Optimierung

Vorlesung 5

Optimierung mit Nebenbedingungen

Minimierung mit Gleichheitsrestriktionen

Gegeben: Funktion $f, h_1, h_2, \ldots, h_m : \mathbb{R}^n \to \mathbb{R}$

Ziel:

$$\min_{x \in \mathbb{R}^n} f(x)$$

$$\textbf{wobei } h_1(x) = 0, h_2(x) = 0, \ldots, h_m(x) = 0$$

Unrestringierter Fall: Notwendige Bedingung für lokales Minimum x^*

- keine Abstiegsrichtung bei x^*: $\nabla f(x^*) = 0$

Restringierter Fall: Notwendige Bedingung für lokales Minimum x^*

- x^* ist **zulässig**: $\forall i$: $h_i(x^*) = 0$
- keine **zulässige** Abstiegsrichtung s bei x^*

Notwendige Bedingung

Restringierter Fall: Notwendige Bedingung für lokales Minimum x^*

- x^* ist **zulässig**: $\forall i$: $h_i(x^*) = 0$
- keine **zulässige** Abstiegsrichtung s

Zulässiger inkrementeller Schritt δ:

$$h_i(x^* + \delta) = h_i(x^*) = 0$$

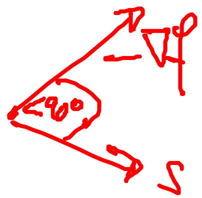
Taylor-Entwicklung bei x^*:

$$0 = h_i(x^* + \delta) = \underbrace{h_i(x^*)}_{=0} + \underbrace{\delta^\top \nabla h_i(x^*)}_{(=0)} + \cancel{o(\|\delta\|)}$$

Zulässige Richtungen s sind orthogonal zu $\nabla h_i(x^*)$: $\quad s^\top \nabla h_i(x^*) = 0$

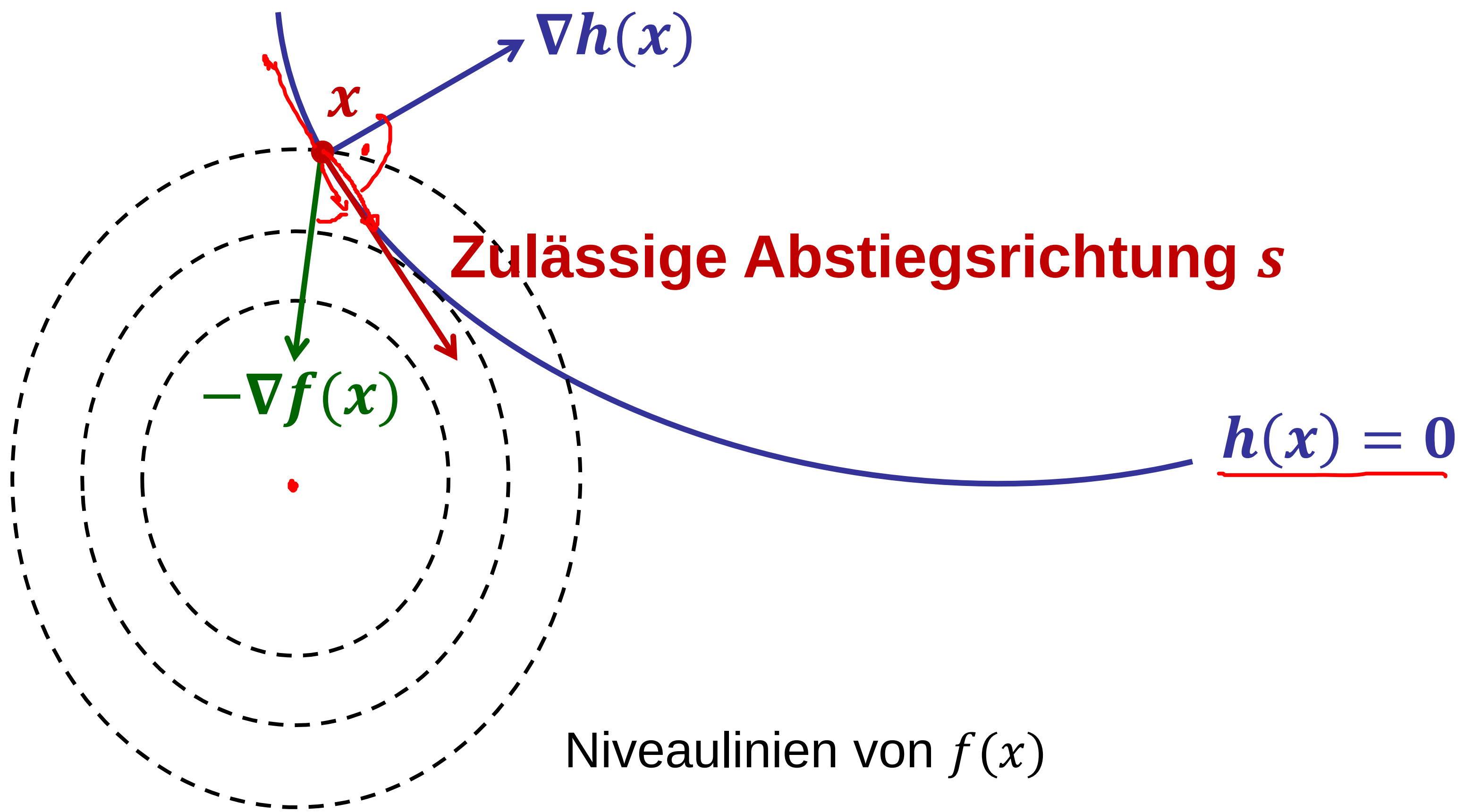
Zulässige Abstiegsrichtung s:

$$\forall i: \; s^\top \nabla h_i(x^*) = 0, \qquad s^\top \nabla f(x^*) < 0$$



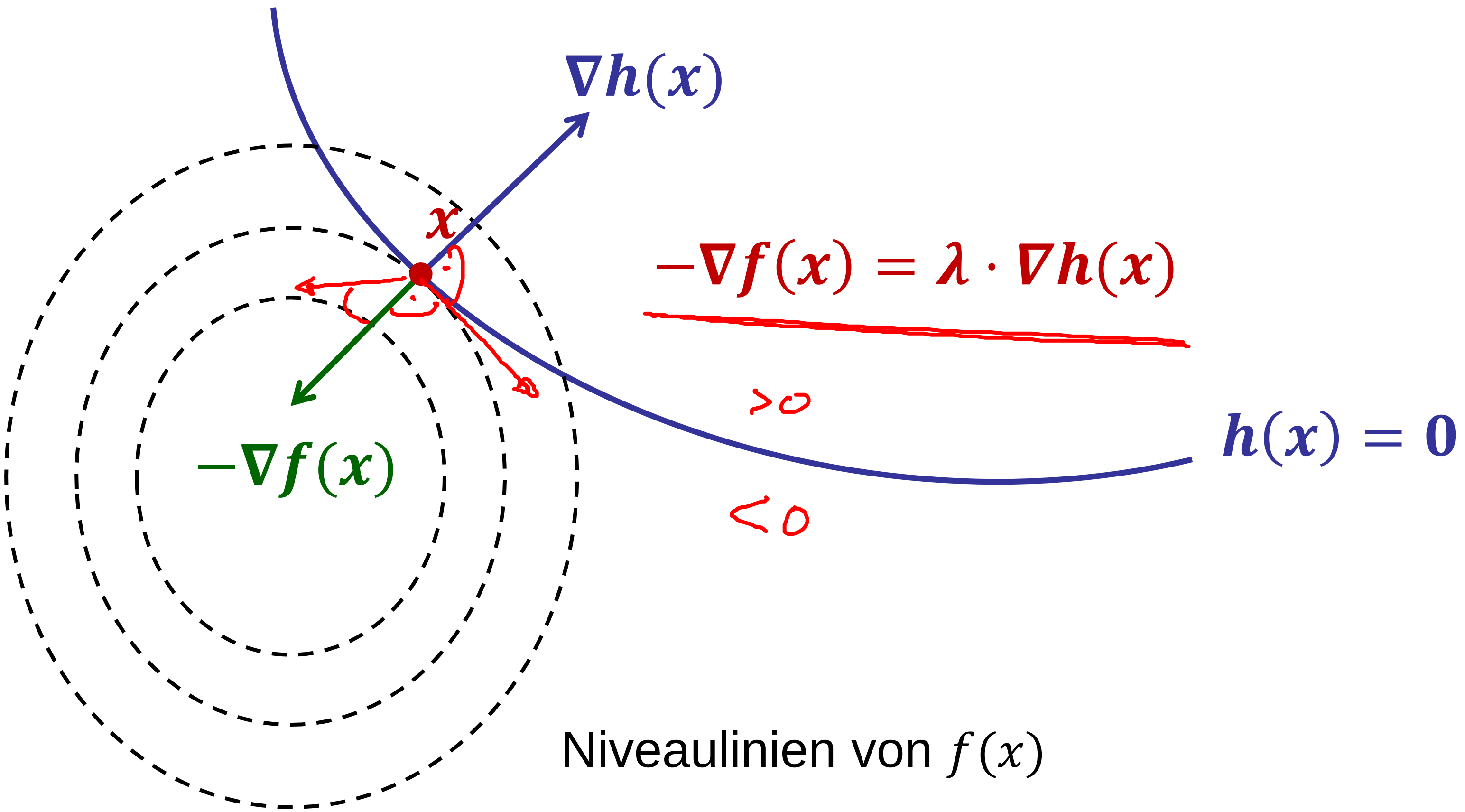
Zulässige Abstiegsrichtung s:

$$\forall i:\ s^\top \nabla h_i(x^*) = 0, \qquad s^\top \nabla f(x^*) < 0$$



Keine zulässige Abstiegsrichtung s:

$$\forall s: \quad \text{Falls } \forall i: \; s^\top \nabla h_i(x^*) = 0, \qquad \text{dann } s^\top \nabla f(x^*) = 0$$



Keine zulässige Abstiegsrichtung

Keine zulässige Abstiegsrichtung s:

$$\forall s: \quad \text{Falls } \forall i: \; s^{\top} \nabla h_i(x^*) = 0, \qquad \text{dann } s^{\top} \nabla f(x^*) = 0 \qquad (*)$$

Satz (Lagrange Multiplikatoren): Unter Annahme von Regularität, gibt es bei x^* keine zulässige Abstiegsrichtung genau dann, wenn λ_i^* existieren, so dass

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \cdot \nabla h_i(x^*) = 0. \qquad (**)$$

Beweis: $-s^{\top} \nabla f(x^*) = \sum \lambda_i^* \underbrace{s^{\top} \nabla h_i(x^*)}_{=0} \Longrightarrow s^{\top} \nabla f(x^*) = 0$

Annahme: keine zul. Abs.-richtung und keine λ_i^*, so dass $(**)$ gilt

$$-\nabla f(x^*) = \sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) + \mu \qquad \mu^{\top} \nabla h_i(x^*) = 0 \quad (\forall i)$$

$$\mu \neq 0$$

$s = \mu$ $\quad$ s ist zulässig

$$-s^{\top} \nabla f(x^*) = + \underbrace{s^{\top} s}_{>0} \Longrightarrow s^{\top} \nabla f(x^*) < 0$$

Lagrange Multiplikatoren

Satz (Lagrange Multiplikatoren): Unter Annahme von Regularität, gibt es bei x^* keine zulässige Abstiegsrichtung genau dann, wenn λ_i^* existieren, so dass

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \cdot \nabla h_i(x^*) = 0.$$

und $h_i(x^*) = 0 \;\; \forall i$

Regularität: Insbesondere müssen $\nabla h_i(x^*)$ linear unabhängig sein

Methode von Lagrange:

Finde Vektoren x^* und λ^*, für welche

$$\boldsymbol{\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \cdot \nabla h_i(x^*) = 0}$$

$$\boldsymbol{h_i(x^*) = 0, \qquad i = 1, \dots, m}$$

- $n + m$ Unbekannte, $n + m$ Gleichungen

Beispiel

Lagrange: $\nabla f(x^*) + \sum_i \lambda_i^* \cdot \nabla h_i(x^*) = 0, \quad h_i(x^*) = 0 \ (i = 1, \ldots, m)$

Beispiel:

$$\min f(x) = x_1 + x_2$$

$$\text{unter } h(x) = x_1^2 - x_2 = 0$$

$$\nabla f(x) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \nabla h(x) = \begin{pmatrix} 2x_1 \\ -1 \end{pmatrix}$$

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix} + \lambda \begin{pmatrix} 2x_1 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \Longrightarrow \lambda^* = 1$$

$$1 + 2x_1 = 0 \quad \Longrightarrow x_1^* = -1/2$$

$$\frac{1}{4} - x_2 = 0 \quad \Longrightarrow x_2^* = 1/4$$

Lagrange Methode

Lagrange Funktion:

$$L(x,\lambda) := f(x) + \sum_{i=1}^{m} \lambda_i \cdot h_i(x)$$

Gradient:

$$\frac{\partial}{\partial x_j} L(x,\lambda) = \frac{\partial}{\partial x_j} f(x) + \sum_{i=1}^{m} \lambda_i \cdot \frac{\partial}{\partial x_j} h_i(x)$$

$$\frac{\partial}{\partial \lambda_j} L(x,\lambda) = h_j(x)$$

Lagrange Methode:

$\nabla L(x,\lambda) = 0$

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \cdot \nabla h_i(x^*) = 0, \qquad h_i(x^*) = 0 \quad (i = 1, \dots, m)$$

Lagrange Methode

Lagrange Funktion:

$$\underline{L(x,\lambda)} := f(x) + \sum_{i=1}^{m} \lambda_i \cdot h_i(x)$$

Lagrange Methode:

Notwendige Bedingung für lokales Minimum bei regulärem Punkt x^*:

$$\underline{\boldsymbol{\nabla L(x^*, \lambda^*) = 0}}$$

- Reduziert das Problem auf das Finden eines stationären Punktes eines unrestringierten Problems!
- Methoden: (konj.) Gradientenmethode, Newton, Quasi-Newton, ...

Optimierung mit Ungleichheitsrestriktionen

Gegeben: Funktion $f, h_1, h_2, \dots, h_m : \mathbb{R}^n \to \mathbb{R}$

Ziel:

$$\min_{x \in \mathbb{R}^n} f(x)$$

$$\textbf{wobei } g_1(x) \leq 0, g_2(x) \leq 0, \dots, g_p(x) \leq 0$$

Notwendige Bedingung für lokales Minimum x^*:

- x^* ist **zulässig**: $\forall i$: $g_i(x^*) \leq 0$
- keine **zulässige** Abstiegsrichtung s bei x^*

$g_i(x^*) < 0$

Aktive Nebenbedingungen:

$$I^* := \{i \mid g_i(x^*) = 0\}$$

- Nur aktive Nebenbedingungen beschränken die gültigen Richtungen

Zulässige Richtung

Aktive Nebenbedingung g_i an Punkt x^*:

$$g_i(x^*) = 0$$

Zulässige Richtung s and Punkt x^*:

- Für genug kleines $\delta > 0$ gilt:

$$g_i(x^* + \delta \cdot s) \leq 0$$

- Taylor-Approximation:

$$\underbrace{g_i(x^* + \delta \cdot s)}_{\leq 0} \approx \overset{=0}{g_i(x^*)} + \underbrace{\delta \cdot s^\top \nabla g_i(x^*)}_{\leq 0} \leq 0$$

- Zulässige Richtung s bei x^* (notwendige Bedingung):

$$\boldsymbol{s^\top \nabla g_i(x^*) \leq 0}$$

Zulässige Abstiegsrichtung s bei x^*:

$$\boldsymbol{s^\top \nabla f(x^*) < 0} \quad \text{und} \quad \boldsymbol{\forall i \in I^*: s^\top \nabla g_i(x^*) \leq 0}$$

Lokales Minimum: Notwendige Bedingung

Bei **regulärem lokalem Minimum $\boldsymbol{x}^*$** hat folgendes System keine Lösung:

$$\left.\begin{array}{ll} s^\top \nabla g_i(x^*) \leq 0, & \forall i \in I^* \\ \boxed{s^\top \nabla f(x^*) < 0} & \end{array}\right. \qquad \text{(A)}$$

Hinreichende Bedingung, damit (A) keine Lösung hat:

$$-\nabla f(x^*) = \sum_{i \in I^*} \lambda_i \cdot \nabla g_i(x^*), \qquad \text{mit } \lambda_i \geq 0 \text{ für } i \in I^* \qquad \text{(B)}$$

$$-s^\top \nabla f(x^*) = \sum_{i \in I^*} \lambda_i \underbrace{s^\top \nabla g_i(x^*)}_{\leq 0} \leq 0$$

$$s^\top \nabla f(x^*) \geq 0$$

Lemma von Farkas:

- (B) ist auch eine notwendige Bedingung, damit (A) keine Lösung hat

Lemma von Farkas

Satz (Lemma von Farkas): Für beliebige Vektoren $\boldsymbol{a_1}, \dots, \boldsymbol{a_k}$ und $\boldsymbol{g}$ ist die Menge

$$S = \{\boldsymbol{s} \mid \boldsymbol{s}^\top \boldsymbol{g} < 0 \;\; \text{und} \;\; \boldsymbol{s}^\top \boldsymbol{a_i} \leq 0 \text{ for all } i = 1, \dots, k\}$$

genau dann leer, falls $\lambda_i \geq 0$ existieren, so dass

$a_i = \nabla g_i(x^*)$

$g = \nabla f(x^*)$

$$-\boldsymbol{g} = \sum_{i=1}^{k} \lambda_i \boldsymbol{a_i}\,.$$

- Äquivalent zum **Dualitätssatz** der **linearen Programmierung**

Anwendung hier: System

$$s^\top \nabla f(x^*) < 0, \qquad s^\top \nabla g_i(x^*) \leq 0, \qquad \forall i \in I^*$$

hat genau dann keine Lösung, wenn

$$-\nabla f(x^*) = \sum_{i \in I^*} \lambda_i \cdot \nabla g_i(x^*)\,, \qquad \text{mit } \lambda_i \geq 0 \text{ für } i \in I^*$$

Lokales Minimum: Notwendige Bedingung

Notwendige Bedingung für (reguläres) lokales Minimum bei x^*:

$$-\nabla f(x^*) = \sum_{i \in I^*} \lambda_i \cdot \nabla g_i(x^*), \qquad \text{mit } \lambda_i \geq 0 \text{ für } \underline{i \in I^*}$$

<u>**Karush-Kuhn-Tucker**</u> **Satz:** Sei x^* ein reguläres, lokales Minimum. Dann existieren Multiplikatoren λ_i^*, derart dass

$$\nabla f(x^*) + \sum_{i=1}^{p} \lambda_i^* \nabla g_i(x^*) = 0$$

$$\lambda_i^* \geq 0, \qquad i = 1, \ldots, p \qquad i \notin I^* : g_i(x^*) \neq 0$$

$$\longrightarrow \lambda_i^* g_i(x^*) = 0, \qquad i = 1, \ldots, p \qquad \Longrightarrow \lambda_i^* = 0$$

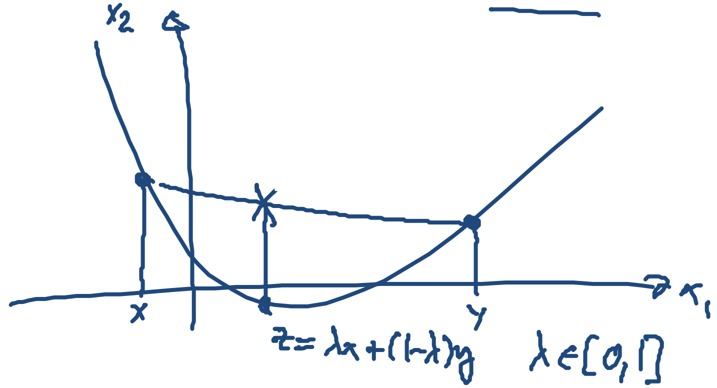
<u>Karush-Kuhn-Tucker Punkt x^*</u>

Konvexität

Zur Erinnerung:

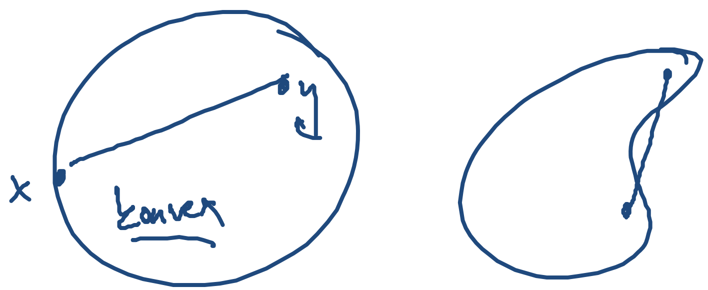
- Funktion $f : \mathbb{R}^n \to \mathbb{R}$ heisst konvex, falls für alle $x, y \in \mathbb{R}^n$ und $\lambda \in [0,1]$:

$$f(\lambda x + (1-\lambda)y) \leq \underline{\lambda f(x)} + (1-\lambda)\underline{f(y)}$$



- Eine Menge $S \subseteq \mathbb{R}^n$ heisst, falls für alle $x, y \in S$ und $\lambda \in [0,1]$:

$$\lambda x + (1-\lambda)y \in S$$



Konvexe Mengen

Der Schnitt zweier konvexer Mengen ist konvex:

$$S_1, S_2 \text{ konvex} \implies S_1 \cap S_2 \text{ konvex}$$

Konvexe Nebenbedingung:

$$\text{Funktion } g_i(x) \text{ konvex} \implies \{x : g_i(x) \le 0\} \text{ ist konvex}$$

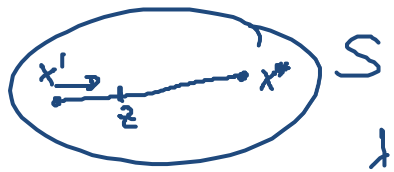
$g_i(x) \le 0$ $\quad$ $g_i(y) \le 0 \implies g_i(\lambda x + (1-\lambda)y) \le 0$

$g_i(\lambda x + (1-\lambda)y) \le \lambda g(x) + (1-\lambda) g(y) \le 0$

Konvexe Minimierung:

$$\min_{x \in S} f(x), \qquad \text{wobei Fkt. } f(x) \text{ und Menge } S \text{ konvex sind}$$

Konvexe Minimierung: **lokales Minimum $x' \in S$ ist auch globales Minimum:**



$\lambda \in (0,1)$

- Globales Minimum $x^* \in S$
- Konvexität von S: Verbindungslinie in S
- Konvexität von f: Betrachte $z = \lambda x' + (1-\lambda)x^*$ auf Verbindungslinie:

$$f(z) = f(\lambda x' + (1-\lambda)x^*) \le \lambda f(x') + (1-\lambda) f(x^*) < f(x')$$

(falls $f(x') > f(x^*)$)

Konvexe Optimierung: Hinreichende Bedingung

Karush-Kuhn-Tucker Satz: Sei x^* ein reguläres, lokales Minimum. Dann existieren Multiplikatoren λ_i^*, derart dass

$$\nabla f(x^*) + \sum_{i=1}^{p} \lambda_i^* \nabla g_i(x^*) = 0$$

$$\lambda_i^* \geq 0, \qquad i = 1, \dots, p$$

$$\lambda_i^* g_i(x^*) = 0, \qquad i = 1, \dots, p$$

Hinreichende Karush-Kuhn-Tucker Bedingungen

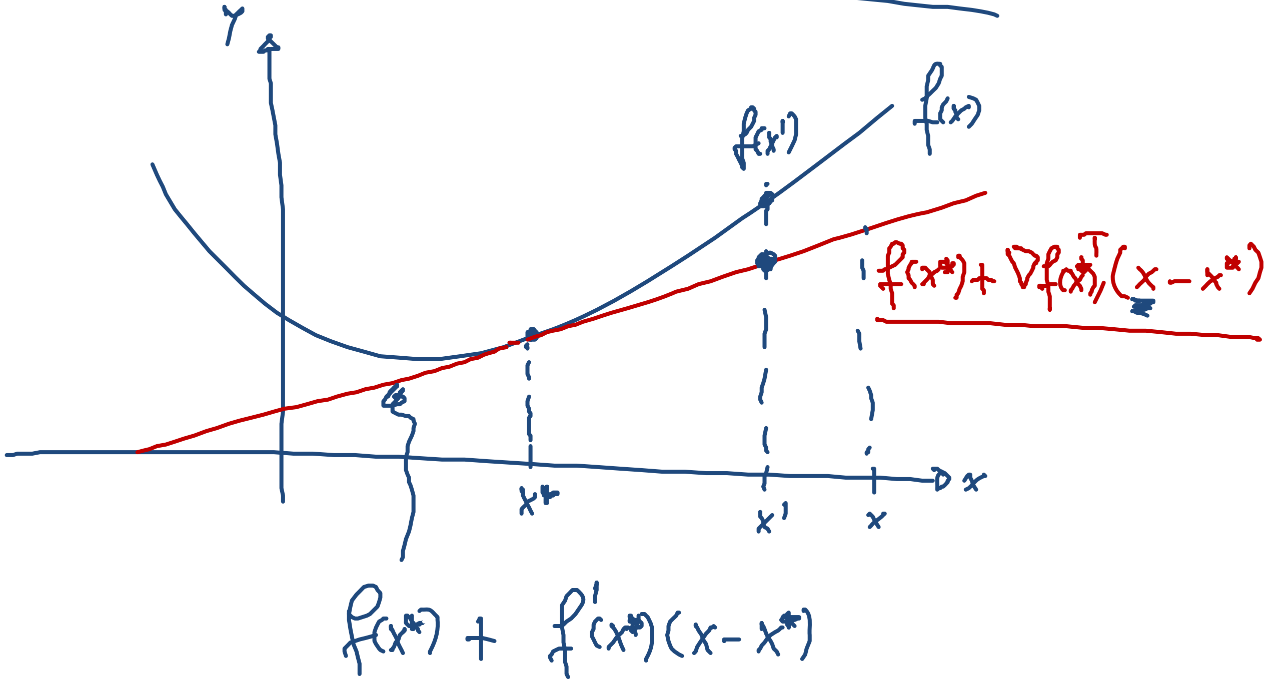
Satz: Falls $f(x), g_1(x), \dots, g_p(x)$ ***konvex*** sind, dann ist jeder Punkt x^*, welcher die Karush-Kuhn-Tucker Bedingungen erfüllt, ein globales Minimum.

Konvexe Optimierung: Hinreichende Bedingung

Satz: Falls $f(x), g_1(x), \dots, g_p(x)$ ***konvex*** sind, dann ist jeder Punkt x^*, welcher die Karush-Kuhn-Tucker Bedingungen erfüllt, ein globales Minimum.

Eigenschaft konvexer Funktionen:

$$\forall x', x^*: \ f(x') \geq f(x^*) + \nabla f(x^*)^{\top}(x' - x^*)$$



Konvexe Optimierung: Hinreichende Bedingung

Satz: Falls $f(x), g_1(x), \ldots, g_p(x)$ ***konvex*** sind, dann ist jeder Punkt x^*, welcher die Karush-Kuhn-Tucker Bedingungen erfüllt, ein globales Minimum.

Betrachte Karush-Kuhn-Tucker Punkt $\underline{x^*}$ und ein beliebiges $\underline{x'} \neq x^*$:

zulässig

Konvexität:

$$f(x') \geq f(x^*) + \nabla f(x^*)^\top (x' - x^*)$$
$$\forall i: g_i(x') \geq g_i(x^*) + \nabla g_i(x^*)^\top (x' - x^*)$$

$g_i(x') \leq 0$, $\lambda_i^* \geq 0$

$$f(x') \geq f(x') + \sum_{i=1}^{p} \lambda_i^* g_i(x')$$

$$\geq f(x^*) + \nabla f(x^*)^\top (x' - x^*) + \sum_{i=1}^{p} \lambda_i^* \left(g_i(x^*) + \nabla g_i(x^*)^\top (x' - x^*) \right)$$

$$= f(x^*) + \left[\nabla f(x^*) + \sum_{i=1}^{p} \lambda_i^* \nabla g_i(x^*) \right]^\top (x' - x^*) + \sum_{i=1}^{p} \lambda_i^* g_i(x^*)$$

Konvexe Optimierung: Hinreichende Bedingung

Satz: Falls $f(x), g_1(x), \dots, g_p(x)$ ***konvex*** sind, dann ist jeder Punkt x^*, welcher die Karush-Kuhn-Tucker Bedingungen erfüllt, ein globales Minimum.

Betrachte Karush-Kuhn-Tucker Punkt x^* und ein beliebiges $x' \neq x^*$:

$$f(x') \geq f(x^*) + \Bigg[\underbrace{\nabla f(x^*) + \sum_{i=1}^{p} \lambda_i^* \nabla g_i(x^*)}_{=0}\Bigg]^{\mathsf{T}} (x' - x^*) + \sum_{i=1}^{p} \underbrace{\lambda_i^* g_i(x^*)}_{=0}$$

$= f(x^*)$

$f(x') \geq f(x^*)$

Punkt x^* erfüllt Karush-Kuhn-Tucker Bedingungen:

$$\nabla f(x^*) + \sum_{i=1}^{p} \lambda_i^* \nabla g_i(x^*) = 0$$

$$\lambda_i^* \geq 0 \text{ und } \lambda_i^* g_i(x^*) = 0, \qquad i = 1, \dots, p$$

Beispiel

Ist $x^* = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$ Optimallösung des folgenden Minimierungsproblems?

$$\min f(x) = -x_1 - x_2$$

$$g_1(x) = x_1^2 - x_2 \leq 0$$
$$g_2(x) = x_1^2 + x_2^2 - 1 \leq 0$$

Minimierung mit allgemeinen Nebenbedingungen

Gegeben: Funktion $f, h_1, h_2, \dots, h_m : \mathbb{R}^n \to \mathbb{R}$

Ziel:

$$\min_{x \in \mathbb{R}^n} f(x)$$

$$\textbf{wobei } h_1(x) = 0, h_2(x) = 0, \dots, h_m(x) = 0$$
$$g_1(x) \le 0, g_2(x) \le 0, \dots, g_p(x) \le 0$$

Notwendige Bedingung für (reguläres) lokales Minimum x^*:

Falls $\underline{s^\top \nabla h_i(x^*) = 0}$ für $i = 1, \dots, m$ und $\underline{s^\top \nabla g_i(x^*) \le 0}$ für $i \in I^*$, dann gilt

$$\underline{s^\top \nabla f(x^*) \ge 0}.$$

Farkas-Lemma: Es gibt Multiplikatoren λ_i^* $(i = 1, \dots, m)$ und $\underline{\mu_i^* \ge 0}$ $(i \in I^*)$,

$$\underline{\nabla f(x^*) + \sum_i^m \lambda_i^* \nabla h_i(x^*) + \sum_{i \in I^*} \mu_i^* \nabla g_i(x^*) = 0.}$$