# Vorlesung
# **Einführung in die Mathematische Optimierung**
(Wintersemester 2013/14)

Kapitel 3: Optimalitätsbedingungen für konvexe Optimierungsprobleme

Volker Kaibel

Otto-von-Guericke Universität Magdeburg

(Version vom 12. November 2013)

# Gliederung

# Das Setup

- Konvexes Optimierungsproblem:

$$\min\{f(x) \,:\, x \in X\}$$

- $f : \mathbb{R}^n \to \mathbb{R}$ differenzierbare konvexe Zielfunktion
- Zulässige Lösungen:

$$X = \left\{ x \in X^{(0)} \,:\, \begin{array}{ll} g^{(i)}(x) \leq 0 & \text{für alle } i \in [m] \\ h^{(i)}(x) = 0 & \text{für alle } i \in [p] \end{array} \right\}$$

- $X^{(0)} \subseteq \mathbb{R}^n$ abgeschlossene konvexe Menge (einfache Struktur)
- $g^{(1)}, \ldots, g^{(m)} : \mathbb{R}^n \to \mathbb{R}$ differenzierbare konvexe Funktionen
- $h^{(1)}, \ldots, h^{(p)} : \mathbb{R}^n \to \mathbb{R}$ affine Funktionen
- Also: $X$ abgeschlossene konvexe Menge

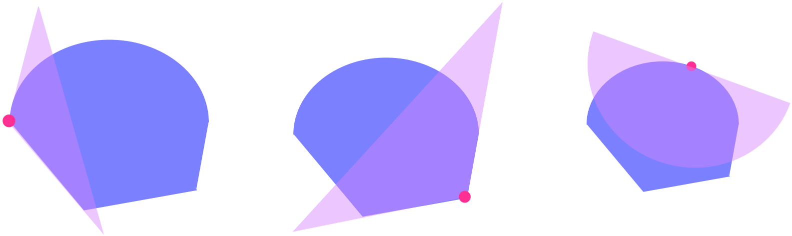
# Radialkegel

**Definition 3.1**

Für eine konvexe Menge $X \subseteq \mathbb{R}^n$ und $x^\star \in X$ heißt

$$\mathrm{K}_{x^\star}(X) \ := \ \mathrm{cone}(X - \{x^\star\})$$

der **Radialkegel** (**Kegel der zulässigen Richtungen**) von $X$ in $x$.



**Beobachtung 3.2**

*Ist $x^\star$ ein innerer Punkt von $X$, so gilt $\mathrm{K}_{x^\star}(X) = \mathbb{R}^n$.*

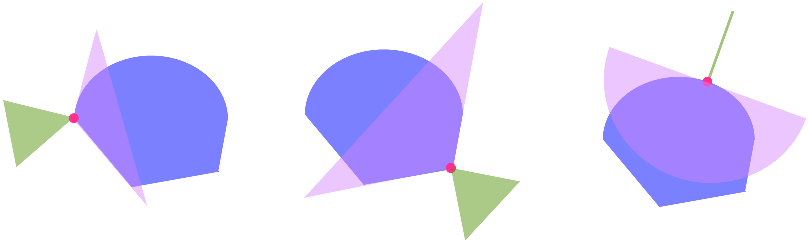
# Normalenkegel

**Definition 3.3**

Für eine konvexe Menge $X \subseteq \mathbb{R}^n$ und $x \in X$ heißt

$$\mathrm{N}_{x^\star}(X) \;:=\; \mathrm{K}_{x^\star}(X)^\circ$$

der **Normalenkegel** von $X$ in $x^\star$.



**Beobachtung 3.4**

*Ist $x^\star$ ein innerer Punkt von $X$, so gilt $\mathrm{N}_{x^\star}(X) = \{\mathbb{O}_n\}$.*

# Radial- und Normalenkegel an Polyeder

Für $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ und $x^\star \in \mathbb{R}^n$ mit $Ax^\star \leq b$:

$$\mathrm{Eq}_{Ax \leq b}(x^\star) := \{i \in [m] \,|\, \langle A_{i,\star}, x^\star \rangle = b_i\} \subseteq [m]$$

## Bemerkung 3.5

*Für* $x^\star \in \mathrm{P}^{\leq}(A, b) = \{x \in \mathbb{R}^n \,|\, Ax \leq b\}$ *(mit* $A \in \mathbb{R}^{m \times n}$ *und* $b \in \mathbb{R}^m$*) gilt:*

$$\begin{aligned} \mathrm{K}_{x^\star}(\mathrm{P}^{\leq}(A, b)) &= \{y \in \mathbb{R}^n \,|\, \langle A_{i,\star}, y \rangle \leq 0 \textit{ für alle } i \in \mathrm{Eq}_{Ax \leq b}(x^\star)\} \\ \mathrm{N}_{x^\star}(\mathrm{P}^{\leq}(A, b)) &= \mathrm{ccone}\{A_{i,\star} \,|\, i \in \mathrm{Eq}_{Ax \leq b}(x^\star)\} \end{aligned}$$

*Insbesondere sind also Radial- und Normalenkegel an Polyeder polyedrisch.*

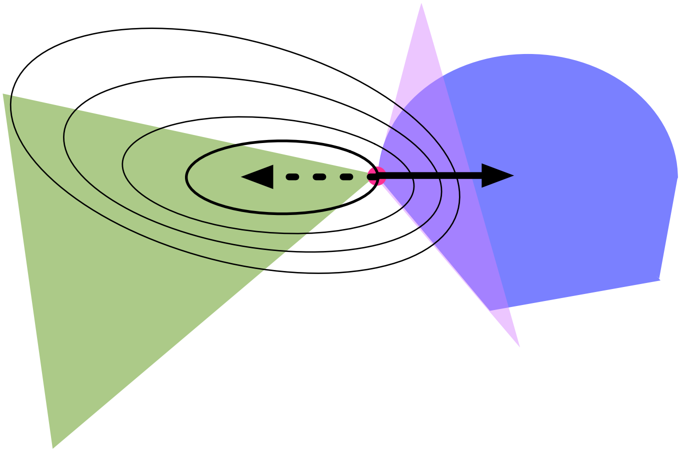
# Differenzierbare Optimierung über konvexen Mengen

**Satz 3.6**

*Sei* $f : X \to \mathbb{R}$ *eine differenzierbare Funktion auf einer konvexen Mengen* $X \subseteq \mathbb{R}^n$.

1. *Nimmt* $f$ *in* $x^\star \in X$ *ein lokales Minimum über* $X$ *an, so ist*
$$-\operatorname{grad}_{x^\star} f \in \mathrm{N}_{x^\star}(X)\,.$$
2. *Ist* $f$ *konvex und gilt* $-\operatorname{grad}_{x^\star} f \in \mathrm{N}_{x^\star}(X)$, *so gilt*
$$f(x^\star) = \min\{f(x) \,|\, x \in X\}\,.$$

# Differenzierbare Optimierung über konvexen Mengen



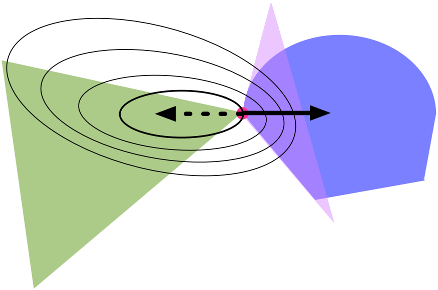
# Satz 3.6 für konvexe Zielfunktionen

**Korollar 3.7**

*Die differenzierbare konvexe Funktion $f$ nimmt in $x^\star \in X$ genau dann ihr (globales) Minimum über der konvexen Menge $X$ an, wenn*

$$-\operatorname{grad}_{x^\star} f \ \in \ \mathrm{N}_{x^\star}(X)$$

*gilt.*



# Einige Normalenkegel

## Lemma 3.8

*Für konvexe Kegel* $K \subseteq \mathbb{R}^n$ *und* $x^\star \in K$ *ist*

$$\mathrm{N}_{x^\star}(K) = \{y \in K^\circ \mid \langle x^\star, y\rangle = 0\}.$$

## Bemerkung 3.9

1. *Für alle* $x^\star \in \mathbb{R}^n_+$ *ist*
$$\mathrm{N}_{x^\star}(\mathbb{R}^n_+) = \{y \in \mathbb{R}^n_- \mid y_i = 0 \text{ für alle } i \in [n] \text{ mit } x^\star_i > 0\}.$$
2. *Für alle* $X^\star \in \mathbb{S}^k_+$ *ist*
$$\mathrm{N}_{X^\star}(\mathbb{S}^k_+) = \{Y \in \mathbb{S}^k_- \mid \langle X^\star, Y\rangle = 0\}.$$

## Bemerkung 3.10

*Für konvexe Kegel* $K_i \subseteq \mathbb{R}^{n_i}$ $(i \in [r])$ *und* $(x^{(1)}, \dots, x^{(r)}) \in K_1 \times \cdots \times K_r$ *gilt*

$$\mathrm{N}_{(x^{(1)},\dots,x^{(r)})}(K_1 \times \cdots \times K_r) = \mathrm{N}_{x^{(1)}}(K_1) \times \cdots \times \mathrm{N}_{x^{(r)}}(K_r).$$

# Das Setup

$(X_0, (g_i)_{i\in[m]}, (h_i)_{i\in[p]})$ (für $m, p \geq 0$) erfülle:

- $X_0 \subseteq \mathbb{R}^n$ konvex
- $g_i : \mathbb{R}^n \to \mathbb{R}$ konvex und differenzierbar (für $i \in [m]$).
- $h_i : \mathbb{R}^n \to \mathbb{R}$ affin (für $i \in [p]$).

Menge der zulässigen Lösungen:

$X = \{x \in X_0 \mid g_i(x) \leq 0 \text{ für alle } i \in [m], h_i(x) = 0 \text{ für alle } i \in [p]\}$

Mit $G_i := g_i^{-1}(\mathbb{R}_-) \subseteq \mathbb{R}^n$ und $H_i := h_i^{-1}(\{0\}) \subseteq \mathbb{R}^n$:

$$X = X_0 \cap \bigcap_{i\in[m]} G_i \cap \bigcap_{i\in[p]} H_i$$

# Der Normalenkegel für reguläre Tripel

**Lemma 3.11**

*Falls* $(X_0, (g_i)_{i\in[m]}, (h_i)_{i\in[p]})$ *regulär ist:*

$$\mathrm{N}_{x^\star}(X) = \mathrm{N}_{x^\star}(X_0) + \sum_{i\in[m]} \mathrm{N}_{x^\star}(G_i) + \sum_{i\in[p]} \mathrm{N}_{x^\star}(H_i)$$

$(X_0, (g_i)_{i\in[m]}, (h_i)_{i\in[p]})$ **regulär**, falls:

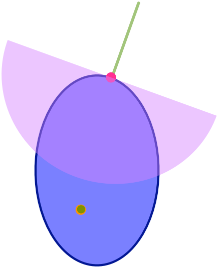
1. Die Menge $X_0$ ist ein Polyeder und die Funktionen $g_1, \dots, g_m$ sind affin oder
2. Die Menge $X \cap \operatorname{int}(X_0)$ ist nicht leer, und die Funktionen $g_1, \dots, g_m$ sind affin oder
3. Es gibt ein $x^{(s)} \in X$ mit $x^{(s)} \in \operatorname{int}(X_0)$ falls $p \neq 0$, für das $g_i(x^{(s)}) < 0$ für alle $i \in [m]$ gilt (**Slater-Bedingung**).

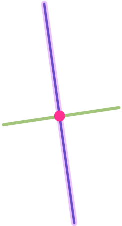
# Eine differenzierbare konvexe Nebenbedingung

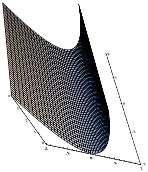
## Satz 3.12

*Ist* $g : \mathbb{R}^n \to \mathbb{R}$ *differenzierbar und konvex und gibt es ein* $x^{(s)} \in \mathbb{R}^n$ *mit* $g(x^{(s)}) < 0$, *so ist für alle* $x^\star \in \mathbb{R}^n$ *mit* $g(x^\star) \leq 0$

$$\mathrm{N}_{x^\star}(g^{-1}(\mathbb{R}_-)) \; = \; \begin{cases} \operatorname{cone}\{\operatorname{grad}_{x^\star} g\} & \textit{falls } g(x^\star) = 0 \\ \{\mathbb{O}_n\} & \textit{falls } g(x^\star) < 0 \end{cases} .$$







# Karush-Kuhn-Tucker (differenzierbar, konvex) ...

**Voraussetzungen:**

- $f, g_1, \ldots, g_m : \mathbb{R}^n \to \mathbb{R}$ konvex und differenzierbar
- $h_1, \ldots, h_p : \mathbb{R}^n \to \mathbb{R}$ affin
- $X_0 \subseteq \mathbb{R}^n$ konvex
- $(X_0, (g_i)_{i \in [m]}, (h_i)_{i \in [p]})$ reguläres Tripel
- $X \subseteq \mathbb{R}^n$ sei die Menge aller $x \in X_0$ mit
  - $g_i(x) \leq 0$ für alle $i \in [m]$ und
  - $h_i(x) = 0$ für alle $i \in [p]$.

# ... Karush-Kuhn-Tucker (differenzierbar, konvex)

**Satz 3.13**

*Ein Punkt $x^\star \in X$ ist genau dann Optimallösung von*

$$\min\{f(x) \mid x \in X\},$$

*wenn es Multiplizierer $\lambda_1, \dots \lambda_m \in \mathbb{R}_+$ und $\mu_1, \dots, \mu_p \in \mathbb{R}$ mit*

$$\mathrm{grad}_{x^\star} f + \sum_{i=1}^{m} \lambda_i \, \mathrm{grad}_{x^\star} g_i + \sum_{i=1}^{p} \mu_i \, \mathrm{grad}_{x^\star} h_i \in -\mathrm{N}_{x^\star}(X_0) \qquad (1)$$

*und*

$$\lambda_i = 0 \quad \textit{für alle } i \in [m] \textit{ mit } g_i(x^\star) < 0 \qquad (2)$$

*gibt.*

# KKT für LP (1. Variante)

**Satz 3.14 (Satz vom komplementären Schlupf I)**

*Seien $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ und $c \in \mathbb{R}^n$. Ein Punkt $x^\star \in \mathbb{R}^n_+$ mit $Ax^\star = b$ ist genau dann Optimallösung von*

$$\min\{\langle c, x\rangle \mid Ax = b, x \in \mathbb{R}^n_+\},$$

*wenn es $\mu \in \mathbb{R}^m$ gibt mit $\mu^\mathsf{T} A \leq c^\mathsf{T}$ und*

$$\mu^\mathsf{T} A_{\star,j} = c_j \quad \text{für alle } j \in [n] \text{ mit } x_j^\star > 0\,.$$

# KKT für LP (2. Variante)

**Satz 3.15 (Satz vom komplementären Schlupf II)**

*Seien $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ und $c \in \mathbb{R}^n$. Ein Punkt $x^\star \in \mathbb{R}^n$ mit $Ax^\star \leq b$ ist genau dann Optimallösung von*

$$\max\{\langle c, x\rangle \mid Ax \leq b, x \in \mathbb{R}^n\},$$

*wenn es einen Vektor $\lambda \in \mathbb{R}^m_+$ gibt mit $\lambda^\mathsf{T} A = c^\mathsf{T}$ und*

$$\lambda_i = 0 \quad \textit{für alle } i \in [m] \textit{ mit } \langle A_{i,\star}, x^\star\rangle < b_i\,.$$

# KKT für SDP

## Satz 3.16

*Seien* $A^{(1)}, \dots, A^{(p)} \in \mathbb{R}^{k\times k}$, $b \in \mathbb{R}^k$ *und* $C \in \mathbb{R}^{k\times k}$, *und gebe es eine positiv definite symmetrische Matrix* $X^{(s)} \in \mathbb{S}^k$ *mit* $\langle A^{(i)}, X^{(s)}\rangle = b_i$ *für alle* $i \in [p]$.

*Eine Matrix* $X^\star \in \mathbb{S}^k_+$ *mit* $\langle A^{(i)}, X^\star\rangle = b_i$ *für alle* $i \in [p]$ *ist genau dann Optimallösung von*

$$\min\{\langle C, X\rangle \mid \langle A^{(i)}, X\rangle = b_i \text{ für alle } i \in [p], X \in \mathbb{S}^k_+\}\,,$$

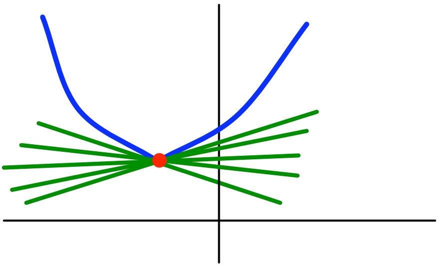
*wenn es einen Vektor* $\mu \in \mathbb{R}^p$ *gibt mit*

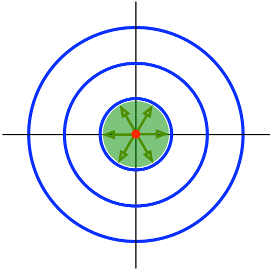
$$Y := C - \sum_{i=1}^{p} \mu_i A^{(i)} \in \mathbb{S}^k_+ \quad \text{und} \quad \langle X^\star, Y\rangle = 0\,.$$

# Konvexe nicht-differenzierbare Probleme

- Sind $f, g^{(1)}, \ldots, g^{(m)} : \mathbb{R}^n \to \mathbb{R}$ zwar konvex, aber (vielleicht) nicht differenzierbar, so gilt ein mit Hilfe von *Subgradienten* formulierbares Analogon von Satz 3.17.
- Z.B.: [Ruszczyński, Thm. 3.34]
- $y \in \mathbb{R}^n$ *Subgradient* von $f$ in $x^\star \in \mathbb{R}^n$ (d.h. $y \in \mathrm{SGRAD}_{x^\star}(f)$):

$$f(x) \;\geq\; f(x^\star) + \langle y, x - x^\star \rangle \quad \text{für alle } x \in \mathbb{R}^n$$





# Karush-Kuhn-Tucker (konvex) ...

**Voraussetzungen:**

- $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ konvex
- $g_1, \ldots, g_m : \mathbb{R}^n \to \mathbb{R}$ konvex
- $h_1, \ldots, h_p : \mathbb{R}^n \to \mathbb{R}$ affin
- $X_0 \subseteq \mathbb{R}^n$ konvex
- $(X_0, (g_i)_{i \in [m]}, (h_i)_{i \in [p]})$ reguläres Tripel
- $X \subseteq \mathbb{R}^n$ sei die Menge aller $x \in X_0$ mit
  - $g_i(x) \leq 0$ für alle $i \in [m]$ und
  - $h_i(x) = 0$ für alle $i \in [p]$.
- $f$ stetig in wenigstens einem Punkt von $X$

# . . . Karush-Kuhn-Tucker (konvex)

**Satz 3.17**

*Ein Punkt* $x^\star \in X$ *ist genau dann Optimallösung von*

$$\min\{f(x) \,|\, x \in X\}\,,$$

*wenn es Multiplizierer* $\lambda_1, \ldots \lambda_m \in \mathbb{R}_+$ *und* $\mu_1, \ldots, \mu_p \in \mathbb{R}$ *mit*

$$\Big(\mathrm{SGRAD}_{x^\star}(f) + \sum_{i=1}^{m} \lambda_i \,\mathrm{SGRAD}_{x^\star}(g_i) + \sum_{i=1}^{p} \mu_i \,\mathrm{grad}_{x^\star} h_i\Big) \cap \big(-\mathrm{N}_{x^\star}(X_0)\big) \neq \varnothing$$

*und*

$$\lambda_i = 0 \quad \textit{für alle } i \in [m] \textit{ mit } g_i(x^\star) < 0$$

*gibt.*

# Differenzierbare nicht-konvexe Probleme

- Sind $f, g^{(1)}, \dots, g^{(m)}, h^{(1)}, \dots, h^{(p)} : \mathbb{R}^n \to \mathbb{R}$ zwar stetig differenzierbar, aber (vielleicht) nicht konvex, so sind (unter geeigneten Regularitätsbedingungen) die KKT-Bedingungen in Satz 3.17 *notwendig* für das Vorliegen eines lokalen Minimums.
- Z.B.: [Ruszczyński, Thm. 3.25]
- Herleitung: *Tangentialkegel* statt Radialkegel