Notwendige Bedingung für Extrempunkt unter Nebenbedingungen

Theorem 0.1 *Sei $\Omega \subset \mathbb{R}^n$ und $f, g_1, \ldots, g_m \in C^1(\Omega)$ mit $m < n$. Sei $N_g = \{x \in \mathbb{R}^n \,|\, g_1(x) = \ldots = g_m(x) = 0\}$ eine Lösungsmannigfaltigkeit und f habe unter der Bedingung $g_1 = \ldots = g_m = 0$ an der Stelle $c \in N_g$ ein lokales Extremum. Dann gibt es eindeutig bestimmte Zahlen $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$, so daß*

$$\nabla f(c) - \sum_{j=1}^{m} \lambda_j \nabla g_j(c) = 0,$$

d.h. die Funktion $f - \sum_{j=1}^{m} \lambda_j \, g_j$ hat an der Stelle c einen stationären Punkt.

Proof: Nach etwaiger Umordnung und Umbennung der Variablen kann die Lösungsmannigfaltigkeit N_g durch die Gleichung

$$g(x, y) = 0, \quad \text{mit} \quad x \in \mathbb{R}^{n-m},\ y \in \mathbb{R}^m,$$

beschrieben werden (y ist der Teil der Variablen nach denen wir auflösen). Nach dem Satz über implizite Funktionen gibt es daher eine Umgebung $U \times V$ ($U \subset \mathbb{R}^{n-m}$, $V \subset \mathbb{R}^m$) von $c = (a, b)$ ($a \in \mathbb{R}^{n-m}, b \in \mathbb{R}^m$) und eine stetig differenzierbare Abbildung $\phi : U \to V$ mit $\phi(a) = b$ und

$$g(x, y) = 0 \quad \Longleftrightarrow \quad y = \phi(x) \quad \text{für} \quad (x, y) \in U \times V.$$

Es gilt daher $g_k(x, \phi(x)) = 0$ in U für $k = 1, \ldots, m$. Wir definieren $h : U \to \mathbb{R}^n$ mit

$$h(x) = \begin{pmatrix} x \\ \phi(x) \end{pmatrix} \quad \text{und } \tilde{g}(x) := g(x, \phi(x)) = g(h(x)), \qquad \tilde{f}(x) := f(x, \phi(x)) = f(h(x)).$$

Für die Jacobimatrix J_h von h gilt für $x \in U$,

$$\frac{\partial h_k}{\partial x_i}(x) = \begin{cases} 1 & \text{wenn } k = i \text{ und } 1 \leq \ k \ \leq n - m, \\ 0 & \text{wenn } k \neq i \text{ und } 1 \leq \ k \ \leq n - m, . \\ \frac{\partial \phi_{k-(n-m)}}{\partial x_i}(x) & \text{wenn } n - m + 1 \leq k \leq n \end{cases}$$

Aus $\tilde{g}(x) = g(x, \phi(x)) = 0$ für $x \in \mathcal{U}$ folgt mit der Kettenregel insbesondere für $a \in U$,

$$0 = J_{\tilde{g}}(a) = J_g(h(a))\, J_h(a) = \begin{pmatrix} [(\nabla g_1)(h(a))]^T \\ \vdots \\ [(\nabla g_m)(h(a))]^T \end{pmatrix} \begin{pmatrix} \mathbb{1}_{n-m} \\ [\nabla \phi_1(x)]^T \\ \vdots \\ [\nabla \phi_m(x)]^T \end{pmatrix},$$

wobei $\mathbb{1}_{n-m} \in \mathbb{R}^{(n-m)\times(n-m)}$ die Einheitsmatrix ist. Es folgt daher mit $h(a) = c$ für die (k, l)-Komponente von $J_{\tilde{g}}(a)$ mit $z = (x, y)$ und $1 \leq k \leq m$, $1 \leq l \leq n - m$,

$$\begin{aligned} 0 &= \frac{\partial g_k}{\partial z_l}(c) + \sum_{i=n-m+1}^{n} \frac{\partial g_k}{\partial z_i}(c)\, \frac{\partial \phi_{i-(n-m)}}{\partial x_l}(a) \\ &= \langle e_l, \nabla g_k(c) \rangle + \sum_{i=n-m+1}^{n} \frac{\partial \phi_{i-(n-m)}}{\partial x_l}(a)\ \langle e_i, \nabla g_k(c) \rangle \\ &= \left\langle e_l + \sum_{i=n-m+1}^{n} \frac{\partial \phi_{i-(n-m)}}{\partial x_l}(a) e_i, \nabla g_k(c) \right\rangle \end{aligned}$$

Da nach Voraussetzung c ein lokales Extremum von f unter den Nebenbedingungen $g_1 = \ldots = g_m = 0$ ist, gilt

$$0 = J_{\tilde{f}}(a) = J_f(h(a))\, J_h(a) = [\nabla f(c)]^T \begin{pmatrix} \mathbb{1}_{n-m} \\ [\nabla \phi_1(x)]^T \\ \vdots \\ [\nabla \phi_m(x)]^T \end{pmatrix}.$$

Für die l-te Komponente von $J_{\tilde{f}}(a) \in \mathbb{R}^{n-m}$ gilt daher

$$
\begin{aligned}
0 &= \frac{\partial f}{\partial z_l}(c) + \sum_{i=n-m+1}^{n} \frac{\partial f}{\partial z_i}(c) \, \frac{\partial \phi_{i-(n-m)}}{\partial x_l}(a) \\
&= \langle e_l, \nabla f(c) \rangle + \sum_{i=n-m+1}^{n} \frac{\partial \phi_{i-(n-m)}}{\partial x_l}(a) \; \langle e_i, \nabla f(c) \rangle \\
&= \left\langle e_l + \sum_{i=n-m+1}^{n} \frac{\partial \phi_{i-(n-m)}}{\partial x_l}(a) e_i, \nabla f(c) \right\rangle
\end{aligned}
$$

Wir definieren $(n-m)$ Vektoren $v_l \in \mathbb{R}^m$,

$$
v_l := e_l + \sum_{i=n-m+1}^{n} \frac{\partial \phi_{i-(n-m)}}{\partial x_l}(a) \, e_i, \quad l = 1, \ldots, n-m.
$$

Dann gilt

$$
\langle v_l, \nabla f(c) \rangle = 0, \text{ und } \langle v_l, \nabla g_k(c) \rangle = 0, \qquad l = 1, \ldots, n-m, \quad k = 1, \ldots, m.
$$

Daher liegen die Vektoren $\nabla f(c), \nabla g_1(c), \ldots, \nabla g_m(c)$ im Orthogonalraum von $W := \operatorname{span}\{v_1, \ldots, v_{n-m}\}$. Die Vektoren $v_1, \ldots, v_{n-m}$ sind linear unabhängig ($A := (v_1, \ldots, v_{n-m}) = J_h(a)$ und $J_h(a)$ hat maximalen Rang $(n-m)$) und damit gilt $\dim W = n - m$ und $\dim W^\perp = m$ (zur Erinnerung: $\mathbb{R}^n = W \oplus W^\perp$).

Die Vektoren $\nabla g_1(c), \ldots, \nabla g_m(c)$ sind nach Voraussetzung (N_g ist Lösungsmannigf.) linear unabhängig (J_g hat maximalen Rang m) und bilden daher eine Basis von $W^\perp$. Daher existieren eindeutig bestimmte Koeffizienten $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$ (Basiskoeffizienten), so daß

$$
\nabla f(c) = \sum_{j=1}^{m} \lambda_j \nabla g_j(c).
$$

□

Hinreichende Bedingung fuer lokales Minimum unter Nebenbedingungen

Theorem 0.2 *Sei $\Omega \subset \mathbb{R}^n$ und $f, g_1, \ldots, g_m \in C^2(\Omega)$ mit $m < n$. und $N_g = \{x \in \mathbb{R}^n \,|\, g_1(x) = \ldots = g_m(x) = 0\}$ eine Lösungsmannigfaltigkeit. Hinreichend für ein lokales Minimum von f unter den Nebenbedingungen $g_1 = \ldots = g_m = 0$ an der Stelle $c \in N_g$ ist die Existenz von Zahlen $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$, so daß für*

$$
\phi(x) := f(x) - \sum_{j=1}^{m} \lambda_j \, g_j(x),
$$

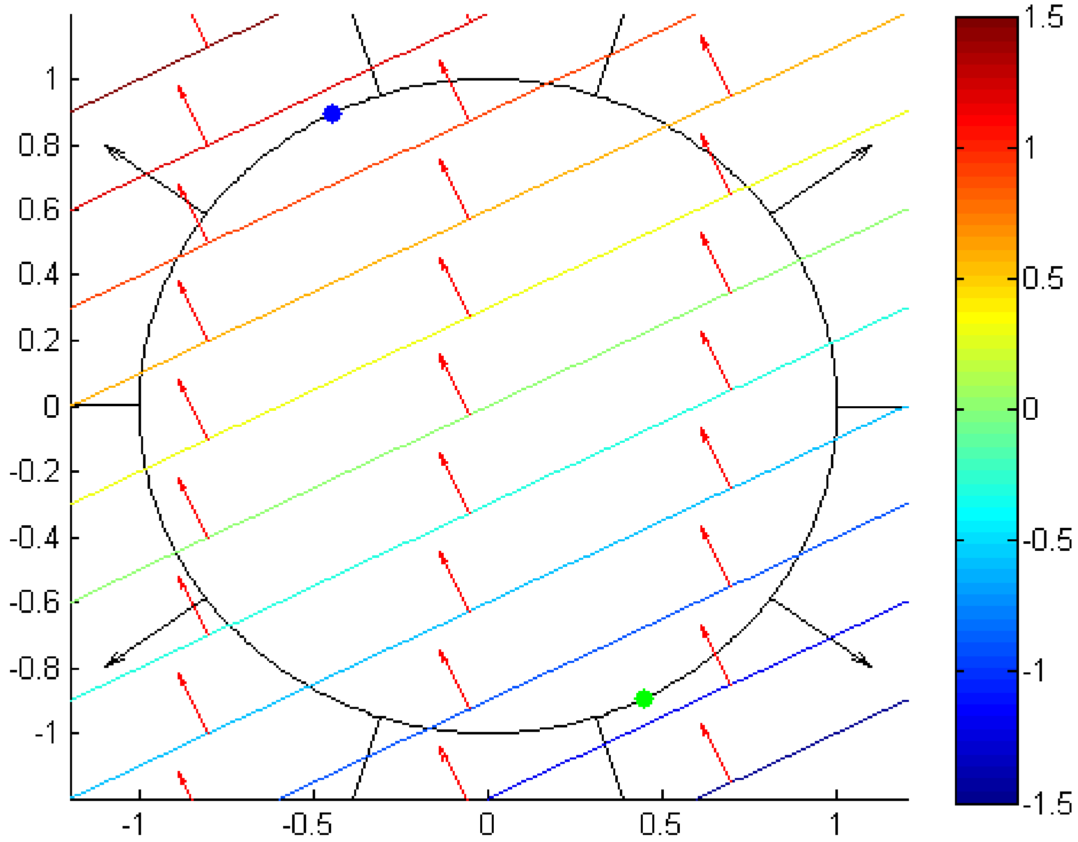
gilt

1. $\nabla \phi(c) = 0$,
2. $\langle v, (H\phi)(c)v \rangle > 0$ *für alle* $v \in W^\perp$ *mit* $v \neq 0$ *wobei* $W = \operatorname{span}\{\nabla g_1(c), \ldots, \nabla g_m(c)\}$.

Geometrische Motivation (kein Beweis !)

Um die Situation zu vereinfachen, nehmen wir an, daß wir nur eine Nebenbedingung $g(x) = 0$ haben. Sei c ein Punkt mit $g(c) = 0$. Das Taylorpolynom erster Ordnung ist am Punkt c

$$
g(c) + \langle \nabla g(c), x - c \rangle = \langle \nabla g(c), x - c \rangle
$$



Die Punkte $x \in \mathbb{R}^n$ mit $\langle \nabla g(c), x - c\rangle = 0$ bilden die Tangentialebene an die Nullstellenmenge (Normalenvektor ist $\nabla g(c)$). Das Taylorpolynom erster Ordnung von f an der Stelle c ist

$$f(c) + \langle \nabla f(c), x - c\rangle .$$

Falls $\nabla f(c) \neq \lambda \nabla g(c)$ dann existiert ein $d \in \mathbb{R}^n$ mit $\langle \nabla g(c), d - c\rangle = 0$ aber $\langle d - c, \nabla f(c)\rangle \neq 0$. Daraus folgt aber, daß wir ein (infinitesimales) Stück Weg, $\gamma : (-\epsilon, \epsilon) \to \mathbb{R}^n$, $\gamma(t) = c + td$ entlang der Tangentialebene (resp. entlang N_g) laufen können und dabei f wächst ($\langle d - c, \nabla f(c)\rangle > 0$) oder kleiner wird ($\langle d - c, \nabla f(c)\rangle < 0$), siehe Taylorpolynom erster Ordnung.

Daraus folgt, daß an einem extremalen Punkt der Gradient von f und g kollinear sein müssen. Bei mehreren Ungleichungen betrachten wir den Schnitt aller Tangentialebenen, dann muß $\nabla f(c)$ orthogonal zu den Tangentialebenen von allen Nebenbedingungsfunktionen stehen.

Im Schaubild wird folgendes Problem illustriert: finde Extrempunkte von $f(x) = \langle a, x\rangle$ unter der Nebenbedingung $g(x) = \|x\|_2^2 = 1$ d.h. finde das Maximum/Minimum der linearen Funktion auf dem Kreis. Gezeigt werden die Niveaumengen von f und der Gradient von f (rot), $\nabla f(x) = a$. In schwarz wird die Nullstellenmenge N_g gezeigt (der Einheitskreis) und der Gradient von g, $\nabla g(x) = 2x$, an ausgewählten Punkten von N_g. Das lokale (und hier globale) Maximum von f wird angenommen an der blauen Stelle, $x^* = \frac{a}{\|a\|_2}$, und das lokale (und hier globale) Minimum von f an der grünen Stelle, $x^* = -\frac{a}{\|a\|_2}$. Es ist klar ersichtlich, daß an diesen Punkten der Gradient von f und g kollinear sind.