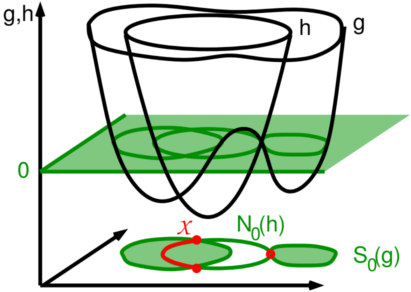
Inhalt

7.1 Restringierte Nichtlineare Optimierung

Aufgabenstellung:

$$(P)\quad \begin{array}{llll} \min & f(x) & & \text{Zielfunktion} \\ \text{s.t.} & h_i(x) = 0 & i \in \mathcal{E} & \text{Gleichungsnebenbedingungen} \\ & g_i(x) \le 0 & i \in \mathcal{I} & \text{Ungleichungsnebenbedingungen} \\ & x \in \mathbb{R}^n & & \text{Grundmenge (meist } \mathbb{R}^n\text{, manchmal } [0,1]^n) \end{array}$$

mit f, g_i, $h_i : \mathbb{R}^n \to \mathbb{R}$ hinreichend glatt und $\mathcal{E}$, $\mathcal{I}$ endliche Indexmengen. Die zulässige Menge $\mathcal{X} := \{x \in \mathbb{R}^n : h_i(x) = 0\ (i \in \mathcal{E}), g_i(x) \le 0\ (i \in \mathcal{I})\}$ ist Schnitt der Niveaulinien/-mengen, $\mathcal{X} = \bigcap_{i\in\mathcal{E}} N_0(h_i)\ \cap\ \bigcap_{i\in\mathcal{I}} S_0(g_i)$.



7.1 Restringierte Nichtlineare Optimierung

Aufgabenstellung:

$$(P)\qquad \begin{array}{llll} \min & f(x) & & \text{Zielfunktion} \\ \text{s.t.} & h_i(x) = 0 & i \in \mathcal{E} & \text{Gleichungsnebenbedingungen} \\ & g_i(x) \le 0 & i \in \mathcal{I} & \text{Ungleichungsnebenbedingungen} \\ & x \in \mathbb{R}^n & & \text{Grundmenge (meist } \mathbb{R}^n\text{, manchmal } [0,1]^n) \end{array}$$

mit f, g_i, $h_i : \mathbb{R}^n \to \mathbb{R}$ hinreichend glatt und $\mathcal{E}$, $\mathcal{I}$ endliche Indexmengen.
Die zulässige Menge $\mathcal{X} := \{x \in \mathbb{R}^n : h_i(x) = 0 \ (i \in \mathcal{E}), g_i(x) \le 0 \ (i \in \mathcal{I})\}$ ist Schnitt der Niveaulinien/-mengen, $\mathcal{X} = \bigcap_{i\in\mathcal{E}} N_0(h_i) \ \cap \ \bigcap_{i\in\mathcal{I}} S_0(g_i)$.
Wir suchen ein lokal optimales $x^* \in \mathcal{X}$, also ein x^* mit

$$f(x^*) \le f(x) \quad \forall x \in \mathcal{X} \cap B_\varepsilon(x^*) := \{x : \|x - x^*\| \le \varepsilon\} \text{ für ein } \varepsilon > 0.$$

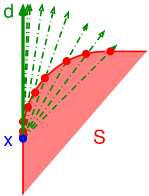
Für Algorithmen ist diese Beschreibung ungeeignet. Wir benötigen eine algebraische Beschreibung der lokalen Optimalität, die für gegebenes $\bar{x} \in \mathbb{R}^n$ nur auf den Funktionswerten und Gradienten (und eventuell ∇^2) von f, g_i, h_i in $\bar{x}$ beruht (Orakel 1./2. Ordnung für f, g_i, h_i).
Idee: Beschreibe die Richtungen, in die man sich von $\bar{x}$ aus innerhalb $\mathcal{X}$ noch ein kleines Stück bewegen kann. Ist die Richtungsableitung in all diesen Richtungen positiv, gibt es in der Nähe keinen besseren Punkt.

Inhalt

7.2 Zulässige Richtungen und Tangentialkegel

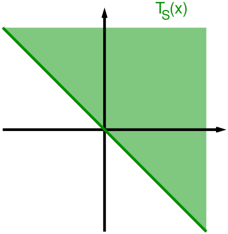
Für eine Menge $S \subseteq \mathbb{R}^n$ und ein $x \in S$ heißt $d \in \mathbb{R}^n$ zulässige Richtung (Tangentialrichtung), wenn es eine Folge $x^{(k)} \in S$ $(k \in \mathbb{N})$ mit $\lim_{k\to\infty} x^{(k)} = x$ und $\alpha_k \geq 0$ gibt mit $d = \lim_{k\to\infty} \alpha_k(x^{(k)} - x)$.

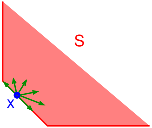
[Die α_k strecken die Vektoren auf die Länge von d]



Die Menge aller zulässigen Richtungen für ein x in S bildet den Tangentialkegel $T_S(x)$ von S in x.

[$T_S(x)$ ist Kegel, denn für $\alpha \geq 0$ ist mit $d \in T_S(x)$ auch $\alpha d \in T_S(x)$: Für $d = \lim_{k\to\infty} \alpha_k(x^{(k)} - x)$ ist $\alpha d = \lim_{k\to\infty} \bar{\alpha}_k(x^{(k)} - x)$ mit $\bar{\alpha}_k := \alpha\alpha_k$.]

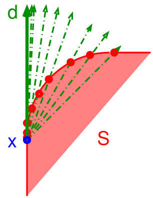




7.2 Zulässige Richtungen und Tangentialkegel

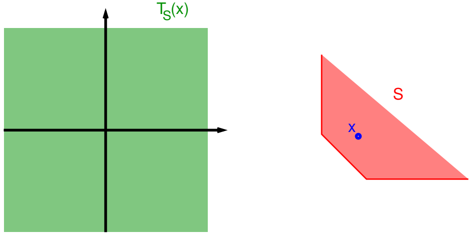
Für eine Menge $S \subseteq \mathbb{R}^n$ und ein $x \in S$ heißt $d \in \mathbb{R}^n$ zulässige Richtung (Tangentialrichtung), wenn es eine Folge $x^{(k)} \in S$ $(k \in \mathbb{N})$ mit $\lim_{k\to\infty} x^{(k)} = x$ und $\alpha_k \geq 0$ gibt mit $d = \lim_{k\to\infty} \alpha_k(x^{(k)} - x)$.

[Die α_k strecken die Vektoren auf die Länge von d]



Die Menge aller zulässigen Richtungen für ein x in S bildet den Tangentialkegel $T_S(x)$ von S in x.

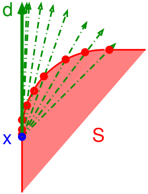
[$T_S(x)$ ist Kegel, denn für $\alpha \geq 0$ ist mit $d \in T_S(x)$ auch $\alpha d \in T_S(x)$: Für $d = \lim_{k\to\infty} \alpha_k(x^{(k)} - x)$ ist $\alpha d = \lim_{k\to\infty} \bar{\alpha}_k(x^{(k)} - x)$ mit $\bar{\alpha}_k := \alpha\alpha_k$.]



7.2 Zulässige Richtungen und Tangentialkegel

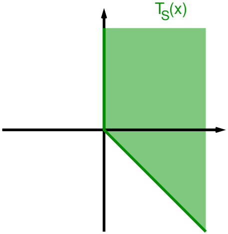
Für eine Menge $S \subseteq \mathbb{R}^n$ und ein $x \in S$ heißt $d \in \mathbb{R}^n$ zulässige Richtung (Tangentialrichtung), wenn es eine Folge $x^{(k)} \in S$ $(k \in \mathbb{N})$ mit $\lim_{k\to\infty} x^{(k)} = x$ und $\alpha_k \geq 0$ gibt mit $d = \lim_{k\to\infty} \alpha_k(x^{(k)} - x)$.

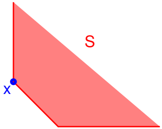
[Die α_k strecken die Vektoren auf die Länge von d]



Die Menge aller zulässigen Richtungen für ein x in S bildet den Tangentialkegel $T_S(x)$ von S in x.

[$T_S(x)$ ist Kegel, denn für $\alpha \geq 0$ ist mit $d \in T_S(x)$ auch $\alpha d \in T_S(x)$: Für $d = \lim_{k\to\infty} \alpha_k(x^{(k)} - x)$ ist $\alpha d = \lim_{k\to\infty} \bar{\alpha}_k(x^{(k)} - x)$ mit $\bar{\alpha}_k := \alpha\alpha_k$.]

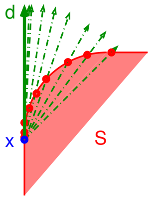




7.2 Zulässige Richtungen und Tangentialkegel

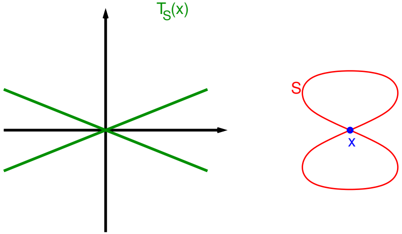
Für eine Menge $S \subseteq \mathbb{R}^n$ und ein $x \in S$ heißt $d \in \mathbb{R}^n$ zulässige Richtung (Tangentialrichtung), wenn es eine Folge $x^{(k)} \in S$ ($k \in \mathbb{N}$) mit $\lim_{k\to\infty} x^{(k)} = x$ und $\alpha_k \geq 0$ gibt mit $d = \lim_{k\to\infty} \alpha_k(x^{(k)} - x)$.

[Die α_k strecken die Vektoren auf die Länge von d]



Die Menge aller zulässigen Richtungen für ein x in S bildet den Tangentialkegel $T_S(x)$ von S in x.

[$T_S(x)$ ist Kegel, denn für $\alpha \geq 0$ ist mit $d \in T_S(x)$ auch $\alpha d \in T_S(x)$: Für $d = \lim_{k\to\infty} \alpha_k(x^{(k)} - x)$ ist $\alpha d = \lim_{k\to\infty} \bar{\alpha}_k(x^{(k)} - x)$ mit $\bar{\alpha}_k := \alpha\alpha_k$.]



Der Polarkegel

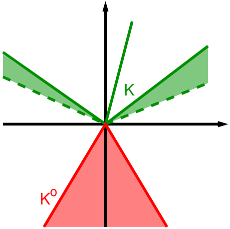
Ist $K \subseteq \mathbb{R}^n$ ein Kegel, nennt man

$$K^\circ := \{p \in \mathbb{R}^n : p^T d \le 0 \quad \forall d \in K\}$$

den Polarkegel zu K.

[Für K konvex ist $K^\circ = -K^*$ der negative Dualkegel. Er wird auch Normalkegel zu K genannt.]

Beachte: Aus $K \subseteq \bar{K}$ folgt $\bar{K}^\circ \subseteq K^\circ$.



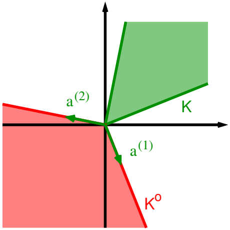
Gegeben $K = \{d \in \mathbb{R}^n : (a^{(i)})^T d \le 0, i = 1, \ldots, m\}$, was ist K°?

Setze dazu $A := [a^{(1)}, \ldots, a^{(m)}]$.

$$p \in K^\circ \Leftrightarrow$$

$$\begin{array}{rl} 0 = \max & p^T d \\ \text{s.t.} & A^T d \le 0 \\ & d \text{ frei} \end{array} \quad \Leftrightarrow \quad \begin{array}{rl} 0 = \min & 0^T \lambda \\ \text{s.t.} & A\lambda = p \\ & \lambda \ge 0 \end{array}$$

$$\Leftrightarrow \exists \lambda \ge 0 : A\lambda = p$$



Also ist $K^\circ = \{p \in \mathbb{R}^n : p = \sum_{i=1}^m \lambda_i a^{(i)}, \lambda_i \ge 0\}$.

Inhalt

7.3 Notwendige Optimalitätsbedingung 1. Ordnung

Sei $\mathcal{X} \subseteq \mathbb{R}^n$ und betrachte: $(P) \quad \min_{x\in\mathcal{X}} f(x)$

Ist $x^* \in \mathcal{X}$ lokales Optimum und $d \in T_{\mathcal{X}}(x^*)$ mit $d = \lim \alpha_k(x^{(k)} - x^*)$, dann gilt, wegen $\mathcal{X} \ni x^{(k)} \to x^*$, für große k

$$\begin{aligned}
& f(x^*) \leq f(x^{(k)}) = f(x^*) + \nabla f(x^*)^T(x^{(k)} - x^*) + \mathbf{o}(\|x^{(k)} - x^*\|) \\
\Rightarrow \quad & \nabla f(x^*)^T(x^{(k)} - x^*) \geq -\mathbf{o}(\|x^{(k)} - x^*\|) \qquad | \cdot \alpha_k \geq 0 \\
\Rightarrow \quad & \underbrace{\nabla f(x^*)^T[\alpha_k(x^{(k)} - x^*)]}_{\lim \to \nabla f(x^*)^T d} \geq -\underbrace{\mathbf{o}(\alpha_k\|x^{(k)} - x^*\|)}_{\lim \to \mathbf{o}(\|d\|)=0}
\end{aligned}$$

Satz (Notwendige Optimalitätsbedingung 1. Ordnung)

Ist x^* ein lokales Optimum von (P), so ist jede Richtungsableitung in eine zulässige Richtung nichtnegativ,

$$\nabla f(x^*)^T d \geq 0 \quad \forall d \in T_{\mathcal{X}}(x^*).$$

[Wie in der freien Optimierung ist dies nicht hinreichend!]

Äquivalent: $-\nabla f(x^*) \in T_{\mathcal{X}}(x^*)^\circ$, der negative Gradient liegt im Polarkegel.

Ein $\bar{x} \in \mathcal{X}$ mit $-\nabla f(\bar{x}) \in T_{\mathcal{X}}(\bar{x})^\circ$ heißt stationärer Punkt von (P).

Das Ziel der Optimierungsalgorithmen ist, stationäre Punkte zu finden, aber lässt sich $T_{\mathcal{X}}(x)$ durch die Gradienten in x beschreiben?

Inhalt

7.4 Der linearisierte Tangentialkegel

Schreibt man $\mathcal{X} = \{x \in \mathbb{R}^n : h_i(x) = 0\ (i \in \mathcal{E}), g_i(x) \leq 0\ (i \in \mathcal{I})\}$ als

$\mathcal{X} = \bigcap_{i \in \mathcal{E}} N_0(h_i)\ \cap\ \bigcap_{i \in \mathcal{I}} S_0(g_i)$, sieht man leicht

$$d \in T_{\mathcal{X}}(x) \quad \Rightarrow \quad d \in T_{N_0(h_i)}(x)\ (i \in \mathcal{E}) \text{ und } d \in T_{S_0(g_i)}(x)\ (i \in \mathcal{I}).$$

[Alle $x^{(k)} \in \mathcal{X}$ für d sind auch in den anderen Mengen.]

7.4 Der linearisierte Tangentialkegel

Schreibt man $\mathcal{X} = \{x \in \mathbb{R}^n : h_i(x) = 0\ (i \in \mathcal{E}), g_i(x) \leq 0\ (i \in \mathcal{I})\}$ als

$$\mathcal{X} = \bigcap_{i\in\mathcal{E}} N_0(h_i)\ \cap\ \bigcap_{i\in\mathcal{I}} S_0(g_i), \quad \text{sieht man leicht}$$

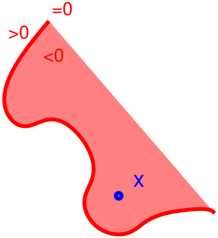
$$d \in T_{\mathcal{X}}(x) \quad \Rightarrow \quad d \in T_{N_0(h_i)}(x)\ (i \in \mathcal{E}) \text{ und } d \in T_{S_0(g_i)}(x)\ (i \in \mathcal{I}).$$

Für jedes einzelne h_i und g_i ist der Tangentialkegel meist gut darstellbar:

<u>1. $g_i(x) < 0$:</u> $T_{S_0(g_i)}(x) = \mathbb{R}^n$

Keine Einschränkung der Richtungen!

Eine Ungleichung mit $g_i(x) < 0$ heißt inaktiv (in x).



7.4 Der linearisierte Tangentialkegel

Schreibt man $\mathcal{X} = \{x \in \mathbb{R}^n : h_i(x) = 0\ (i \in \mathcal{E}), g_i(x) \le 0\ (i \in \mathcal{I})\}$ als

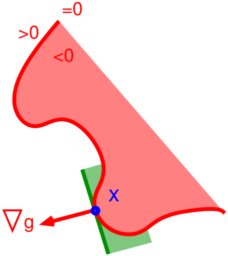
$$\mathcal{X} = \bigcap_{i\in\mathcal{E}} N_0(h_i) \ \cap\ \bigcap_{i\in\mathcal{I}} S_0(g_i), \quad \text{sieht man leicht}$$

$$d \in T_{\mathcal{X}}(x) \quad\Rightarrow\quad d \in T_{N_0(h_i)}(x)\ (i \in \mathcal{E}) \ \text{ und } \ d \in T_{S_0(g_i)}(x)\ (i \in \mathcal{I}).$$

Für jedes einzelne h_i und g_i ist der Tangentialkegel meist gut darstellbar:

1. $g_i(x) < 0$: $T_{S_0(g_i)}(x) = \mathbb{R}^n$
Keine Einschränkung der Richtungen!
Eine Ungleichung mit $g_i(x) < 0$ heißt inaktiv (in x).

2. $g_i(x) = 0$: $T_{S_0(g_i)}(x) = \{d \in \mathbb{R}^n : \nabla g_i(x)^T d \le 0\}$,
falls $\nabla g_i(x) \ne 0$! Ist $\nabla g_i(x) = 0$, gilt $T_{S_0(g_i)}(x) \subseteq \mathbb{R}^n$
und Gleichheit nur, wenn x lokales Maximum von g_i ist!
Eine Ungleichung mit $g_i(x) = 0$ heißt aktiv (in x).

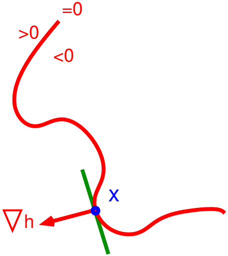


7.4 Der linearisierte Tangentialkegel

Schreibt man $\mathcal{X} = \{x \in \mathbb{R}^n : h_i(x) = 0\ (i \in \mathcal{E}), g_i(x) \le 0\ (i \in \mathcal{I})\}$ als $\mathcal{X} = \bigcap_{i\in\mathcal{E}} N_0(h_i)\ \cap\ \bigcap_{i\in\mathcal{I}} S_0(g_i)$, sieht man leicht

$$d \in T_{\mathcal{X}}(x) \quad \Rightarrow \quad d \in T_{N_0(h_i)}(x)\ (i \in \mathcal{E}) \text{ und } d \in T_{S_0(g_i)}(x)\ (i \in \mathcal{I}).$$

Für jedes einzelne h_i und g_i ist der Tangentialkegel meist gut darstellbar:



1. $g_i(x) < 0$: $T_{S_0(g_i)}(x) = \mathbb{R}^n$
Keine Einschränkung der Richtungen!
Eine Ungleichung mit $g_i(x) < 0$ heißt inaktiv (in x).

2. $g_i(x) = 0$: $T_{S_0(g_i)}(x) = \{d \in \mathbb{R}^n : \nabla g_i(x)^T d \le 0\}$,
falls $\nabla g_i(x) \neq 0$! Ist $\nabla g_i(x) = 0$, gilt $T_{S_0(g_i)}(x) \subseteq \mathbb{R}^n$ und Gleichheit nur, wenn x lokales Maximum von g_i ist!
Eine Ungleichung mit $g_i(x) = 0$ heißt aktiv (in x).

3. $h_i(x) = 0$: $T_{N_0(h_i)}(x) = \{d \in \mathbb{R}^n : \nabla h_i(x)^T d = 0\}$,
falls $\nabla h_i(x) \neq 0$ (für $\nabla h_i(x) = 0$ nur $T_{N_0(h_i)}(x) \subseteq \mathbb{R}^n$).

Im Folgenden bezeichnet $\mathcal{A}(x) = \{i \in \mathcal{I} : g_i(x) = 0\}$ die aktive Menge.
Der Schnitt der von den Gradienten der h_i und aktiven g_i induzierten Unter- und Halbräume ist der linearisierte Tangentialkegel zu (P) in x,

$$\boxed{T_P(x) := \{d \in \mathbb{R}^n : \nabla h_i(x)^T d = 0\ (i \in \mathcal{E}), \nabla g_i(x)^T d \le 0\ (i \in \mathcal{A}(x))\}.}$$

$$\boxed{T_P(x)^\circ := \{p \in \mathbb{R}^n : p = \sum_{i\in\mathcal{E}} \mu_i \nabla h_i(x) + \sum_{i\in\mathcal{A}(x)} \lambda_i \nabla g_i(x), \mu \in \mathbb{R}^{\mathcal{E}}, \lambda \in \mathbb{R}_+^{\mathcal{A}(x)}\}.}$$

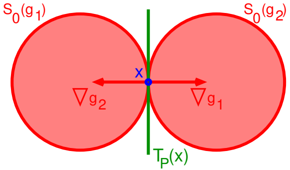
Wann ist $T_{\mathcal{X}}(x) = T_P(x)$?

Wir wissen, $T_{\mathcal{X}}(x) \subseteq T_P(x)$ für alle x, aber für ein bestimmtes $\bar{x}$ kann $T_P(\bar{x})$ zu groß sein, wenn z.B. $\nabla g_i(\bar{x}) = 0$ für ein $i \in \mathcal{A}(\bar{x})$ oder wenn $\bar{x}$ ungünstig am Rand mehrerer Niveaumengen liegt. Dann gilt nur $T_P(x)^\circ \subseteq T_{\mathcal{X}}(x)^\circ$ und für ein Minimum $\bar{x}$ mit $-\nabla f(\bar{x}) \in T_{\mathcal{X}}(\bar{x})^\circ$ ist eventuell $-\nabla f(\bar{x}) \notin T_P(\bar{x})^\circ$, also ist $\bar{x}$ schlecht als stationär erkennbar!

Bsp: $\bar{x} = 0$ und 2 Nebenbedingungen:
$g_1(x) := \frac{1}{2}\|x + e_1\|^2 - 1 \le 0$, $\nabla g_1(\bar{x}) = e_1$,
$g_2(x) := \frac{1}{2}\|x - e_1\|^2 - 1 \le 0$, $\nabla g_2(\bar{x}) = -e_1$,

$$\Rightarrow\ T_P(\bar{x}) = \{d \in \mathbb{R}^n : e_1^T d \le 0, -e_1^T d \le 0\}$$
$$= \{d \in \mathbb{R}^n : d_1 = 0\},$$

aber $T_{\mathcal{X}}(\bar{x}) = \{0\}$. Betrachte $f(x) := x_1^2 + x_2$.



In $\bar{x} \in \mathcal{X}$ ist die Regularitätsbedingung der linearen Unabhängigkeit (*linear independence constraint qualification*, kurz LICQ) erfüllt, wenn die Gradienten $\nabla h_i(\bar{x})$ $(i \in \mathcal{E})$ und $\nabla g_i(\bar{x})$ $(i \in \mathcal{A}(\bar{x}))$ linear unabhängig sind.

Satz

Für ein $\bar{x} \in \mathcal{X}$ sei (LICQ) erfüllt, dann gilt $T_{\mathcal{X}}(\bar{x}) = T_P(\bar{x})$.

[Der Beweis zeigt für jedes $d \in T_P(\bar{x})$ die Existenz einer geeigneten Folge $x^{(k)} \in \mathcal{X}$ mittels des Satzes über implizite Funktionen.]

Inhalt

7.5 Existenz von Lagrange-Multiplikatoren

Ein Minimum x^*, das (LICQ) erfüllt, kann über $-\nabla f(x^*) \in T_P(x^*)^\circ$ als stationär erkannt werden.

Satz (Karush-Kuhn-Tucker)

Sei x^* ein lokales Minimum von (P), in dem (LICQ) erfüllt ist. Dann gibt es (eindeutige) Lagrange-Multiplikatoren $\mu^* \in \mathbb{R}^{\mathcal{E}}$ und $\lambda^* \in \mathbb{R}_+^{\mathcal{I}}$, sodass

$$\begin{aligned} \nabla f(x^*) + \sum_{i\in\mathcal{E}} \mu_i^* \nabla h_i(x^*) + \sum_{i\in\mathcal{I}} \lambda_i^* \nabla g_i(x^*) &= 0 \\ \lambda_i^* g_i(x^*) &= 0 \quad i \in \mathcal{I} \quad \textit{[Kompl.]} \end{aligned}$$

Beweis: Da x^* lokales Minimum ist, gilt $-\nabla f(x^*) \in T_{\mathcal{X}}(x^*)^\circ$.
Wegen (LICQ) ist $T_P(x^*)^\circ = T_{\mathcal{X}}(x^*)^\circ$, also gibt es $\mu^* \in \mathbb{R}^{\mathcal{E}}$ und $\lambda^* \in \mathbb{R}^{\mathcal{A}(x^*)}$ mit $-\nabla f(x^*) = \sum_{i\in\mathcal{E}} \mu_i^* \nabla h_i(x^*) + \sum_{i\in\mathcal{A}(x^*)} \lambda_i^* \nabla g_i(x^*)$.
Für $i \in \mathcal{I} \setminus \mathcal{A}(x^*)$ setze $\lambda_i^* := 0$. □

„In einem lokalen Minimum, in dem (LICQ) erfüllt ist, liegt der negative Gradient der Zielfunktion im Kegel, der von den Gradienten der aktiven Nebenbedingungen aufgespannt wird.“

Warum „Lagrange“-Multiplikatoren?

Lagrange-Funktion und Sattelpunkte

Bestrafung der Verletzung von $h_i(x) = 0$ mit Multiplikatoren $\mu_i \in \mathbb{R}$ und von $g_i(x) \le 0$ mit $\lambda_i \ge 0$ in der Zielfunktion ergibt die Lagrange-Funktion

$$\mathcal{L}(x, \mu, \lambda) = f(x) + \sum_{i \in \mathcal{E}} \mu_i h_i(x) + \sum_{i \in \mathcal{I}} \lambda_i g_i(x).$$

Es gilt $\quad v(P) = \inf_{x \in \mathbb{R}^n} \sup_{\mu \in \mathbb{R}^{\mathcal{E}}, \lambda \in \mathbb{R}^{\mathcal{I}}_+} \mathcal{L}(x, \mu, \lambda)$

[$\bar{x} \in \mathcal{X}$: $\sup_{\mu \in \mathbb{R}^{\mathcal{E}}, \lambda \in \mathbb{R}^{\mathcal{I}}_+} \mathcal{L}(\bar{x}, \mu, \lambda) = f(\bar{x})$, da $\lambda_i = 0$ für $g_i(\bar{x}) < 0$ optimal,
$\bar{x} \notin \mathcal{X}$: $\sup_{\mu \in \mathbb{R}^{\mathcal{E}}, \lambda \in \mathbb{R}^{\mathcal{I}}_+} \mathcal{L}(\bar{x}, \mu, \lambda) = \infty$, da $h_i(\bar{x}) \neq 0$ oder $g_i(\bar{x}) > 0$ für ein i]

Lagrange-Funktion und Sattelpunkte

Bestrafung der Verletzung von $h_i(x) = 0$ mit Multiplikatoren $\mu_i \in \mathbb{R}$ und von $g_i(x) \leq 0$ mit $\lambda_i \geq 0$ in der Zielfunktion ergibt die Lagrange-Funktion

$$\mathcal{L}(x, \mu, \lambda) = f(x) + \sum_{i \in \mathcal{E}} \mu_i h_i(x) + \sum_{i \in \mathcal{I}} \lambda_i g_i(x).$$

Es gilt $\quad v(P) = \inf_{x \in \mathbb{R}^n} \sup_{\mu \in \mathbb{R}^{\mathcal{E}}, \lambda \in \mathbb{R}_+^{\mathcal{I}}} \mathcal{L}(x, \mu, \lambda)$

Ein Punkt $(\bar{x}, \bar{\mu}, \bar{\lambda}) \in \mathbb{R}^n \times \mathbb{R}^{\mathcal{E}} \times \mathbb{R}_+^{\mathcal{I}}$ heißt (lokaler) Sattelpunkt von $\mathcal{L}$, falls

$$\mathcal{L}(x, \bar{\mu}, \bar{\lambda}) \;\geq\; \mathcal{L}(\bar{x}, \bar{\mu}, \bar{\lambda}) \;\geq\; \mathcal{L}(\bar{x}, \mu, \lambda)$$

für alle $\lambda \geq 0$, μ und alle x in einer Umgebung von $\bar{x}$.

Eigenschaften von Sattelpunkten $(\bar{x}, \bar{\mu}, \bar{\lambda})$ der Lagrange-Funktion:

$(\bar{\mu}, \bar{\lambda})$ ist Maximum des linearen Problems $\max_{\mu \in \mathbb{R}^{\mathcal{E}}, \lambda \in \mathbb{R}_+^{\mathcal{I}}} \mathcal{L}(\bar{x}, \mu, \lambda)$:

$$\nabla_\mu \mathcal{L}(\bar{x}, \bar{\mu}, \bar{\lambda}) = 0 \;\Leftrightarrow\; h_i(\bar{x}) = 0$$

$$\nabla_\lambda \mathcal{L}(\bar{x}, \bar{\mu}, \bar{\lambda}) \leq 0 \text{ und } \nabla_\lambda \mathcal{L}(\bar{x}, \bar{\mu}, \bar{\lambda})^T \bar{\lambda} = 0 \;\Leftrightarrow\; g_i(\bar{x}) \leq 0 \text{ und } g_i(\bar{x}) \bar{\lambda}_i = 0$$

[Komplementarität: $(i \notin \mathcal{A}(\bar{x}) \Rightarrow \bar{\lambda}_i = 0)$ und $(\bar{\lambda}_i > 0 \Rightarrow g_i(\bar{x}) = 0)$]

$\bar{x}$ ist lokales Minimum des unrestringierten Problems $\min_{x \in \mathbb{R}^n} \mathcal{L}(x, \bar{\mu}, \bar{\lambda})$:

$$\nabla_x \mathcal{L}(\bar{x}, \bar{\mu}, \bar{\lambda}) = 0 \;\Leftrightarrow\; \nabla f(\bar{x}) + \textstyle\sum_{i \in \mathcal{E}} \bar{\mu}_i \nabla h_i(\bar{x}) + \sum_{i \in \mathcal{I}} \bar{\lambda}_i \nabla g_i(\bar{x}) = 0,$$

also $-\nabla f(\bar{x}) \in T_P(\bar{x})^\circ \subseteq T_{\mathcal{X}}(\bar{x})^\circ$ und $\bar{x}$ ist lokales Minimum von (P).

Die KKT-Bedingungen

Optimierungsverfahren suchen nach Lösungen der KKT-Bedingungen

$$
(KKT) \quad
\begin{array}{lrcll}
\nabla_x \mathcal{L} = & \nabla f(x) + \sum \mu_i \nabla h_i(x) + \sum \lambda_i \nabla g_i(x) & = & 0, & \\
\nabla_\mu \mathcal{L} = & h_i(x) & = & 0, & i \in \mathcal{E}, \\
\nabla_\lambda \mathcal{L} = & g_i(x) & \le & 0, & i \in \mathcal{I}, \\
 & \lambda_i & \ge & 0, & i \in \mathcal{I}, \\
 & g_i(x)\lambda_i & = & 0, & i \in \mathcal{I},
\end{array}
$$

also nach stationären Punkten $(x, \mu, \lambda) \in \mathbb{R}^n \times \mathbb{R}^{\mathcal{E}} \times \mathbb{R}^{\mathcal{I}}_+$ der Lagrange-Funktion. Jeder solche Punkt ist auch stationärer Punkt von (P) (ein Sattelpunkt sogar lokales Minimum), und jeder stationäre Punkt von (P), in dem (LICQ) erfüllt ist, lässt sich so finden.

Gilt $T_P(x^*) = T_{\mathcal{X}}(x^*)$ für ein lokales Optimum x^*, dann gibt es Lagrange-Multiplikatoren (μ^*, λ^*), sodass (x^*, μ^*, λ^*) die KKT-Bedingungen erfüllt. (LICQ) ist dafür hinreichend, aber nicht notwendig. Es gibt schwächere Bedingungen und in Spezialfällen sind keine notwendig:

- Sind alle Nebenbedingungen g_i und h_i linear/affin, dann gilt $T_P(x) = T_{\mathcal{X}}(x)$ für alle $x \in \mathcal{X}$.
- Sind alle h_i affin, alle g_i konvex und existiert ein Slater-Punkt $\bar{x} \in \mathcal{X}$ mit $g_i(\bar{x}) < 0$ $(i \in \mathcal{I})$, dann gilt $T_P(x) = T_{\mathcal{X}}(x)$ für alle $x \in \mathcal{X}$.

Spezialfall: Opt.-Bed. für (glatte) konvexe Optimierung

Ein (glattes) konvexes Optimierungsproblem hat die Gestalt

$$(CP)\quad \begin{array}{lllll} \min & f(x) & & & f \text{ konvex, glatt} \\ \text{s.t.} & h_i(x) = 0 & i \in \mathcal{E} & & \text{jedes } h_i \text{ affin} \\ & g_i(x) \leq 0 & i \in \mathcal{I} & & \text{jedes } g_i \text{ konvex, glatt} \\ & x \in \Omega & & & \mathbb{R}^n \text{ oder „einfaches" Polyeder} \end{array}$$

Die zulässige Menge $\mathcal{X}$ ist Schnitt konvexer Mengen, also konvex.

Satz

Sei (CP) ein konvexes Optimierungsproblem und es gebe einen Slater-Punkt oder alle g_i seien affin, dann sind die KKT-Bedingungen notwendig und hinreichend für die Optimalität.

Bsp: Innere-Punkte-Verfahren für LP $\min\{c^Tx : Ax = b, x \geq 0\}$.
$f(x) := c^Tx - \mu \log x$ (μ Barriereparameter!), $h_i(x) := [b - Ax]_i$, $\Omega := \mathbb{R}^n_+$,
Lagrange-Funktion (Multiplikator y für h): $\mathcal{L}(x, y) := f(x) + y^T(b - Ax)$
KKT-Bedingungen (notwendig und hinreichend für $x > 0$):
$\nabla_x \mathcal{L} = c - \mu x^{-1} - A^T y = 0$
$\nabla_y \mathcal{L} = b - Ax = 0$
Mit $z := \mu x^{-1}$ bzw. $z \circ x = \mu \mathbf{1} \quad \longrightarrow \quad$ primal-duales KKT-System.

Inhalt

7.6 Notwendige Optimalitätsbedingung 2. Ordnung

Sei x^* ein lokales Minimum von (P), in dem (LICQ) erfüllt ist, und seien (μ^*, λ^*) die Lagrange-Multiplikatoren. Da $\nabla f(x^*)^T d \geq 0$ für alle $d \in T_P(x^*) = T_{\mathcal{X}}(x^*)$, sind nun die Richtungen interessant mit

$$0 = d^T \nabla f(x^*) \overset{(KKT)}{=} \sum \mu_i^* \underbrace{d^T \nabla h_i(x^*)}_{=0} + \sum \lambda_i^* \underbrace{d^T \nabla g_i(x^*)}_{\leq 0},$$

also Richtungen aus dem Teilkegel

$$T_P'(x^*, \lambda^*) := \{d \in T_P(x^*) : d^T \nabla g_i(x^*) = 0 \text{ mit } \lambda_i^* > 0 \ (i \in \mathcal{A}(x^*))\}.$$

Satz über implizite Funktionen: Für jedes $d \in T_P'$ gibt es $x^{(k)} \in \mathcal{X}$, $\alpha_k \geq 0$ mit $d = \lim \alpha_k (x^{(k)} - x^*)$ und $f(x^{(k)}) = \mathcal{L}(x^{(k)}, \mu^*, \lambda^*)$.
Setze $h^{(k)} := x^{(k)} - x^* \to 0$, nutze $\mathcal{L}_* := \mathcal{L}(x^*, \mu^*, \lambda^*) = f(x^*) =: f_*$,

$$\mathcal{L}(x^* + h^{(k)}, \mu^*, \lambda^*) = \underbrace{\mathcal{L}_*}_{=f_*} + \underbrace{\nabla_x \mathcal{L}_*^T}_{=0 \ (KKT)} h^{(k)} + \frac{1}{2} \underbrace{(h^{(k)})^T \nabla_{xx} \mathcal{L}_* h^{(k)}}_{\Rightarrow \lim \to d^T \nabla_{xx} \mathcal{L}_* d \geq 0} + \mathbf{o}(\|h\|^2) \geq f_*$$

Satz (Notwendige Optimalitätsbedingung 2. Ordnung)

Ist x^* ein lokales Minimum von (P), das (LICQ) erfüllt, und sind (μ^*, λ^*) die Lagrange-Multiplikatoren zu x^*, dann gilt

$$d^T \nabla_{xx} \mathcal{L}(x^*, \mu^*, \lambda^*) d \geq 0 \quad \text{für alle } d \in T_P'(x^*, \lambda^*).$$

Hinreichende Optimalitätsbedingungen

Wieder nutzt man, dass die Lagrange-Funktion für die korrekten Multiplikatoren ein gutes unrestringiertes Modell für (P) um x^* ist.

Satz (Hinreichende Optimalitätsbedingungen)

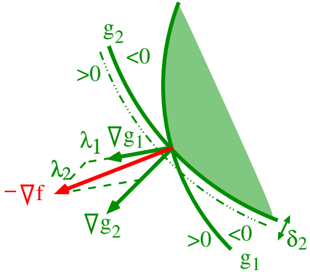
Erfüllt $x^* \in \mathcal{X}$ mit (μ^*, λ^*) die KKT-Bedingungen und gilt

$$d^T \nabla_{xx} \mathcal{L}(x^*, \mu^*, \lambda^*) d > 0 \quad \text{für alle } d \in T'_P(x^*, \lambda^*) \setminus \{0\},$$

so ist x^* ein lokales Minimum von (P).

Inhalt

7.7 Sensitivität



Wie ändert sich die Lösung, wenn man an den rechten Seiten der Nebenbedingungen wackelt?

Sind die Multiplikatoren in einer Umgebung der Lösung eindeutig (dies benötigt (LICQ) und strenge Komplementarität, $\lambda_i^* > 0 \Leftrightarrow g_i(x^*) = 0$), geben sie dazu viel Information.

Satz (Sensitivität von Optimallösungen)

Für $\delta \in \mathbb{R}^{\mathcal{E}\cup\mathcal{I}}$ bezeichne (P_δ)

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & h_i(x) = \delta_i, \quad i \in \mathcal{E}, \\ & g_i(x) \le \delta_i, \quad i \in \mathcal{I}, \\ & x \in \mathbb{R}^n. \end{array}$$

Für (P_0) erfülle ein Punkt x^* (LICQ), die hinr. Opt.-Bed. und strenge Komplementarität bzgl. der Lagrange-Mult. (μ^*, λ^*). Dann gibt es eine Umgebung $U \subseteq \mathbb{R}^{\mathcal{E}\cup\mathcal{I}}$ um $\delta = 0$ und eine stetige Funktion $x^*(\delta)$ mit $x^*(0) = x^*$ und der Eigenschaft, dass für jedes $\delta \in U$ der Punkt $x^*(\delta)$ lokales Minimum von (P_δ) ist und $\nabla_\delta[f(x^*(\cdot))](0) = -\begin{bmatrix} \mu^* \\ \lambda^* \end{bmatrix}$.

Für sehr kleine δ ändert sich der Funktionswert also um etwa $-\delta^T \begin{bmatrix} \mu^* \\ \lambda^* \end{bmatrix}$.

Ein gutes Maß für den Einfluss einer Ungleichung ist $\lambda_i \|\nabla g_i\|$ (Skalierung!).