Differenzierbarkeit als lokale affin lineare Approximierbarkeit und lokale Extrema

Eindimensionaler Fall

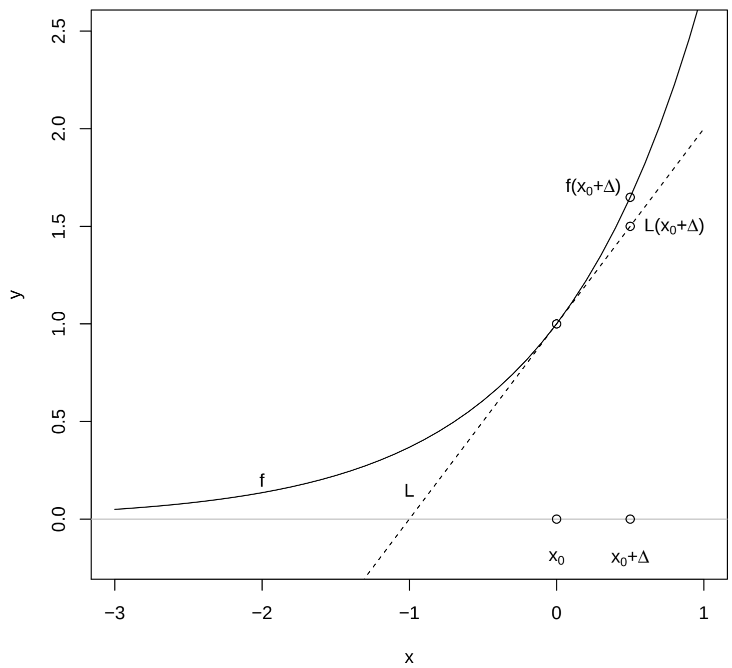
Ist eine Funktion $f : \mathbb{R} \longrightarrow \mathbb{R}$ differenzierbar in einem Punkt x_0, so ist f

$\underbrace{\text{in einer Umgebung von } x_0}_{\text{lokal}}$ affin linear approximierbar,

d.h. für hinreichend kleine Δ kann f durch die affin lineare Funktion

$$L : \mathbb{R} \longrightarrow \mathbb{R} : x_0 + \Delta \mapsto f(x_0) + f'(x_0) \cdot \Delta$$

approximiert werden.



Dabei ist der Fehler $R(\Delta) = f(x_0 + \Delta) - L(x_0 + \Delta)$ der affin linearen Approximation für sehr kleine Δ vernachlässigbar, genauer gilt

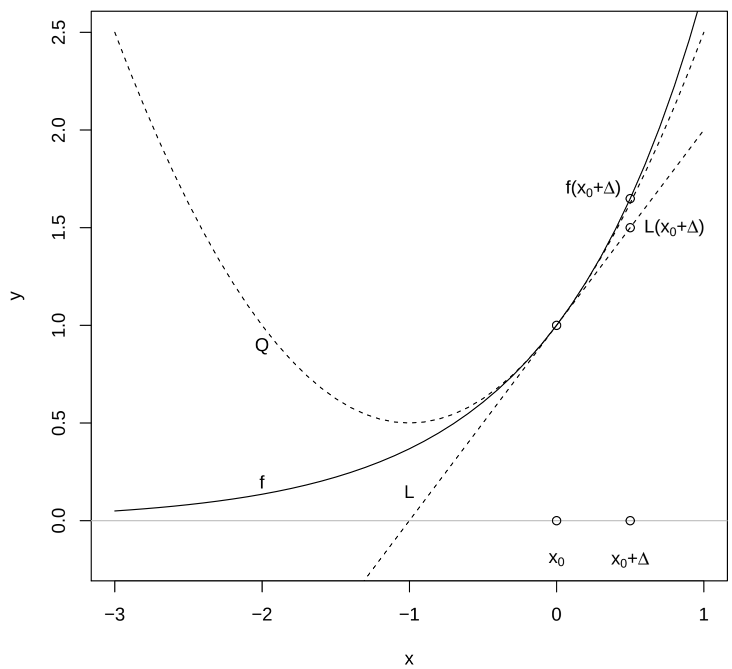
$$\lim_{\Delta \to 0} \frac{f(x_0 + \Delta) - L(x_0 + \Delta)}{\Delta} = 0,$$

d.h. der Fehler geht schneller als linear gegen Null, wenn Δ gegen Null geht.

Bezieht man auch die zweite Ableitung von f mit ein, so kann man die Funktion f auch mit einer quadratischen Funktion

$$Q(x_0 + \Delta) = f(x_0) + f'(x_0)\Delta + \frac{f''(x_0)}{2} \cdot \Delta^2$$

approximieren:



Dies macht verständlich, warum man die 2. Ableitung für die Untersuchung auf Extrema verwendet:

Gilt $f'(x_0) = 0$, so gilt in einer Umgebung um x_0 (also falls Δ hinreichend klein):

$$
\begin{aligned}
f(x_0 + \Delta) &\approx f(x_0) + \underbrace{f'(x_0) \cdot \Delta}_{=0} + \frac{f''(x_0)}{2} \cdot \Delta^2 \\
&= f(x_0) + f''(x_0) \cdot \Delta^2 \begin{cases} > f(x_0) & \text{falls } f''(x_0) > 0 \quad (\Longrightarrow \text{ lokales Minimum}) \\ < f(x_0) & \text{falls } f''(x_0) < 0 \quad (\Longrightarrow \text{ lokales Maximum}) \end{cases}
\end{aligned}
$$

Bemerkung:

Allgemein kann man unter Einbeziehung von höheren Ableitungen die Funktion f durch ein Polynom k-ten Grades approximieren:

$$f(x_0 + \Delta) \approx f(x_0) + \sum_{i=1}^{k} \frac{f^{(i)}(x_0)}{i!} \cdot \Delta^k.$$

Dies wird als Taylorapproximation bezeichnet.

Mehrdimensionaler Fall

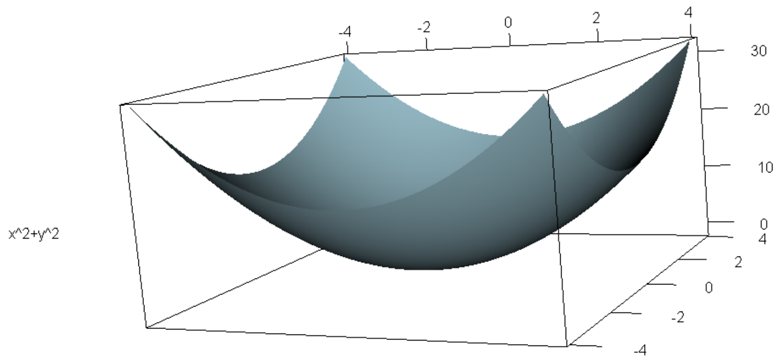
Sei $f : \mathbb{R}^p \longrightarrow \mathbb{R}^q$ eine Funktion. Dann heißtf total differenzierbar im Punkt $x_0 \in \mathbb{R}^p$, wenn es eine affin lineare Abbildung $L : \mathbb{R}^p \longrightarrow \mathbb{R}^q$ gibt mit:

$$\lim_{||\Delta|| \to 0} \frac{f(x_0 + \Delta) - L(x_0 + \Delta)}{||\Delta||} = 0,$$

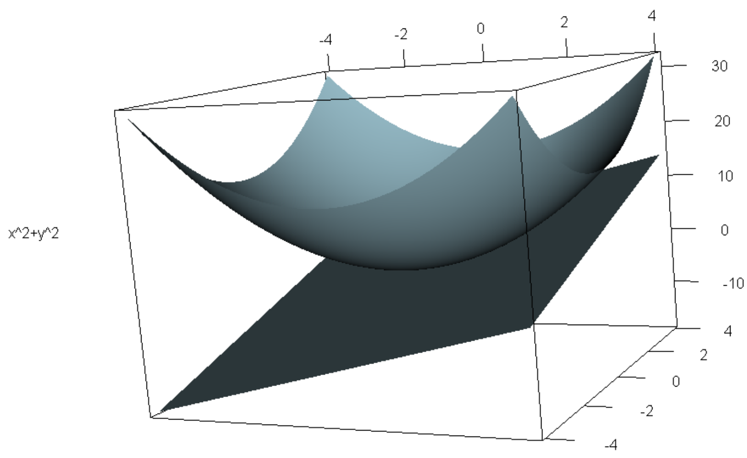
d.h. wenn sich f affin linear approximieren lässt mit einem Fehler, der stärker als linear gegen 0 geht.

Ist f total differenzierbar, so ist die gesuchte affin lineare Abbildung im Wesentlichen durch die Jacobi-matrix (an der Stelle x_0) gegeben:

$$L(x_0 + \Delta) = f(x_0) + J \cdot \Delta.$$



Paraboloid $f(x, y) = x^2+y^2$



Paraboloid $f(x, y) = x^2 + y^2$ mit approximierender linearer Funktion (Ebene)

Achtung:

Δ ist jetzt ein p-dimensionaler Vektor.

Lokale Extrema im mehrdimensionalen Fall

Im Folgenden betrachten wir jetzt speziell Funktionen von $\mathbb{R}^p$ nach $\mathbb{R}$. Analog zum eindimensionalen Fall kann die Analyse von lokalen Extrema wieder über die quadratische Approximation von f erfolgen. Dazu nutzen wir die Jakobimatirx J (die dann eine $1 \times p$ - Matrix ist, also letztendlich nichts anderes als $(\bigtriangledown f)^T$) und die Hessematrix H an der Stelle x_0:

$$f(x_0 + \Delta) \approx f(x_0) + J\Delta + \frac{1}{2}\Delta^T H \Delta.$$

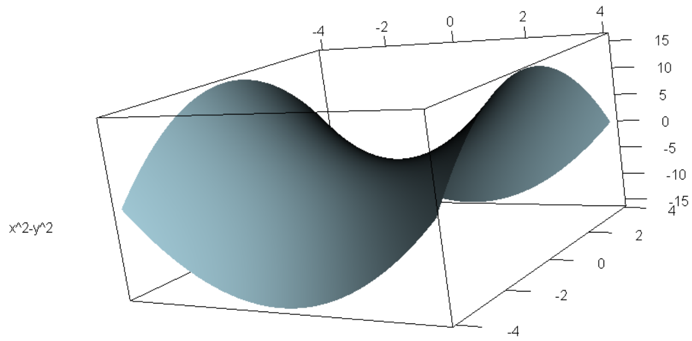
Analog zum eindimensionalen Fall muss für ein lokales Extremum an der Stelle x_0 die Jacobimatrix Null sein. Also folgt:

$$f(x_0 + \Delta) \approx f(x_0) + \frac{1}{2}\Delta^T H \Delta.$$

Jetzt ist aber für den letzte Term

$$\frac{1}{2}\Delta^T H \Delta = \frac{1}{2}\left(H_{11}\Delta_1^2 + H_{12}\Delta_1\Delta_2 + H_{21}\Delta_2\Delta_1 + \dots H_{pp}\Delta_p^2\right)$$

nicht so leicht zu analysieren, ob dieser immer größer bzw. kleiner als 0 ist.



Sattelfläche $f(x, y) = x^2 - y^2$

Definition
Sei A eine symmetrische $p \times p$-Matrix.

1. A heißt positiv definit, wenn für alle Vektoren $v \neq 0$ gilt: $v^t Av > 0$.
2. A heißt positiv semidefinit, wenn für alle Vektoren $v \neq 0$ gilt: $v^t Av \geq 0$.
3. A heißt negativ definit, wenn für alle Vektoren $v \neq 0$ gilt: $v^t Av < 0$.
4. A heißt negativ semidefinit, wenn für alle Vektoren $v \neq 0$ gilt: $v^t Av \leq 0$.
5. Wenn Vektoren $v_1, v_2 \in \mathbb{R}^p$ existieren mit $v_1^T Av_1 > 0$ und $v_2^T Av_2 < 0$, dann heißt A indefinit.

Satz (Notwendiges Kriterium, vgl. Satz 15.2, S.276)
Sei $U \subseteq \mathbb{R}^p$ offen und $f : U \longrightarrow \mathbb{R}$ zweimal stetig differenzierbar. Liegt bei $x_0 \in U$ ein lokales

Extremum vor, so gilt notwendigerweise $\bigtriangledown f(x_0) = 0$.

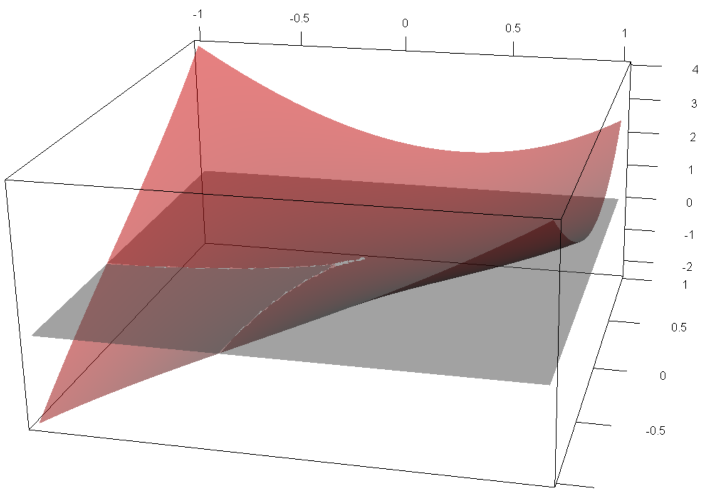
Satz (Hinreichendes Kriterium vgl. Satz 15.4, S.278)
Sei $U \subseteq \mathbb{R}^p$ offen und $f : U \longrightarrow \mathbb{R}$ zweimal stetig differenzierbar und gelte weiter $\bigtriangledown f(x_0) = 0$ für $x_0 \in U$.

1. Ist die Hesse-Matrix von f an der Stelle x_0 negativ (positiv) definit, so liegt in x_0 ein lokales Maximum (Minimum) vor.

2. Ist die Hesse-Matrix von f an der Stelle x_0 indefinit, so liegt kein lokales Extremum vor.

Bemerkungen:

- Im mehrdimensionalen Fall ist es nicht mehr so einfach möglich, die Funktion in der Umgebung einer potentiellen Extremstelle x_0 mit $(\nabla f)(x_0) = 0$ zu betrachten um sicherzustellen, dass es sich wirklich um einen Extrempunkt handelt. Während man im eindimensionalen Fall nur links und rechts vom potentiellen Extremum die Funktion analysieren müsste, hätte man im mehrdimensionalen Fall unendlich viele Richtungen zu untersuchen, es würde auch nicht ausreichen, alle Hauptrichtungen $e_1, \ldots, e_p$ zu analysieren, wie folgendes Beispiel zeigt:

  Dargestellt ist in roter Farbe die Funktion $f(x, y) = (x-y)^2+0.3\cdot(x+y)^3$, deren Gradient im Punkt $(0, 0)$ verschwindet. Grau hinterlegt ist die Funktion $g = 0$. Man sieht, dass, ausgehend vom Punkt $(0, 0)$, die Funktion sowohl in den Hauptrichtungen $(1, 0)$ bzw $(-1, 0)$, als auch in den Hauptrichtungen $(0, 1)$ bzw. $(0, -1)$ wächst, nicht jedoch z.B. in die Richtung $\frac{1}{\sqrt{2}}(-1, -1)$.

- Wie man auf Definitheit testet wird in den Sätzen 12.1 bis 12.3 (S.205-207) bzw. explizit in Korollar 15.1 (S.278) beschrieben:

$$\vdots$$

Extrema unter Nebenbedingungen

Problemstellung: Maximiere oder minimiere eine Funktion $f : \mathbb{R}^p \longrightarrow \mathbb{R}$ unter gewissen Nebenbedingungen, d.h. die Menge der zulässigen Argumente, für die die Funktion f analysiert wird, ist eingeschränkt. Die zulässige Menge wird im Folgenden mit Z bezeichnet.

Arten von Nebenbedingungen:

1. Ungleichungsnebenbedingungen: $g(v) \leq c$, wobei $v \in \mathbb{R}^p$ und c eine Konstante.

   Beispiel:
   Ein Schäfer hat $100m$ Zaun um eine rechteckige Fläche für seine Schafe abzustecken. Er will die Fläche, die den Schafen zur Verfügung steht, maximieren, Da er ja nur $100m$ Zaun besitzt muss der Umfang der abgesteckten Rechteckfläche kleinergleich $100m$ sein. Die Nutzenfunktion wäre gegeben durch

   $$f : \mathbb{R}^2 \longrightarrow \mathbb{R} : v = (l, b) \mapsto l \cdot b,$$

   wobei l die Länge und b die Breite der abgesteckten Fläche beschreiben. Die Nebenbedingung lautet dann: $g(v) = 2l + 2b \leq 100$.

2. Gleichheitsnebenbedingungen: $g(v) = c$.

   Beispiel:
   Ein Verkehrsleitsystem soll einen Verkehrsstrom von $i = 100$ Pkw pro Minute auf insgesamt 3 verschiedene Umleitungen aufteilen um den Verkehrsfluss zu minimieren. Die Nebenbedingung würde hier lauten : $g(v) = i_1 + i_2 + i_3 = 100$, wobei i_1 bis i_3 den Verkehrsstrom auf Umleitung 1 bis 3 beschreibe.

Bemerkungen:

- Oft kann eine Ungleichungsnebenbedingung in eine Gleichungsnebenbedingung umgewandelt werden: in obigem Beispiel ist leicht zu erkennen, dass die maximale Weidefläche nur bei Verwendung der gesamten Zaunlänge zu erreichen ist, es kann also auch mit der Gleichungsnebenbedingung $g(v) = 2l + 2v = 100$ gearbeitet werden.

- Eine Gleichheitsnebenbedingung $\tilde{g}(v) = c$ kann immer in die Form $g(v) = 0$ gebracht werden, indem die Konstante c auf die linke Seite gebracht wird, also: $g(v) := \tilde{g}(v) - c$.

- Es ist auch möglich, dass mehrere Nebenbedingungen vorliegen, z.B. $g_1(v) = g_2(v) = \ldots = g_k(v) = 0$.

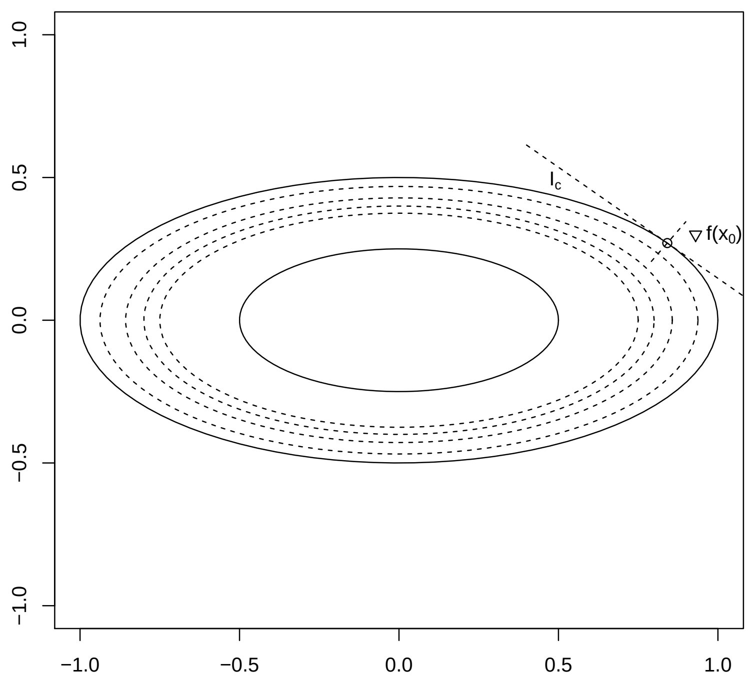
- Oft wird implizit gefordert, dass der Vektor v bzw. einzelne Komponenten von v nichtnegativ sind (z.B. bei Mengen oder Anzahlen. Dies ist dann bei der Lösung zu überprüfen bzw. als Ungleichungsnebenbedingung mit aufzunehmen.)

Bestimmung potentieller lokaler Extrema unter Nebenbedingungen mittels des Verfahrens von Lagrange

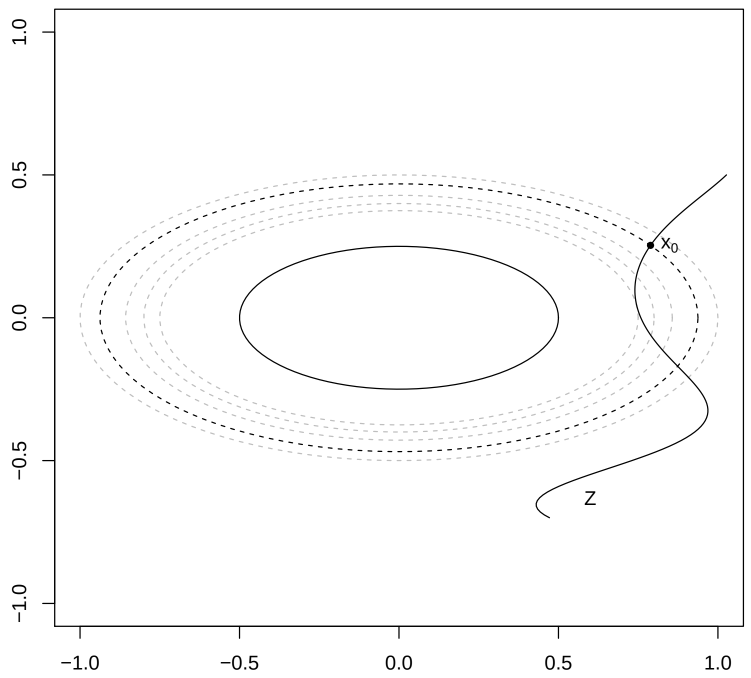
Vorbemerkungen zum Zusammenhang zwischen Höhenlinien und Gradient:

- Entlang der Höhenlinien ist die Funktion konstant.
- Um so höher die Änderung der Funktion in eine Koordinatenrichtung e_i (in einem Punkt x_0), desto grösser der entsprechende Eintrag $(\bigtriangledown f(x_0))_i$ des Gradienten.
- Die Richtung des Gradienten beschreibt die Richtung des stärksten Anstieges der Funktion f im Punkt x_0.
- Die Richtung der Höhenlinie in einem Punkt x_0 steht immer senkrecht zum Gradienten im Punkt x_0.

Illustration:



Idee zur Herleitung notwendiger Bedingungen:



- Wenn man auf der Kurve Z der zulässigen Punkte entlang geht, dann kann bei $x_0 \in Z$ nur ein lokales Extremum vorliegen, wenn die durch den Punkt x_0 verlaufende Isolinie von f parallel zu dieser Kurve verläuft, ansonsten könnte man noch ein Stückchen entlag der Kurve Z entlang laufen und könnte so die Zielfunktion f noch ein bischen vergrößern bzw. verkleinern.

- Die Menge Z ist nichts anderes als die Isolinie der Funktion g zum Niveau c.

- Die Isolinien von g und f müssen also in einem Extrempunkt x_0 parallel verlaufen.

- Dies ist gleichbedeutend damit, dass die Gradienten $\bigtriangledown f(x_0)$ und $\bigtriangledown g(x_0)$ parallel sind, d.h. es muss eine Konstante $\lambda \in \mathbb{R}$ geben mit:

$$\bigtriangledown f(x_0) = \lambda \bigtriangledown g(x_0).$$

- Zusätzlich muss natürlich weiterhin die Nebenbedingung $g(x_0) = c$ gelten.

Vorgehen zur Ermittlung von Punkten x_0, die die obengenannten notwendigen Bedingungen erfüllen (Verfahren nach Lagrange)

Gegeben Sei das Optimierungsproblem:
Maximiere/minimiere die Funktion $f : \mathbb{R}^p \longrightarrow \mathbb{R}$ unter der Nebenbedingung $g(v) = 0$.

1. Falls eine Ungleichungsnebenbedingung gegeben ist, versuche diese in eine Gleichungsnebenbedingung umzuwandeln.

2. Eine Gleichungsnebenbedingung $\tilde{g}(v) = c$ kann immer in die Form $g(v) = 0$ gebracht werden indem man das c auf die linke Seite bringt, also $g(v) := \tilde{g}(v) - c$.

3. Stelle die Lagrangefunktion auf:

$$L(v, \lambda) = f(v) - \lambda g(v)$$

4. Bestimme den Gradienten $\bigtriangledown L$

5. Finde alle Punkte (x_0, λ) mit $\bigtriangledown L(x_0, \lambda) = 0$, d.h. löse ein Gleichungssystem mit $p + 1$ Gleichungen. Die ersten p Gleichungen stellen sicher, dass $\bigtriangledown f(x_0) = \lambda \bigtriangledown g(x_0)$ gilt und die letzte Gleichung $\frac{\partial L}{\partial \lambda}(x_0) = 0$ stellt sicher, dass die Nebenbedingung erfüllt ist, denn $\frac{\partial L}{\partial \lambda} = -g$.

Bemerkung:
Dieses Verfahren funktioniert auch analog für mehrere Nebenbedingungen, vergleiche dazu das Vorgehen auf S.286:

$$\vdots$$