

## 9 Optimierung mehrdimensionaler reeller Funktionen $f : \mathbb{R}^n \rightarrow \mathbb{R}$

### 9.1 Optimierung ohne Nebenbedingungen

Ein Optimum zu suchen heißt, den größten oder den kleinsten Wert zu suchen.

Wir suchen also ein  $x_0 \in \mathbb{R}^n$ , sodass

- $f(x_0) \leq f(x)$  für alle  $x \in \mathbb{R}^n$  (dann heißt  $x_0$  *globales Minimum*)
- $f(x_0) \geq f(x)$  für alle  $x \in \mathbb{R}^n$  (dann heißt  $x_0$  *globales Maximum*)

Die Funktion  $f$  heißt auch *Zielfunktion*.

Die Formulierung bzw. das Finden einer „geeigneten“ Zielfunktion ist meist der schwierigste Teil einer Optimierungsaufgabe.

Zunächst suchen wir aber nach **lokalen** Minima bzw. Maxima:

Suche  $x_0 \in \mathbb{R}^n$  so, dass

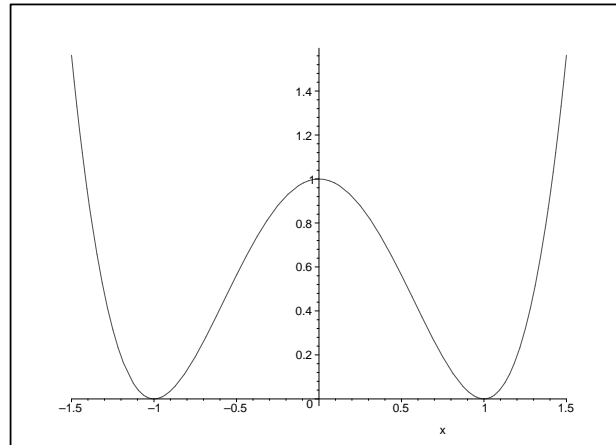
- $f(x_0) \leq f(x)$  für alle  $x$  aus einer Umgebung  $U_\varepsilon(x_0)$  um  $x_0$  ( $x$  nahe bei  $x_0$ ):  
 $x_0$  ist dann ein *lokales Minimum*
- $f(x_0) \geq f(x)$  für alle  $x$  aus einer Umgebung  $U_\varepsilon(x_0)$  um  $x_0$  ( $x$  nahe bei  $x_0$ ):  
 $x_0$  ist dann ein *lokales Maximum*.

Wie finden wir solche Punkte?

#### BEISPIEL 9.1

$n = 1$ :  $f : \mathbb{R} \rightarrow \mathbb{R}$ .

Wir berechnen die Ableitung  $f'(x)$  und setzen sie gleich 0:  $f'(x) = 0$  (horizontale Tangente, kein Anstieg/Abstieg im Extremum). Punkte, die diese Forderung erfüllen, sind Kandidaten für Extremwerte. Zur genauen Bestimmung berechnen wir  $f''(x_0)$ , die zweite Ableitung eines Kandidaten  $x_0$ , d. h., das Krümmungsverhalten der Funktion in bzw. um  $x_0$ . Falls  $f''(x_0) > 0$ , so ist  $x_0$  ein lokales Minimum, ist  $f''(x_0) < 0$ , so ist  $x_0$  ein lokales Maximum. Erhalten wir aber  $f''(x_0) = 0$ , so ist es notwendig weiterzurechnen, eventuell befinden wir uns an einer Sattelstelle, das ist jedoch genauer zu untersuchen.



**Abbildung 8:** Plot der Funktion  $(x - 1)^2(x + 1)^2$ .

Sei etwa  $f(x) = (x - 1)^2(x + 1)^2$ .

Dann ist  $f'(x) = 4x(x - 1)(x + 1)$  und dieser Ausdruck verschwindet in  $x_0 = -1$ ,  $x_1 = 0$  und  $x_2 = 1$ . Alle drei Punkte sind stationäre Punkte, also Kandidaten für Extremalstellen. Wir berechnen die zweite Ableitung und erhalten  $f''(x) = 12x^2 - 4$ . Wir stellen fest, dass  $f''(-1) = f''(1) = 8 > 0$  und  $f''(0) = -4$ . Also sind die Stellen  $x_0$  und  $x_2$  lokale Minima und die Stelle  $x_1$  ist ein lokales Maximum (siehe Abbildung 8).

Was passiert bei  $n > 1$ ?

Eine lokale Betrachtung ist möglich mittels (partieller) Ableitungen (falls die Funktion differenzierbar ist, sonst wird es komplizierter). (Partielle) Ableitungen werden in Gradienten  $\nabla f(x)$  zusammengefasst. Dieser kann auch als Richtungsvektor im  $\mathbb{R}^n$  betrachtet werden. Betrachten wir Niveaumengen der Funktion  $f(x)$ , so stellt sich heraus, dass  $\nabla f(x)$  immer „senkrecht“ zu den Niveaumengen steht (egal an welcher Stelle).

Außerdem zeigt der Gradient immer in die Richtung, in der die Funktion vom Punkt  $x$  aus gesehen am steilsten ansteigt.

Falls also  $\nabla f(x) = 0$ , dann gibt es auch lokal keinen An/Abstieg und damit eventuell ein lokales Minimum oder Maximum.

Auf diese Art können wir auch im mehrdimensionalen Raum  $\mathbb{R}^n$  Kandidaten für Extremalstellen finden.

Zur Bestimmung, ob ein Minimum oder Maximum vorliegt, brauchen wir wie im eindimensionalen Fall höhere Ableitungen.

## Mehrdimensionaler Taylor

Ähnlich wie im eindimensionalen Fall, kann auch für mehrdimensionale Funktionen unter

bestimmten Bedingungen lokal eine Approximation mit Hilfe von Ableitungen an einer bestimmten Stelle berechnet werden. Meist wird nur eine Approximation zweiten Grades benötigt (siehe auch Abschnitt 1.2.3 Mittelwertsätze):

### THEOREM 9.1

Sei  $f$  zweimal total differenzierbar bei  $x_0$ . Dann gilt:

$$f(x) = f(x_0) + \nabla f(x_0)(x - x_0) + \frac{1}{2}(x - x_0)^t \nabla^2 f(x_0)(x - x_0) + \text{Rest},$$

wobei  $\frac{\text{Rest}}{\|x - x_0\|^2} \rightarrow 0$  mit  $\|x - x_0\| \rightarrow 0$ .

### THEOREM 9.2 (NOTWENDIGE BEDINGUNG FÜR LOKALE OPTIMA)

Sei  $x_0$  lokales Minimum (Maximum) von  $f$  und  $f$  bei  $x_0$  differenzierbar, dann gilt  $\nabla f(x_0) = 0$ .

BEWEIS. [eigentlich nur Beweisskizze für Minimum]

Wir wissen, dass  $\nabla f(x_0)$  in Richtung des steilsten Anstiegs zeigt.

Angenommen,  $\nabla f(x_0) \neq 0$ . Sei  $x = x_0 - \lambda[\nabla f(x_0)]^t$ , mit  $\lambda > 0$ , klein.

Daraus folgt dann, dass auch  $\|x - x_0\| = |\lambda| \|\nabla f(x_0)\|$  klein ist. Also

$$\begin{aligned} f(x_0) \leq f(x) &= f(x_0) + \nabla f(x_0)(x - x_0) + \text{Rest (siehe Taylor)} \\ &= f(x_0) - \lambda \nabla f(x_0) \nabla f(x_0)^t + \text{Rest (da } x - x_0 = (-\lambda)[\nabla f(x_0)]^t) \\ &= f(x_0) - \lambda \|\nabla f(x_0)\|^2 + \text{Rest, dieser Rest ist aber sehr klein} \\ &< f(x_0) \text{ und daher ein Widerspruch!} \end{aligned}$$

□

Umgekehrt gilt:

### THEOREM 9.3 (HINREICHENDE BEDINGUNG FÜR LOKALE OPTIMA)

Sei  $f$  zweimal stetig differenzierbar bei  $x_0$ .

Ist die erste Ableitung  $\nabla f(x_0) = 0$  und die zweite Ableitung  $H = H_f(x_0) = \nabla^2 f(x_0)$  (die Hesse'sche Matrix) positiv definit (d. h.,  $u^t H u > 0$ , für alle  $u \neq 0$ ), dann ist  $x_0$  eine lokale Minimalstelle.

Ist die erste Ableitung  $\nabla f(x_0) = 0$  und die zweite Ableitung  $H = \nabla^2 f(x_0)$  (die Hesse'sche Matrix) negativ definit (d. h.,  $u^t H u < 0$ , für alle  $u \neq 0$ ), dann ist  $x_0$  eine lokale Maximalstelle

BEWEIS. [wieder nur Beweisskizze für Minimum]

$$\begin{aligned} f(x) &= f(x_0) + \frac{1}{2} u^t H u + \text{Rest (klein)} \\ &\geq f(x_0) \text{ für } x \text{ nahe bei } x_0 \text{ (da } u^t H u > 0) \end{aligned}$$

□

### BEMERKUNG 9.1

$H$  positiv definit heißt, dass die quadratische Form  $u^t H u$ , die in der Taylorapproximation den quadratischen Term repräsentiert, positiv gekrümmt ist (d. h., die Steigung nimmt

– im eindimensionalen Fall – von „links nach rechts“ zu), also konvex ist.  $u^t H u > 0$  bedeutet, dass alle Eigenwerte der Matrix  $H = H_f(x_0)$  positiv sein müssen, da  $u^t H u > 0$  für alle Vektoren  $u$  gelten muss, also gilt speziell für Eigenvektoren  $v$  und den zugehörigen Eigenwert  $\lambda$

$$v^t H v = v^t \lambda v = \lambda v^t v = \lambda \|v\|^2 > 0,$$

also muss  $\lambda > 0$  sein, da die Länge  $\|v\|$  immer positiv ist.

Analog gilt für negativ definite Matrizen, dass alle Eigenwerte negativ sein müssen.

#### DEFINITION 9.1

Eine  $n \times n$ -Matrix  $A = (a_{ij})_{i,j=1}^n$  hat  $n$  Hauptminoren  $A_k$ ,  $k = 1, \dots, n$ .

Ein *Hauptminor*  $A_k$  ist definiert als  $A_k = (a_{ij})_{i,j=1}^k$ .

#### THEOREM 9.4 (KRITERIUM VON SYLVESTER)

Eine symmetrische Matrix ist

- positiv definit, wenn alle ihre Hauptminoren positiv sind.
- negativ definit, wenn ihre Hauptminoren alternierend negativ und positiv sind, d. h.,  $A_1 < 0$ ,  $A_2 > 0$ ,  $A_3 < 0$ ,  $A_4 > 0$ ,  $\dots$ ,  $A_{2k} > 0$ ,  $A_{2k+1} < 0$ ,  $\dots$

#### BEISPIEL 9.2

Suche die lokalen Minima der Funktion  $f(x_1, x_2, x_3) = x_1^3 + 2x_1^2 x_2 + x_1 x_2 x_3 + x_2^2 - 2x_2 + x_2 x_3^2$ . Dann ist

$$\nabla f \left( \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \right) = \begin{pmatrix} 3x_1^2 + 4x_1 x_2 + x_2 x_3 \\ 2x_1^2 + x_1 x_3 + 2x_2 - 2 + x_3^2 \\ x_1 x_2 + 2x_2 x_3 \end{pmatrix}$$

Setzen wir  $\nabla f(x_1, x_2, x_3) = 0$ , so erhalten wir:

$$\begin{array}{lll} \text{I :} & 3x_1^2 + 4x_1 x_2 + x_2 x_3 & = 0 \\ \text{II :} & 2x_1^2 + x_1 x_3 + 2x_2 - 2 + x_3^2 & = 0 \\ \text{III :} & x_1 x_2 + 2x_2 x_3 & = 0 \end{array}$$

Aus III :  $x_2(x_1 + 2x_3) = 0$  folgt nun, dass entweder (1)  $x_2 = 0$  oder (2)  $x_1 = -2x_3$ .

Fall(1)  $x_2 = 0$ : Durch Einsetzen folgt aus I, dass  $3x_1^2 = 0$  und daher  $x_1 = 0$ .

Andererseits folgt dann aus II, dass  $-2 + x_3^2 = 0$  und damit  $x_3 = \pm\sqrt{2}$ .

Wir erhalten nun zwei Kandidaten für Extrema:

$$x^{(1)} = \begin{pmatrix} 0 \\ 0 \\ +\sqrt{2} \end{pmatrix}, \quad x^{(2)} = \begin{pmatrix} 0 \\ 0 \\ -\sqrt{2} \end{pmatrix}$$

Fall(2)  $x_1 = -2x_3$ : Durch Einsetzen in I und die einfache Umformung:  $12x_3^2 - 7x_2 x_3 = x_3(12x_3 - 7x_2) = 0$  erhalten wir die zwei Unterfälle (2a)  $x_3 = 0$  bzw. (2b)  $x_3 = \frac{7}{12}x_2$ .

Fall(2a)  $x_3 = 0$ : Durch Einsetzen erhalten wir  $x_1 = 0$  und  $x_2 = 1$ , also einen neuen Kandidaten für ein Extremum

$$x^{(3)} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}.$$

Fall(2b)  $x_3 = \frac{7}{12}x_2$ :

Hier folgt, dass  $x_1 = -\frac{7}{12}x_2$ , und durch Einsetzen in II, dass  $\frac{343}{144}x_2^2 + 2x_2 - 2 = 0$ .

Diese quadratische Gleichung hat zwei Lösungen, die annähernd folgende Werte annehmen:  $-1.428$  und  $+0.588$ . Daraus ergeben sich die beiden letzten Kandidaten (mit ungefähren Koordinaten):

$$x^{(4)} = \begin{pmatrix} 1.666 \\ -1.428 \\ -0.833 \end{pmatrix}, \quad x^{(5)} = \begin{pmatrix} -0.686 \\ 0.588 \\ 0.343 \end{pmatrix}$$

Nun gilt es zu prüfen, ob die obigen Punkte Minima sind. Dazu berechnen wir die Hesse'sche Matrix der zweiten partiellen Ableitungen:

$$H(x) = \nabla^2(x) = \begin{pmatrix} 6x_1 + 4x_2 & 4x_1 + x_3 & x_2 \\ 4x_1 + x_3 & 2 & x_1 + 2x_3 \\ x_2 & x_1 + 2x_3 & 2x_2 \end{pmatrix}$$

Wir setzen nun die Punkte ein, also:

$$H(x^{(1)}) = \begin{pmatrix} 0 & \sqrt{2} & 0 \\ \sqrt{2} & 2 & 2\sqrt{2} \\ 0 & 2\sqrt{2} & 0 \end{pmatrix}, \text{ diese Matrix ist indefinit! } (0, -, \dots)$$

$$H(x^{(2)}) = \begin{pmatrix} 0 & -\sqrt{2} & 0 \\ -\sqrt{2} & 2 & -2\sqrt{2} \\ 0 & -2\sqrt{2} & 0 \end{pmatrix}, \text{ ist indefinit! } (0, -, \dots)$$

$$H(x^{(3)}) = \begin{pmatrix} 4 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 2 \end{pmatrix}, \text{ ist positiv definit } (+, +, +), \text{ also ist } x^{(3)} \text{ lokal minimal.}$$

$$H(x^{(4)}) = \begin{pmatrix} 1.121 & 5.831 & -1.428 \\ 5.831 & 2 & 0 \\ -1.428 & 0 & -2.856 \end{pmatrix}, \text{ ist indefinit! } (+, -, \dots)$$

$$H(x^{(5)}) = \begin{pmatrix} -1.764 & -2.401 & 0.588 \\ -2.401 & 2 & 0 \\ 0.588 & 0 & 1.176 \end{pmatrix}, \text{ ist indefinit! } (-, +, +)$$

Das einzige lokale Minimum ist daher  $x^{(3)} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$  mit  $f(x^{(3)}) = f\left(\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}\right) = -1$ .

Achtung: Durch einfache Überlegungen können wir feststellen, dass  $x^{(3)}$  zwar lokale, aber keine globale Minimalstelle ist:  $f\left(\begin{pmatrix} \lambda \\ 0 \\ 0 \end{pmatrix}\right) = \lambda^3 \rightarrow -\infty, \lambda \rightarrow -\infty$ .

## 9.2 Einfache numerische Optimierungsverfahren

Wir können unser Wissen über Gradienten und den Taylor'schen Lehrsatz auch für iterative Näherungsverfahren zur Bestimmung von Extremwerten verwenden:

### Methoden des steilsten Abstiegs (steepest descent)

Algorithmus:

(1) Wähle ein  $x_1 = \begin{pmatrix} x_{11} \\ \vdots \\ x_{1n} \end{pmatrix}$  als Startpunkt.

(2) Für  $k = 1, 2, 3, \dots$  berechne die Richtung  $d_k := -(\nabla f(x_k))^t$ .  
Minimiere  $\varphi(\lambda) = f(x_k + \lambda d_k)$  durch „eindimensionale Optimierung nach  $\lambda$ .“

Wir erhalten einen optimalen Wert  $\lambda^*$ !

Als nächsten Punkt wählen wir  $x_{k+1} = x_k + \lambda^* d_k$ .

Zunächst gehen wir wie Wasser vor: Wir wählen die Richtung des steilsten Abstiegs. Diese Richtung behalten wir so lange bei, bis die Funktion  $f$  (diese entspricht der „Höhe“) wieder anzusteigen beginnt. Dort sehen wir uns erneut nach dem steilsten Abstieg um (Gradient und Höhengichtlinie sind dort ja immer orthogonal zueinander).

Warum können wir es nicht wirklich genauso wie das Wasser machen? Wir müssten unendlich viele Neuberechnungen der Richtung machen (die Schrittweite müsste infinitesimal klein werden) und wir würden dadurch eine unendliche Rechenzeit erreichen.

Unser Optimierungsverfahren ist hingegen schon recht effizient: Anfangs machen wir große Schritte, erst wenn wir uns dem Optimum nähern, werden die Schritte kleiner.

Das *Steepest descent*-Verfahren verläuft immer in aufeinander folgenden, orthogonalen Richtungen (siehe Abbildung 9).

#### BEISPIEL 9.3

Sei für folgende Funktion  $f(x, y) = x^2 + y^2 - 2x - 6y$  ein Minimum zu bestimmen.

Wir wählen zunächst  $x_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$  als Startpunkt.

Wir berechnen  $(\nabla f(x))^t = \begin{pmatrix} 2x - 2 \\ 2y - 6 \end{pmatrix}$ . Dann ist  $d_1 = -(\nabla f(x_1))^t = -\begin{pmatrix} -2 \\ -6 \end{pmatrix}$  und wir

erhalten:  $\varphi(\lambda) = f(x_1 + \lambda d_1) = f\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix} + \lambda \begin{pmatrix} 2 \\ 6 \end{pmatrix}\right) = 4\lambda^2 + 36\lambda^2 - 4\lambda - 36\lambda = 40\lambda^2 - 40\lambda$

Abgeleitet nach  $\lambda$  erhalten wir  $\varphi'(\lambda) = 80\lambda - 40$ . Die Funktion  $\varphi(\lambda)$  hat also ein Minimum in  $\lambda^* = \frac{1}{2}$ .

Damit ergibt sich der nächste Punkt als  $x_2 = x_1 + \lambda^* d_k = \begin{pmatrix} 0 \\ 0 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} 2 \\ 6 \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$ .

Dieser ist auch schon das Minimum.

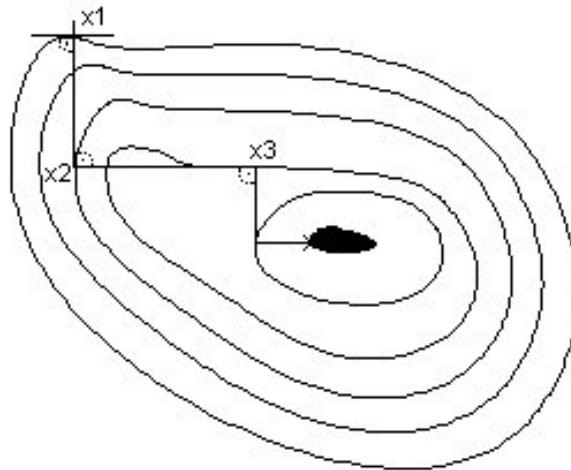


Abbildung 9: Steepest Descent.

### Das mehrdimensionale Newtonverfahren

Wir approximieren die Funktion  $f(x)$  an der Stelle  $x_0$  durch eine quadratische Funktion

$$h(x) = a + b^t x + \frac{1}{2} x^t G x.$$

Es muss dann gelten:

$$\begin{aligned} f(x_1) &= h(x_1) = a + b^t x_1 + \frac{1}{2} x_1^t G x_1 \\ \nabla f(x_1) &= \nabla h(x_1) = b^t + x_1^t G \\ \nabla^2 f(x_1) &= \nabla^2 h(x_1) = G \end{aligned}$$

Wir erhalten daraus:  $b^t = \nabla f(x_1) - x_1^t \nabla^2 f(x_1)$ .

Wenn wir nun das Minimum  $x^*$  von  $h(x)$  berechnen, erhalten wir eine Näherung für das Minimum von  $f$ .

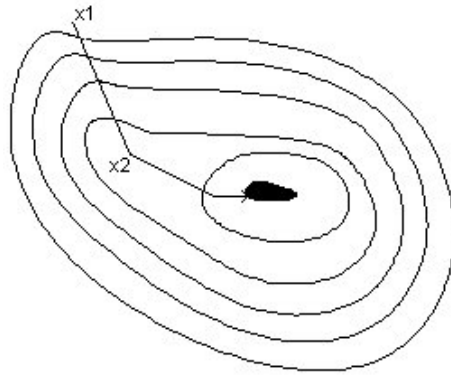
Wir setzen wieder  $\nabla h(x) = 0$  und erhalten:

$$b^t + x^{*t} G = 0 \text{ und daraus } x^* = x_1 - (\nabla^2 f(x_1))^{-1} (\nabla f(x_1))^t.$$

Daraus leitet sich der Algorithmus von Newton und Raphson für den mehrdimensionalen Fall ab.

### Algorithmus nach Newton–Raphson:

- (1) Wähle ein  $x_1 = \begin{pmatrix} x_{11} \\ \vdots \\ x_{1n} \end{pmatrix}$  als Startpunkt.
- (2) Wiederhole  $x_{k+1} = x_k - (\nabla^2 f(x_k))^{-1} (\nabla f(x_k))^t$  bis die gewünschte Genauigkeit erreicht ist.



**Abbildung 10:** Newton-Raphson.

Vorteil: Wir gehen gleich in eine „gute Richtung“ (bei quadratischer Zielfunktion genau in Richtung des Minimums), statt einen Zick-Zack-Kurs zu steuern.

Nachteile:

- Die zweiten Ableitungen müssen bekannt sein.
- Die Hesse'sche Matrix muss immer positiv-definit sein (entspricht strikter Konvexität)
- Der Funktionswert kann zwischendurch sogar ansteigen (das ist bei *steepest descent* nie der Fall), dies können wir aber durch Kombination mit dem *line search* Verfahren vermeiden.
- Die Inversion der Hesse'schen Matrix kann numerisch aufwändig sein.

Das Newton-Raphson Verfahren kann also nur dann angewendet werden, wenn die zweiten Ableitungen bekannt bzw. berechenbar sind.

### 9.3 Optimierung mit Nebenbedingungen

Zusätzlich zur Zielfunktion  $f$  müssen wir noch Restriktionen für die zulässigen Werte betrachten, d. h., es sind nur Werte aus einer Teilmenge des  $\mathbb{R}^n$  zur Konkurrenz zugelassen ( $x \in S \subset \mathbb{R}^n$ ).

Diese Restriktionen könnten in unterschiedlicher Weise beschrieben werden. Im Folgenden gehen wir davon aus, dass unsere Restriktionen mit Hilfe von Gleichungen und Ungleichungen beschrieben werden können.



Unser Optimierungsproblem lässt sich also folgendermaßen formulieren:

$$\begin{aligned} f(x) &\rightarrow \min! \\ g_1(x) &\leq 0 \\ &\vdots \\ g_m(x) &\leq 0 \\ h_1(x) &= 0 \\ &\vdots \\ h_r(x) &= 0 \end{aligned}$$

Wir haben hier ein Minimierungsproblem mit  $m$  Ungleichheits- und  $r$  Gleichheitsnebenbedingungen beschrieben, wobei die Funktionen  $g_i(x)$ ,  $i = 1, \dots, m$  und  $h_j(x)$ ,  $j = 1, \dots, r$  im Prinzip beliebige Funktionen sein können.

#### BEISPIEL 9.4

Es sei ein Wartehäuschen mit folgenden Vorgaben zu konstruieren: Die Rückwand darf höchstens 3m lang sein. Die Gesamtlänge der Wände (Rück- und Seitenwände) soll gleich 4m sein. Die überdachte Fläche ist zu maximieren.

Als mathematische Formulierung erhalten wir dann daraus ( $x_1$  ist die Länge der Hinterwand,  $x_2$  ist die Länge der Seitenwände):

$$\begin{aligned} f(x_1, x_2) &= x_1 x_2 &&\rightarrow \max! \\ g(x_1, x_2) &= x_1 &&\leq 3 \\ h(x_1, x_2) &= x_1 + 2x_2 &= 4 \end{aligned}$$

### Ein wichtiger Spezialfall: Nur Gleichheitsnebenbedingungen

#### THEOREM 9.5

Es sei  $f(x_1, \dots, x_n) : \mathbb{R}^n \rightarrow \mathbb{R}$  stetig differenzierbar.

Der zulässige Bereich  $S$  sei durch Gleichungen beschrieben

$$S = \left\{ x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \mid g_1(x) = 0, \dots, g_m(x) = 0 \right\},$$

die Funktionen  $g_i$  seien stetig differenzierbar und die Matrix der partiellen Ableitungen der  $g_i$  ( $m \times n$ -Matrix) habe den Rang  $m$ ,  $m < n$ .

Dann ist für ein lokales Extremum von  $f\left(\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}\right)$  mit  $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in S$  notwendig, dass

$\nabla f(x) = \sum_{i=1}^m \lambda_i \nabla g_i(x)$ , mit eindeutig bestimmten Koeffizienten  $\lambda_i$  (diese heißen Lagrange'sche Multiplikatoren).

Wir gehen so vor, dass wir die **Lagrange-Funktion**

$$L(x_1, \dots, x_n, \lambda_1, \dots, \lambda_m) = f(x_1, \dots, x_n) - \sum_{i=1}^m \lambda_i g_i(x_1, \dots, x_n)$$

bilden und sowohl nach  $x_1, \dots, x_n$  als auch nach  $\lambda_1, \dots, \lambda_m$  ableiten und die Ableitungen gleich 0 setzen.

Danach müssen wir das entstandene (im Allgemeinen nichtlineare) Gleichungssystem in  $n + m$  Variablen zu lösen.

#### BEISPIEL 9.5

$$\begin{aligned} f(x_1, x_2) &= x_1 x_2 && \rightarrow \max! \\ g(x_1, x_2) &= x_1^2 + x_2^2 - 1 &= 0 \end{aligned}$$

Die Lagrange-Funktion ist dann:

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

Wir leiten nach allen Variablen und  $\lambda$  ab und setzen gleich 0:

$$\begin{aligned} \text{I : } \quad \frac{\partial f}{\partial x_1} &= x_2 - 2\lambda x_1 &= 0 \\ \text{II : } \quad \frac{\partial f}{\partial x_2} &= x_1 - 2\lambda x_2 &= 0 \\ \text{III : } \quad \frac{\partial f}{\partial \lambda} &= -(x_1^2 + x_2^2 - 1) &= 0 \end{aligned}$$

Falls  $x_1 = 0$ , dann folgt aus I, dass auch  $x_2 = 0$ ; das steht aber im Widerspruch zu III. Wir schließen daraus, dass  $x_1 \neq 0$ .

Ähnliches gilt auch für  $x_2$ , es muss also auch  $x_2 \neq 0$  sein.

Falls aber nun  $x_1 \neq 0$ , dann erhalten wir durch Umformung aus I, dass  $2\lambda = \frac{x_2}{x_1}$ . Aus II folgt dann (wegen  $x_2 \neq 0$ ), dass  $2\lambda = \frac{x_1}{x_2}$ . Daraus schließen wir, dass  $x_1 = \pm x_2$ . Eingesetzt in III, erhalten wir dann  $x_1 = \pm \frac{\sqrt{2}}{2} = \pm \frac{1}{\sqrt{2}}$ . Also erhalten wir schließlich 4 Kandidaten für lokale Extremwerte:

$$x^{(1)} = \begin{pmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{pmatrix}, x^{(2)} = \begin{pmatrix} \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} \end{pmatrix}, x^{(3)} = \begin{pmatrix} -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{pmatrix}, x^{(4)} = \begin{pmatrix} -\frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} \end{pmatrix}$$

mit Funktionswerten  $\frac{1}{2}$  bzw.  $-\frac{1}{2}$ .

Wenn wir nicht nur Gleichheits- sondern auch Ungleichheitsnebenbedingungen haben, dann gibt es ein ähnliches, aber etwas komplizierteres Verfahren – das **Karush-Kuhn-Tucker-Verfahren** – mit dem auch solche komplexeren Fälle behandelbar sind.