zweiten partiellen Ableitungen und sind stetig, so heißt $f$ zweimal stetig differenzierbar. Die Jacobische des Gradienten nennen wir* Hessematrix $\nabla^2 f(x)$. *Die Hessematrix ist also die Matrix der zweiten partiellen Ableitungen*

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

*Ist $U \subseteq \mathbb{R}^n$ und $f$ $k$-fach stetig differenzierbar in $U$, so schreiben wir $f \in C^k(U)$.*

**Bemerkung 4.1.5** *Genau genommen ist der von uns hier gewählte Zugang nicht ganz sauber, da nicht klar ist, daß die oben eingeführte Richtungsableitung unabhängig von der Wahl des Weges $c$ ist. Wir berufen uns dafür auf Bekanntes aus der Mathematik für Chemiker und Wirtschaftsinformatiker, da wir zur Herleitung der Zuasammenhänge hier keine Zeit haben.*

**Proposition 4.1.6** *Ist $f$ zweimal stetig differenzierbar, so ist*

$$\frac{\partial f}{\partial d}(p) = \nabla f(p)d.$$

**Beweis.** Anwendung der Kettenregel aus Mathematik für Chemiker und Wirtschaftsinformatiker Kapitel IV. □

**Beispiel 4.1.7** *Wir betrachten die Funktion $f(x, y) = x^4+y^4-5x^2-4y^2+5x+2y-1.5$. Dann ist $\nabla f(x, y) = (4x^3 - 10x + 5, 4y^3 - 8y + 2)$ und*

$$\nabla^2 f(x) := \begin{pmatrix} 12x^2 - 10 & 0 \\ 0 & 12y^2 - 8 \end{pmatrix}.$$

Wir wollen diese Wiederholung abschließen mit einer mehrdimensionalen Konsequenz aus dem Satz von Taylor.

**Satz 4.1.8** *Sei $p \in S \subseteq \mathbb{R}^n$ und $f : S \to \mathbb{R}$ zweimal stetig differenzierbar. Sei $c$ ein zweimal stetig differenzierbarer Weg mit $c(t_0) = p$ und $c'(t_0) = d$. Dann ist*

$$(f\circ c)(t) = f(p)+\nabla f(p)d(t-t_0)+\frac{1}{2}\nabla f(p)c''(t_0)(t-t_0)^2+\frac{1}{2}d^\top\nabla^2 f(p)d(t-t_0)^2+o((t-t_0)^2).$$

**Beweis.** Nach dem Satz von Taylor für Variablen einer Veränderlichen ist $(f\circ c)(t) = f(p)+(f\circ c)'(t_0)(t-t_0)+\frac{1}{2}(f\circ c)''(t_0)(t-t_0)^2+o((t-t_0)^2)$. Nach Definition der Richtungsableitung und Proposition 4.1.6 ist $(f\circ c)'(t_0) = \nabla f(p)d$. Zu zeigen ist also nur $(f\circ c)''(t_0) = \left(((\nabla f)\circ c)\, c'(t)\right)'(t_0) = \nabla f(p)c''(t_0)+d^\top\nabla^2 f(p)d$ (vgl. Produktregel). Dafür betrachten wir die Funktion

$$\begin{aligned} (f\circ c)'(t) &= \nabla f(c(t))c'(t) \\ &= \sum_{i=1}^n c_i'(t)\frac{\partial f}{\partial x_i}(c(t)) \\ &= \sum_{i=1}^n c_i'(t)\left(\frac{\partial f}{\partial x_i}\circ c\right)(t). \end{aligned}$$

Nach Additionsregel dürfen wir Summandenweise ableiten und berechnen zunächst mit der Produktregel und Proposition 4.1.6:

$$(c_i'\frac{\partial f}{\partial x_i}\circ c)'(t_0) = c_i''(t_0)\frac{\partial f}{\partial x_i}(p)+d_i\nabla\frac{\partial f}{\partial x_i}d.$$

Fassen wir die Summanden zusammen, so erhalten wir:

$$\begin{aligned}(f \circ c)''(t_0) &= \sum_{i=1}^{n} \left( c_i''(t_0) \frac{\partial f}{\partial x_i}(p) + d_i \nabla \frac{\partial f}{\partial x_i} d \right) \\ &= \nabla f(p) c''(t_0) + d^\top \nabla^2 f(p) d.\end{aligned}$$

□

## 4.2 Notwendige und hinreichende Bedingungen für Extremwerte

Bei den folgenden Überlegungen wollen wir den zulässigen Bereich $S$ in unsere Überlegungen miteinbeziehen. Kommen wir zurück auf die algorithmische Idee aus der Einleitung dieses Kapitels, so müssen wir irgendwie ausdrücken, was es heißt, daß es in keine Richtung mehr „bergab“ geht.

**Definition 4.2.1** *Sei $S \subseteq \mathbb{R}^n$, $x \in S$ und $d \in \mathbb{R}^n$. Dann heißt $d$* zulässige Richtung *für $x$ bzgl. $S$, wenn es ein $\varepsilon > 0$ und einen differenzierbaren Weg $c : [0, \varepsilon] \to S$, gibt mit $c(0) = x$ und $c'(0) = d$.*

Notwendig für ein lokales Minimum ist, daß es keine zulässige Abstiegsrichtung gibt.

**Proposition 4.2.2 (Notwendige Bedingung erster Ordnung)** *Sei $S \subseteq U \subseteq \mathbb{R}^n$, $U$ offen, $f \in C^1(U)$. Ist dann $x^*$ ein relatives Minimum von $f$ in $S$, so gilt für jede zulässige Richtung $d$ für $x$ bzgl. $S$:*

$$\nabla f(x^*) d \geq 0.$$

**Beweis.** Da $d$ zulässige Richtung ist, gibt es ein $\varepsilon > 0$ und einen differenzierbaren Weg $c : [0, \varepsilon] \to S$ mit $c(0) = x^*$ und $c'(0) = d$. Wir betrachten wieder die Funktion $f \circ c$. Nach Definition der Ableitung ist

$$(f \circ c)'(0) = \lim_{h \to 0} \frac{(f \circ c)(h) - f(x^*)}{h}.$$

Wäre nun $(f \circ c)'(0) < 0$, so folgte hieraus die Existenz eines $0 < \delta < \varepsilon$ mit

$$\frac{(f \circ c)(\alpha) - f(x^*)}{\alpha} < 0 \text{ für alle } 0 < \alpha < \delta$$

und somit $(f \circ c)(\alpha) < f(x^*)$ für alle $\alpha < \delta$. Dies widerspricht laut Proposition 4.1.3 der Minimalität von $x^*$. Also folgt notwendig $(f \circ c)'(0) \overset{4.1.6}{=} \nabla f(x^*) d \geq 0$. □

Für den Spezialfall eines Minimum im Innern von $S$ erhalten wir folgende Aussage, die ganz analog zur notwendigen Bedingung für ein Extremum aus der Kurvendiskussion ist.

**Korollar 4.2.3** *Ist $x^*$ ein relatives Minimum von $f$ im Innern von $S$, d.h. es gibt $\varepsilon > 0$ mit $U_\varepsilon(x) \subseteq S$, so ist $\nabla f(x^*) = 0$.*

**Beweis.** Nach Proposition 4.2.2 gilt $\nabla f(x^*)d \geq 0$ für alle zulässigen Richtungen. Nach Voraussetzung sind alle $d \in \mathbb{R}^n \setminus \{0\}$ zulässig, betrachte dazu jeweils den Weg $c_d(t) = x^* + td$. Da mit jedem $d$ auch $-d$ zulässig ist, schließen wir $\nabla f(x^*)d = 0$ für alle $d \in \mathbb{R}^n$ und somit $\nabla f(x^*) = 0$. □

**Beispiel 4.2.4** *Wir betrachten das Optimierungsproblem*

$$\begin{aligned} &\min f(x_1, x_2) = x_1^2 - x_1 + x_2 + x_1 x_2 \\ &\textit{unter}\ x_1 \geq 0,\ \ x_2 \geq 0. \end{aligned}$$

*Der Gradient dieser Funktion ist $\nabla f(x_1, x_2) = (2x_1 - 1 + x_2, x_1 + 1)$. Im Innern der zulässigen Bereiches $S$ ist $x_1 > 0$, also kann der Gradient nicht verschwinden, folglich hat die Funktion im Innern kein lokales Minimum. Am Rand gilt $x_1 = 0$ oder $x_2 = 0$. Im ersten Fall sind die zulässigen Richtungen genau die Vektoren $d$ mit $d_1 \geq 0$ und wir haben als Gradienten $(x_2 - 1, 1)$. Die Bedingung aus Proposition 4.2.2 kann hier nicht erfüllt werden, da z.B. $(0, -1)$ stets eine zulässige Richtung ist. Im zweiten Fall ist der Gradient $(2x_1 - 1, x_1 + 1)$ und eine Richtung ist zulässig genau dann, wenn $d_2 \geq 0$. Somit ist die Richtung $(1 - 2x_1, 0)$ zulässig im Punkt $(x_1, x_2)$ und wir erhalten als Bedingung $-(2x_1 - 1)^2 \geq 0$. Wir schließen hieraus, daß der einzige Kandidat für ein lokales Minimum $x^* = (\frac{1}{2}, 0)$ ist. Wir haben $f(x^*) = -\frac{1}{4}$. Dieser Wert ist auch das globale Minimum der Funktion $x_1^2 - x_1$ und wegen $x_1, x_2 \geq 0$ haben wir $f(x_1, x_2) \geq x_1^2 - x_1$ also hat $f$ in $x^*$ sogar ein globales Minimum.*

*In Abbildung 4.3 haben wir die Isoquanten der Funktion geplottet. Der Gradient steht stets senkrecht auf diesen Isoquanten. Die Hyperebene definiert durch die Gleichung $\nabla f(x^*)x = 0$ berührt also die Isoquantenfläche.*

In der Kurvendiskussion haben Sie gelernt, daß die zweite Ableitung Auskunft über die Krümmung einer Funktion gibt. Auch das letzte Beispiel deutet an, daß die Isokostenfläche sich von der Tangentialhyperebene „wegkrümmen" muß. Die in der Kurvendiskussion kennengelernten notwendigen Bedingungen zweiter Ordnung für lokale Minima gelten nun für alle Richtungen.

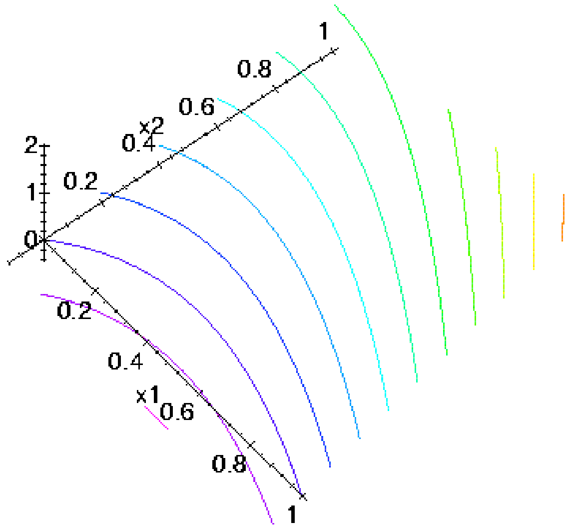



Abbildung 4.3: Isoquanten von 4.2.4

**Proposition 4.2.5 (Notwendige Bedingungen zweiter Ordnung)** *Sei $S \subseteq U \subseteq \mathbb{R}^n$, $U$ offen, $f \in C^2(U)$. Ist dann $x^*$ ein relatives Minimum von $f$ in $S$, so gilt für jedes $d$, für das es ein $\epsilon$ gibt mit $x^* + \alpha d \in S$ für $0 \leq \alpha \leq \varepsilon$:*

*a)* $\nabla f(x^*)d \geq 0$,

*b)* *falls* $\nabla f(x^*)d = 0$, *so ist* $d^\top \nabla^2 f(x)d \geq 0$.

**Beweis.** Wir haben nur die zweite Behauptung zu zeigen. Sei also $c : [0, \varepsilon] \to S$ definiert durch $c(t) = x^* + td$. Dann ist $d$ zulässige Richtung, $c(0) = x^*$ und $c'(0) = d$. Weil $c'(t) = d$ konstant ist, verschwindet $c''(0)$ und nach Satz 4.1.8 ist, $(f \circ c)(\alpha) := f(x^*) + \nabla f(x^*)d\alpha + \frac{1}{2}d^\top \nabla^2 f(x^*)d\alpha^2 + o(\alpha^2) = f(x^*) + \frac{1}{2}d^\top \nabla^2 d\alpha^2 + o(\alpha^2)$ für $\alpha \to 0$. Angenommen nun $d^\top \nabla^2 f(x)d < 0$, dann ist für hinreichend kleines $\alpha$ nach Definition der Landausymbole $\alpha^2 \frac{1}{2} d^\top \nabla^2 f(x)d + o(\alpha^2) < 0$ und somit $(f \circ c)(\alpha) < f(x^*)$ im Widerspruch zur Minimalität von $x^*$. □

Auch hier wollen wir wieder den Fall eines inneren Punktes gesondert notieren.

**Korollar 4.2.6** *Ist $x^*$ ein relatives Minimum von $c$ im Innern von $S$, so gilt für alle $d \in \mathbb{R}^n$.*

*a)* $\nabla f(x^*) = 0$,

*b)* $d^\top \nabla^2 f(x) d \geq 0$.

□

Für die letzte Ungleichung sagen wir auch: die Hessematrix ist *positiv semidefinit* (vgl. Definition 2.5.1).

Ähnlich wie im Eindimensionalen lassen sich die notwendigen Bedingung zweiter Ordnung zu hinreichenden verschärfen.

**Proposition 4.2.7 (Hinreichende Bedingungen zweiter Ordnung)** *Sei $U \subseteq \mathbb{R}^n$, $U$ offen, $f \in C^2(U)$ und $x^* \in U$. Gilt dann*

*a)* $\nabla f(x^*) = 0$,

*b) und $\nabla^2 f(x^*)$ ist positiv definit,*

*so ist $x^*$ ein striktes lokales Minimum von $f$.*

**Beweis.** Wie im Beweis von Proposition 4.2.5 entwickeln wir $f(x^* + \alpha d) = f(x^*) + \alpha^2 \frac{1}{2} d^\top \nabla^2 f(x^*) d + o(\alpha^2)$. Für $\alpha$ hinreichend klein ist wegen $d^\top \nabla^2 f(x^*) d > 0$ folglich $f(x^* + \alpha d) > f(x^*)$. Hieraus folgt, daß $x^*$ auf jedem Strahl ein striktes lokales Minimum ist. Die Behauptung folgt dann aus dem folgenden Satz. □

**Satz 4.2.8** *Sei $x^* \in U \subseteq \mathbb{R}^n$ offen und $f \in C(U)$ eine Funktion. Gibt es dann für alle $d \in \mathbb{R}^n \setminus \{0\}$ ein $\alpha_d > 0$, so daß $0$ striktes lokales Minimum der Funktion $f_d : [0, \alpha_d] \to \mathbb{R}$, definiert durch $f_d(t) := f(x^* + td)$, ist, so ist $x^*$ striktes lokales Minimum von $f$.*

**Beweis.** Sei $S^{n-1} := \{d \in \mathbb{R}^n \mid \|d\| = 1\}$ die Menge aller Einheitsrichtungen. Nach Voraussetzung gibt es für jedes $d \in S^{n-1}$ ein $\alpha_d$ mit $f_d(t) > f_d(0)$ für $0 < t < \alpha_d$. Sei für jedes $d$ ein solches $\alpha_d \in ]0, \infty[$ maximal gewählt. (Beachten Sie, daß das Maximum, falls es von $\infty$ verschieden ist, existiert, da wir $f_d(t) > f_d(0)$ für $t < \alpha_d$ fordern.) Sei nun $U_k := \{d \in S^{n-1} \mid \alpha_d > \frac{1}{k}\}$. Dann gilt $U_k \subseteq U_{k+1}$ und $S^{n-1} = \bigcup_{i=1}^\infty U_i$. Gibt es nun ein $n_0$ mit $S^{n-1} = U_{n_0}$, so ist $x^*$ ein globales Minimum, wenn man $f$ auf $U_{\frac{1}{n_0}}$ einschränkt, also speziell ein lokales Minimum. Nehmen wir also an, daß dies nicht der Fall ist. Da $S^{n-1}$ beschränkt und abgeschlossen ist, können wir eine konvergente Folge $d_k$ mit $d_k \in S^{n-1} \setminus U_k$ konstruieren. Sei $d = \lim_{k\to\infty} d_k$. Sei $k_0 \geq \frac{1}{\alpha_d}$. Dann ist $f(x^* + td) > f(x^*)$ für $0 < t < \frac{1}{n_0}$. Da $f$ stetig ist, gilt diese Ungleichung in einer ganzen Umgebung $U_{\varepsilon_1}(x^* + td) \subseteq \mathbb{R}^n$. Somit ist aber eine ganze Umgebung $U_{\varepsilon_2}(d) \subseteq S^{n-1}$ in $U_{\frac{1}{n_0}}$ enthalten, im Widerspruch dazu, daß $d = \lim_{n\to\infty} d_n$.

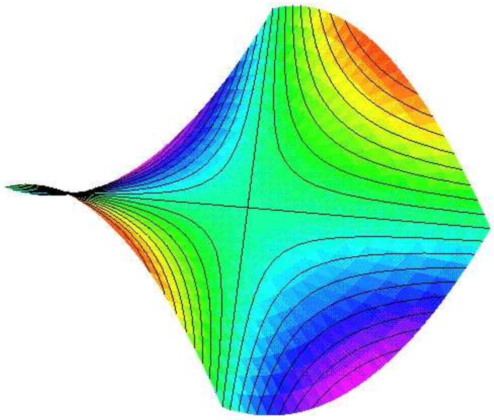
□

## 4.3 Bedingungen für Extrema auf (Un)gleichungsdefinierten Mengen

Wir wollen uns nun mit etwas spezielleren zulässigen Bereichen beschäftigen. Und zwar wollen wir auch hier eine verallgemeinerte Version der Nebenbedingungen der linearen Programmierung betrachten, nämlich Teilmengen des $\mathbb{R}^n$, die sich durch Ungleichungen $g_1(x) \leq 0, \ldots, g_k \leq 0$ mit differenzierbaren Funktionen $g_1, \ldots, g_k : \mathbb{R}^n \to \mathbb{R}$ beschreiben lassen. Wie wir im letzten Kapitel gesehen haben, ist die Situation „im Innern" eines solchen Gebildes relativ einfach. Um die Ränder genauer zu untersuchen wollen wir zunächst gleichungsdefinierte Mengen, sogenannte *Mannigfaltigkeiten*, betrachten.

### 4.3.1 Exkurs Mannigfaltigkeiten und Tangentialräume

Seien $h_1, \ldots, h_k : \mathbb{R}^n \to \mathbb{R}$ differenzierbare Funktionen und $k \leq n$. Wir wollen die Lösungsmenge der Gleichung $h(x) = 0$ untersuchen, wobei $h = (h_1, \ldots, h_n)^\top$. Betrachten wir hierzu als Beispiel die Funktion $h(x, y) = (x^2 - y^2)^2 - a$ für $a \geq 0$. Die betrachtete Menge ist also gerade die Höhenlinie der Funktion $(x^2 - y^2)^2$ zum Wert $a$.



Im allgemeinen können wir erwarten, daß eine Gleichung eine Höhenhyperfläche definiert, da sie einen „Freiheitsgrad einschränkt". Allerdings möchten wir Zerteilungspunkte, wie in der Abbildung im Ursprung zu sehen, ausschließen. Man beachte, daß der Gradient der dargestellten Funktion im Ursprung verschwindet. Bei mehreren Funktionen $h_i$ ist man auf der sicheren Seite, wenn die Gradienten linear unabhängig sind. Dies ist die Aussage des Satzes über implizit definierte Funktionen, der einer der zentralen Sätze der Differentialrechnung mehrerer Veränderlicher ist. Wir werden ihn im Rahmen dieser Vorlesung nicht beweisen.

**Satz 4.3.1 (Satz über implizit definierte Funktionen)** *Sei $T \subseteq \mathbb{R}^{l+k}, x^* \in \mathbb{R}^l$, $y^* \in \mathbb{R}^k$ und $(x^*, y^*)$ ein innerer Punkt von $T$. Sei $h = (h_1, \ldots, h_k) : T \to \mathbb{R}^k$ stetig differenzierbar, die Matrix $\frac{\partial h}{\partial y}(x^*, y^*)$ regulär und $h(x^*, y^*) = 0$. Dann gibt es $\varepsilon_1, \varepsilon_2 > 0$ mit $U_{\varepsilon_1}(x^*) \times U_{\varepsilon_2}(y^*) \subseteq T$ und genau eine stetig differenzierbare Funktion $f : U_{\varepsilon_1} \to U_{\varepsilon_2}$ mit $h(x, f(x)) = 0$ und $Jf(x^*) = -(\frac{\partial h}{\partial y}(x^*, y^*))^{-1} \frac{\partial h}{\partial x}(x^*, y^*)$.*

Der Satz besagt also, daß die Lösungsmenge solcher „gutartiger Gleichungssysteme“ lokal wie ein verformter $\mathbb{R}^l$ aussieht. Überprüfen wir den Spezialfall linearer Gleichungssysteme:

**Beispiel 4.3.2** *Sei $A = (A_1, A_2) \in \mathbb{R}^{(l+k)\times n}, b \in \mathbb{R}^k$ und $A_2$ regulär. Sei $h(x) = Ax - b = A_1x_1 + A_2x_2 - b$. Wir berechnen daraus $x_2 = A_2^{-1}(b_2 - A_1x_1)$.*

Als Anwendung des Satzes über implizit definierte Funktionen werden wir nun zeigen, daß die Menge aller Tangentialvektoren an die Mannigfaltigkeit in einem regulären Punkt $p^*$, das sind die Tangentialvektoren von Wegen, die ganz in der Mannigfaltigkeit verlaufen und $p^*$ enthalten, ein Vektorraum ist.

**Satz 4.3.3** *Sei $T \subseteq \mathbb{R}^{l+k}, h = (h_1, \ldots, h_k) : T \to \mathbb{R}^k$ stetig differenzierbar im inneren Punkt $p^* \in T, h(p^*) = 0$ und die Gradienten $\nabla h_1(p^*), \ldots, \nabla h_k(p^*)$ linear unabhängig. Dann ist die Menge der Tangentialvektoren an die Mannigfaltigkeit $S = \{x \in \mathbb{R}^{l+k} \mid h(x) = 0\}$ in $p^*$ der l-dimensionale Untervektorraum $L$ des $\mathbb{R}^{l+k}$ :*

$$L = \ ker \begin{pmatrix} \nabla h_1(p^*) \\ \vdots \\ \nabla h_k(p^*) \end{pmatrix} = kerJh = \{d \in \mathbb{R}^{l+k} \mid \nabla h_i(p^*)d = 0, i = 1, \ldots, k\}.$$

**Beweis.** Ist $d$ ein Tangentialvektor und $c : [0, \varepsilon] \to S$ mit $c(0) = p^*$ und $c'(0) = d$, so ist $h_i \circ c(t) = 0$ für alle $i$ und $t$ und somit $0 = (h_i \circ c)'(0) = \nabla h_i(p^*)d$. Sei nun umgekehrt $d$ mit $\nabla h_i(p^*)d = 0$ für alle $i$ gegeben. Da die Gradienten von $h_1, \ldots, h_k$ linear unabhängig sind, können wir $k$ Koordinaten $x_{i_1}, \ldots, x_{i_k}$ wählen, so daß die Matrix

$$\frac{\partial h}{\partial(x_{i_1}, \ldots, x_{i_k})} := \begin{pmatrix} \frac{\partial h_1}{\partial x_{i_1}} & \cdots & \frac{\partial h_1}{\partial x_{i_k}} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_k}{\partial x_{i_1}} & \cdots & \frac{\partial h_k}{\partial x_{i_n}} \end{pmatrix}$$

regulär ist. Nach eventueller Umnummerierung können wir annehmen, daß dies die letzten $k$ Koordinaten sind. Wir spalten $d$ und $p^*$ dazu passend auf in $d = (d_1, d_2)$ und $p^* = (p_1, p_2)$. Nach dem Satz 4.3.1 über implizit definierte Funktionen, gibt es $\varepsilon_1, \varepsilon_2 > 0$ mit $U_{\varepsilon_1}(p_1) \times U_{\varepsilon_2}(p_2) \subseteq T$ und genau eine differenzierbare Funktion $f : U_{\varepsilon_1}(p_1) \to U_{\varepsilon_2}(p_2)$ mit $h(x_1, f(x_1)) = 0$ für alle $x_1 \in U_{\varepsilon_1}(p_1)$. Wir betrachten den Weg $c : [0, \frac{\varepsilon_1}{2}] \to S$ definiert durch $c(t) = (c_1(t), c_2(t)) := (p_1 + td_1, f(p_1 + td_1))$. Dann ist $c$ ein differenzierbarer Weg in $S$ mit $c(0) = p^*$. Nach Kettenregel (oder zeilenweise mit Proposition 4.1.6) ist $c_2'(0) = Jf(p_1)d_1$. Nach dem Satz über implizit definierte Funktionen ist

$$Jf(p_1) = -\left(\frac{\partial h}{\partial(x_{l+1}, \ldots, x_{l+k})}\right)^{-1}(p^*)\frac{\partial h}{\partial(x_1, \ldots, x_l)}. \tag{4.1}$$

Da nach Voraussetzung $0 = Jhd = \frac{\partial h}{\partial(x_1,\ldots,x_l)}d_1 + \frac{\partial h}{\partial(x_{l+1},\ldots,x_{l+k})}d_2$ erhalten wir

$$d_2 = -\left(\frac{\partial h}{\partial(x_{l+1}, \ldots, x_{l+k})}\right)^{-1}(p^*)\frac{\partial h}{\partial(x_1, \ldots, x_l)}d_1. \tag{4.2}$$

Insgesamt erhalten wir also

$$c'(0) = (d_1, Jf(p_1)d_1) = (d_1, d_2) = d. \tag{4.3}$$

□

### 4.3.2 Lagrange-Multiplikatoren

Die Erkenntnisse des letzten Paragraphen wollen wir nun dazu benutzen, eine notwendige Bedingung für einen lokalen Extremwert auf einer gleichungsdefinierten Mannigfaltigkeit zu formulieren. Die geometrische Bedeutung des folgenden Satzes ist, daß in einem lokalen Minimum, der Gradient der Zielfunktion senkrecht auf der Mannigfaltigkeit stehen muß.

**Satz 4.3.4** *Seien $h : \mathbb{R}^{l+k} \to \mathbb{R}^k, f : \mathbb{R}^{k+l} \to \mathbb{R}$ differenzierbare Funktionen und $x^* \in S$ sei ein lokales Minimum der Optimierungsaufgabe*

$$\begin{array}{cc} \min & f(x) \\ \textit{unter} & h(x) = 0 \end{array}.$$

*Ferner seien $\nabla h_1(x^*), \ldots, \nabla h_k(x^*)$ linear unabhängig. Dann gibt es $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$ mit*

$$\nabla f(x^*) = \lambda_1 \nabla h_1(x^*) + \ldots + \lambda_k \nabla h_k(x^*).$$

*Die $\lambda_i$ nennt man* Langrange'sche Multiplikatoren.

**Beweis.** Gibt keine solchen Lagrange'schen Multiplikatoren, so liegt $\nabla f(x^*)$ nicht im Vektorraum $L^\perp$ aller Linearkombinationen von $\nabla h_1(x^*), \ldots, \nabla h_k(x^*)$. Wählen wir eine Orthonormalbasis $b_1, \ldots, b_k$ von $L^\perp$, so ist also

$$d^\top := -\nabla f(x^*) + \sum_{i=1}^{k} (\nabla f(x^*) b_i) b_i^\top \neq 0.$$

Das heißt, der Gradient von $f$ hat eine nicht verschwindende Komponente im Tangentialraum, denn offensichtlich ist $d^\top b_i = 0$ für alle $b_i$. Da die $b_i$ den gleichen Raum aufspannen wie die $\nabla h_i(x^*)$, gilt also auch $\nabla h_i(x^*) d = 0$. Nach Satz 4.3.3 gibt es also einen differenzierbaren Weg $c : [0, \varepsilon] \to S$ mit $c(0) = x^*, c'(0) = d$, wobei $S = \{x \in \mathbb{R}^n \mid h(x) = 0\}$. Angenommen nun, $x^*$ wäre ein lokales Minimum, dann gäbe es ein $\alpha > 0$ mit $f(x^*) \leq (f \circ c)(t)$ für alle $t \leq \alpha$. Wir schließen hieraus

$$(f \circ c)'(0) := \lim_{t \searrow 0} \frac{(f \circ c)(t) - f(x^*)}{t} \geq 0. \tag{4.4}$$

Andererseits ist aber nach Proposition 4.1.6

$$\begin{aligned} (f \circ c)'(0) &= \nabla f(x^*) d \\ &= (-d^\top + \sum_{i=1}^{k} (\nabla f(x^*) b_i) b_i^\top) d \\ &= -d^\top d < 0. \end{aligned}$$

Aus diesem Widerspruch folgt die Behauptung. □

Auch hier können wir wieder notwendige Bedingungen 2. Ordnung angeben.

**Satz 4.3.5** *Unter den Voraussetzungen des letzten Satzes gilt: Ist $x^* \in S$ ein lokales Minimum der Optimierungsaufgabe, so gibt es $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$ mit*

*a)* $\nabla f(x^*) = \sum_{i=1}^{k} \lambda_i \nabla h_i(x^*)$ *und*

*b)* *die Matrix $L := \nabla^2 f(x^*) - \sum_{i=1}^{k} \lambda_i \nabla^2 h_i(x^*)$ ist positiv semidefinit auf dem Tangentialraum von $S = \{x \in \mathbb{R}^{l+k} \mid h(x) = 0\}$ in $x^*$, d.h. für alle Vektoren $d \in ker(Jh(x^*))$ gilt $d^\top L d \geq 0$.*

**Beweis.** Wir haben nur die zweite Bedingung zu zeigen. Sei also $Jh(x^*) d = 0$ und $c : [0, \varepsilon] \to S$ ein Weg mit $c(0) = x^*$ und $c'(0) = d$. Da $f$ auch lokales Minimum der Funktion $f \circ c$ ist, gilt bekanntlich $(f \circ c)''(0) \geq 0$. Wir hatten bereits für den Taylorschen Satz ausgerechnet

$$(f \circ c)''(0) = \nabla f(x^*)c''(0) + d^\top \nabla^2 f(x^*)d.$$

Genauso erhalten wir durch zweimaliges Differenzieren:

$$(h_i \circ c)''(0) = \nabla h_i(x^*)c''(0) + d^\top \nabla^2 h_i(x^*)d$$

und somit wegen $(h_i \circ c) \equiv 0 : \nabla h_i(x^*)c''(0) = -d^\top \nabla^2 h_i(x^*)d$. Wir setzen dies unter Ausnutzung von $\nabla f(x^*) = \sum_{i=1}^k \lambda_i \nabla h_i(x)^*$ zusammen zu

$$\begin{aligned} 0 &\leq \nabla f(x^*)c''(0) + d^\top \nabla^2 f(x^*)d \\ &= \sum_{i=1}^k \lambda_i \nabla h_i(x)^* c''(0) + d^\top \nabla^2 f(x^*)d \\ &= \sum_{i=1}^k -\lambda_i d^\top \nabla^2 h_i(x^*)d + d^\top \nabla^2 f(x^*)d \\ &= d^\top \left( \nabla^2 f(x^*) - \sum_{i=1}^k \lambda_i \nabla^2 h_i(x^*) \right) d. \end{aligned}$$

□

Den Beweis für die hinreichenden Bedingungen zweiter Ordnung wollen wir uns sparen.

**Satz 4.3.6** *Seien $h : \mathbb{R}^{k+l} \to \mathbb{R}^k, f : \mathbb{R}^{k+l} \to \mathbb{R}$ differenzierbare Funktionen. Gelte $h(x^*) = 0, \nabla f(x^*) + \sum_{i=1}^k \lambda_i \nabla h_i(x^*)$ für einen Vektor $\lambda$. Sei ferner die Matrix $\nabla^2 f(x^*) + \sum_{i=1}^k \lambda_i \nabla^2 h_i(x^*)$ positiv definit. Dann ist $x^*$ striktes lokales Minimum der Optimierungsaufgabe*

$$\begin{array}{ll} \min & f(x) \\ \textit{unter} & h(x) = 0. \end{array}$$

**Beweis.** Luenberger Seite 227.

### 4.3.3 Kuhn-Tucker Bedingungen

In diesem letzten Paragraphen wollen wir die Ergebnisse des letzten Abschnitts auf Mengen übertragen, die durch Gleichungen und Ungleichungen definiert sind.

Dafür benötigen wir ein Analogon zu Satz 4.3.3. Da die zugehörige Menge anschaulich nicht mehr wie ein „Tangentialraum aussieht" wollen wir hier wieder von zulässigen Richtungen sprechen. Im folgenden setzen wir nicht mehr voraus, daß $n = k + l$!

**Satz 4.3.7** *Seien $h : \mathbb{R}^n \to \mathbb{R}^k$ und $g : \mathbb{R}^n \to \mathbb{R}^l$ stetig differenzierbare Funktionen und*

$$S := \{x \in \mathbb{R}^n \mid h(x) = 0, g(x) \le 0\}.$$

*Für $x \in S$ sei $eq(x) = \{i \in \{1, \dots, l\} \mid g_i(x) = 0\}$. Sei $x^*$ ein* regulärer Punkt von *S, d.h. $\nabla h_1(x^*), \dots, \nabla h_k(x^*), \nabla g_{i_j}(x^*)$ für $i_j \in eq(x^*)$ seien linear unabhängig. Dann ist $d \in \mathbb{R}^n$ eine zulässige Richtung für $x^*$ genau dann, wenn $Jhd = 0$ und $\nabla g_{i_j}(x^*)d \le 0$ falls $i_j \in eq(x^*)$.*

**Beweis.** Ist $d$ eine zulässige Richtung und $c : [0, \varepsilon]$ ein zugehöriger Weg, so folgt wie in Satz 4.3.3 $\nabla h_i(x^*)d = 0$ und $\nabla g_{i_j}d \le 0$ falls $i_j \in eq(x^*)$. Sei nun umgekehrt $d$ gegeben mit $\nabla h_i(x^*)d = 0$ und $\nabla g_{i_j}d \le 0$ falls $i_j \in eq(x^*)$. Sei $\tilde{S} := \{x \in \mathbb{R}^n \mid h(x) = 0$ und $g_{i_j}(x) = 0$ falls $\nabla g_{i_j}d = 0\}$. Dann ist $x^* \in \tilde{S}$ und die Gradienten der definierenden Gleichungen von $\tilde{S}$ sind linear unabhängig in $x^*$. Nach Satz 4.3.3 ist $d$ im Tangentialraum von $\tilde{S}$, also gibt es $c : [0, \varepsilon] \to T$ mit $c(0) = x^*$ und $c'(0) = d$. Um zu zeigen, daß $c$, zumindest nahe bei der Null, in $S$ verläuft, müssen wir nun noch die $g_{i_j}$ mit $\nabla g_{i_j}d < 0$ untersuchen. Nach Proposition 4.1.6 wissen wir $0 > \nabla g_{i_j}(x^*)d = (g_{i_j} \circ c)'(0) = \lim_{t\to 0} \frac{g(c(t))}{t}$ woraus die Behauptung folgt. □

Somit kommen wir nun zum Hauptresultat dieses Kapitels, das besagt, daß in einem relativen Minimum der Gradient der Zielfunktion sekrecht auf allen Gleichungs- und mit Gleichheit angenommenen Ungleichungrestriktionen stehen muß, und im letzten Falle zusätzlich in die zulässige Menge hineinzeigen muß.

**Satz 4.3.8 (Kuhn-Tucker Bedingungen)** *Seien $f : \mathbb{R}^n \to \mathbb{R}, h : \mathbb{R}^n \to \mathbb{R}^k$ und $g : \mathbb{R}^n \to \mathbb{R}^l$ stetig differenzierbar und $x^*$ ein relatives Minimum des Problems*

$$\begin{array}{rl} \min & f(x) \\ \textit{unter} & h(x) = 0 \\ & g(x) \le 0 \end{array}$$

*und ein regulärer Punkt. Dann gibt es Koeffizienten $\lambda_1, \dots, \lambda_k \in \mathbb{R}$ und $\mu_1, \dots, \mu_l \in \mathbb{R}$, $\mu_i \le 0$ mit*

$$\nabla f(x^*) = \sum_{i=1}^{k} \lambda_i \nabla h_i(x^*) + \sum_{j=1}^{l} \mu_j \nabla g_j(x^*)$$

*und $\sum_{j=1}^{l} \mu_j g_j(x^*) = 0$.*

**Beweis.** Die letzte Bedingung besagt offensichtlich, daß $\mu_i \neq 0 \Rightarrow g(x^*) = 0$. Sei wieder $eq(x^*)$ die Menge der „aktiven Bedingungen“. Außerhalb von $eq(x^*)$ setzen wir die $\mu_i$ schon mal Null. Da $x^*$ auch lokales Minimum des Problems

$$\begin{array}{ll} \min & f(x) \\ \text{unter} & h(x) = 0 \\ & g_j(x) = 0 \text{ für } j \in eq(x^*) \end{array}$$

ist, gibt es nach Satz 4.3.4 Langrangeparameter $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$ und $\mu_1, \ldots, \mu_l \in \mathbb{R}$ mit $\nabla f(x^*) = \sum_{i=1}^k \lambda_i \nabla h_i(x^*) + \sum_{j=1}^l \mu_j \nabla g_j(x^*)$. Es bleibt zu zeigen, daß stets $\mu_j \leq 0$ gilt. Angenommen dies wäre nicht der Fall und $\mu_{j_0} > 0$ für $j_0 \in eq(x^*)$. Wir wählen dann eine Orthonormalbasis $b_1, \ldots, b_t$ des von $\nabla h_1(x^*), \ldots, \nabla h_k(x^*)$ und den $\nabla g_j(x^*)$ für $j \in eq(x^*) \setminus \{j_0\}$ aufgespannten Vektorraumes und, da $x^*$ ein regulärer Punkt und $\mu_{j_0} \neq 0$ ist, gilt auch $d^\top := -\nabla f(x^*) + \sum_{i=1}^t (\nabla f(x^*) b_i) b_i^\top \neq 0$. Wir schließen aus dem letzten Satz, daß $d$ eine zulässige Richtung ist, denn $d^\top b_j = 0$ für alle $j$ impliziert $\nabla h_i(x^*) d = 0$ und $\nabla g_j(x^*) d = 0$ für $j \in eq(x^*) \setminus \{j_0\}$ und

$$\begin{aligned} \mu_{j_0} \nabla g_{j_0}(x^*) d &= \left( \nabla f(x^*) - \sum_{i=1}^k \lambda_i \nabla h_i(x^*) - \sum_{\substack{j=1 \\ j \neq i}}^l \mu_j \nabla g_j(x^*) \right) d \\ &= \nabla f(x^*) d \\ &= \nabla f(x^*) \left( -\nabla f(x^*) + \sum_{i=1}^t (\nabla f(x^*) b_i) b_i^\top \right)^\top \\ &= -\|\nabla f(x)\|^2 + \sum_{i=1}^t (\nabla f(x^*) b_i)^2 \\ &= -\|\nabla f(x^*) - \sum_{i=1}^t (f(x^*) b_i) b_i\|^2 \end{aligned}$$

$$= -\| - d\|^2 < 0.$$

Also ist $d$ zulässig und Abstiegsrichtung im Widerspruch zu Proposition 4.2.2. □

**Beispiel 4.3.9** *Wir betrachten das Problem*

$$\begin{array}{ll} \max & f(x, y) = 14x - x^2 + 6y - y^2 + 7 \\ \textit{unter} & g_1(x, y) = x + y - 2 \leq 0 \\ & g_2(x, y) = x + 2y - 3 \leq 0. \end{array}$$

*Wir berechnen*

$$\begin{aligned} \nabla(-f)(x, y) &= (2x - 14, 2y - 6) \\ \nabla g_1(x) &= (1, 1) \\ \nabla g_2(x) &= (1, 2) \end{aligned}$$

*und überprüfen zunächst, wo der Gradient von f verschwindet. Dies ist der Fall in* $(7,3)$*, das nicht im zulässigen Bereich liegt.*

*Falls nur die erste Ungleichung aktiv ist, erhalten wir als Bedingungen:*

$$\begin{aligned}(2x-14, 2y-6) &= (\mu, \mu),\\ x+y &= 2,\end{aligned}$$

*woraus wir* $x = 3$ *und* $y = -1$ *berechnen. Dieser Punkt ist zulässig für* $\mu = -* \leq 0$. *Für die zweite Ungleichung haben wir*

$$\begin{aligned}(2x-14, 2y-6) &= (\mu, 2\mu),\\ x+2y &= 3,\end{aligned}$$

*und somit*

$$\begin{aligned}4x-2y &= 22\\ x+2y &= 3,\end{aligned}$$

*und wir berechnen* $x = 5, y = -1$. *Dieser Punkt ist nicht zulässig.*

*Schließlich betrachten wir noch den Fall, daß beide Ungleichungen aktiv sind. Dies ist der Fall in* $(1,1)$. *Als Kuhn-Tuckerbedingung haben wir dann* $(-12,-4) = (\mu_1+\mu_2, \mu_1+2\mu 2)$*, woraus wir* $\mu_2 = 8, \mu_1 = -20$ *berechnen. Da* $\mu_2 > 0$ *ist, kann hier auch kein lokales Minimum vorliegen.*

*Da die Funktion bei betraglich wachsendem x oder y gegen* $-\infty$ *geht, muß das globale Maximum in einem lokalen Maximum angenommen werden. Da* $(3,-1)$ *der einzige Kandidat hierfür ist, ist das Maximum der Funktion also* $f(3,-1) = 33$.

Wir wollen dieses Kapitel abschließen mit den Bedingungen zweiter Ordnung.

**Satz 4.3.10** *Unter den Bedingungen des letzten Satzes gilt außerdem: die Matrix* $\nabla^2 f(x^*) - \sum_{i=1}^k \lambda_i \nabla^2 h_i^* - \sum_{j=1}^l \mu_j \nabla^2 g_j(x^*)$ *ist positiv semidefinit auf dem Tangentialraum der aktiven Nebenbedingungen von* $x^*$.

**Beweis.** Da $x^*$ auch ein relatives Minimum für das entsprechende gleichungsdefinierte Problem ist, folgt die Behauptung sofort aus Satz 4.3.5. □

Die analogen hinreichenden Bedingungen wollen wir nicht mehr extra anführen.