10 Optimalitätsbedingungen, Box-Beschränkungen

Nichtlineare Optimierung

WS 2020/21

Quizfrage

Wir betrachten die Optimierungsaufgabe

$$\begin{aligned} \text{Minimiere} \quad & x_1 + x_2 \quad , \quad x \in \mathbb{R}^2 \\ \text{unter} \quad & x_2^2 - x_1 \leqslant 0 \\ \text{und} \quad & x_1 \leqslant 10 \end{aligned}$$

Wie nennt man die Menge

$$X := \{x \in \mathbb{R}^2 \,|\, x_2^2 - x_1 \leqslant 0,\; x_1 \leqslant 10\} \; ?$$

0 A zufällige Menge

0 B Lösungsmenge

0 C aktive Menge

7 D zulässige Menge

Quizfrage

Im welchem der folgenden Punkte ist die Nebenbedingung $x_2^2 - x_1 \leqslant 0$ aktiv?

$g(x) = 0$

1 A $x = (2,1)^T$

2 B $x = (4,2)^T$

0 C $x = (1,2)^T$

4 D $x = (0,0)^T$

inaktiv

A: $g(x) = 1^2 - 2 < 0$

B: $g(x) = 2^2 - 4 = 0$

C: $g(x) = 2^2 - 1 > 0$

verletzt

D: $g(x) = 0^2 - 0 = 0$

Quizfrage

Im welchem der folgenden Punkte ist die Nebenbedingung $x_2^2 - x_1 \leqslant 0$ inaktiv?

A $x = (2, 1)^\mathsf{T}$

B $x = (4, 2)^\mathsf{T}$

C $x = (1, 2)^\mathsf{T}$

D $x = (0, 0)^\mathsf{T}$

Quizfrage

Im welchem der folgenden Punkte ist die Nebenbedingung $x_2^2 - x_1 \leqslant 0$ verletzt?

A $x = (2, 1)^\mathsf{T}$

B $x = (4, 2)^\mathsf{T}$

C $x = (1, 2)^\mathsf{T}$

D $x = (0, 0)^\mathsf{T}$

Quizfrage

oft schwer handhabbar

Wozu braucht man den Tangentialkegel (an die zulässige Menge X im Punkt x)

Folge

von x nach x_k

$$\mathcal{T}_X(x) = \left\{ d \in \mathbb{R}^n \;\middle|\; \text{es gibt } \{x_k\} \subset X,\ \underbrace{t_k \searrow 0}_{\text{pos. Nullfolge}},\ \frac{x_k - x}{t_k} \to d \right\}$$

Richtung

notwendigerweise: $x_k \to x$

in der Optimierung?

4 A Für lokale Minimierer x^* gilt: $\nabla f(x^*)^\top d \geqslant 0$ für alle $d \in \mathcal{T}_X(x^*)$.

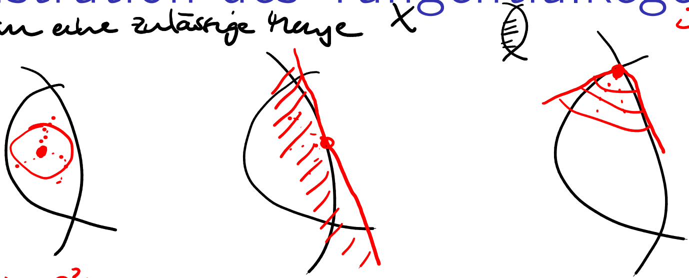
0 B Der Tangentialkegel beschreibt die Menge X lokal.

1 C Für lokale Minimierer x^* gilt: $\nabla f(x^*)^\top d \leqslant 0$ für alle $d \in \mathcal{T}_X(x^*)$.

1 D Der Tangentialkegel enthält die lokalen Minimierer.

Illustration des Tangentialkegels

an eine zulässige Menge X im Punkt x



$T_X(x) = \mathbb{R}^2$

$T_X(x)$ = Halbraum (abgeschlossen) in $\mathbb{R}^2$

$T_X(x)$ = Sektor in $\mathbb{R}^2$

tangentiale Richtungen d:

$(x_k) \subset X$, $t_k \downarrow 0$, $d = \lim\limits_{k \to \infty} \dfrac{x_k - x}{t_k}$

Quizfrage

leichter handhabbar, berechenbar

Wozu braucht man den Linearisierungskegel (an die zulässige Menge X im Punkt x)

$$\mathcal{T}_X^{\text{lin}}(x) = \left\{ d \in \mathbb{R}^n \;\middle|\; \begin{array}{l} \nabla g_i(x)^\mathsf{T} d \leqslant 0 \text{ für alle } i \in \mathcal{A}(x), \\ \nabla h_j(x)^\mathsf{T} d = 0 \text{ für alle } j = 1, \ldots, p \end{array} \right\}$$

in der Optimierung?

4 **A** Für lokale Minimierer x^* gilt: $\nabla f(x^*)^\mathsf{T} d \geqslant 0$ für alle $\boxed{d \in \mathcal{T}_X^{\text{lin}}(x^*).}$ Das sind io zulässige Richtungen!

0 **B** Der Linearisierungskegel beschreibt die Menge X lokal. Das hätte man gerne!

1 **C** Für lokale Minimierer x^* gilt: $\nabla f(x^*)^\mathsf{T} d \leqslant 0$ für alle $d \in \mathcal{T}_X^{\text{lin}}(x^*)$.

1 **D** Der Lineariserungskegel enthält die lokalen Minimierer.

Illustration des Linearisierungskegels

hängt von der Beschreibung der Menge X ab

$X = \{0\} \subset \mathbb{R} \Rightarrow J_X(0) = \{0\}$

$\boxed{h(x) = x}$

$\nabla h(x) = 1$

$\nabla h(0) = 1$

$J_X^{lin}(0) = \{d \in \mathbb{R} : 1 \cdot d = 0\} = \{0\}$

$= J_X(0)$

$J_X^{lin}(0)^\circ = J_X(0)^\circ = \mathbb{R}$

$\boxed{h(x) = x^2}$

$\nabla h(x) = 2x$

$\nabla h(0) = 0$

$J_X^{lin}(0) = \{d \in \mathbb{R} : 0 \cdot d = 0\} = \mathbb{R}$

$\supsetneq J_X(0)$

$J_X^{lin}(0)^\circ = \{0\}$

$J_X(0)^\circ = \mathbb{R}$

$\boxed{g(x) = x^2}$

$\nabla g(x) = 2x$

$\nabla g(0) = 0$

$J_X^{lin}(x) = \{d \in \mathbb{R} : 0 \cdot d \le 0\} = \mathbb{R}$

$\supsetneq J_X(0)$

$\longleftrightarrow$ dito

Notwendige Optimalitätsbedingung

Es sei x^* ein lokales Minimum von

Minimiere $f(x)$ unter $g(x) \le 0$, $h(x) = 0$.

<u>Dann gilt:</u> $\nabla f(x^*)^T d \geq 0$ für alle $d \in \mathcal{T}_X(x^*)$.

$\Leftrightarrow$ $\boxed{-\nabla f(x^*) \in \mathcal{T}_X(x^*)^\circ}$ „duale Beschreibung"

Da $\mathcal{T}_X(x^*)$ und damit auch $\mathcal{T}_X(x^*)^\circ$ i.A. schwer handhabbar sind, würde man lieber $\mathcal{T}_X^{lin}(x^*)$ bzw. $\mathcal{T}_X^{lin}(x^*)^\circ$ verwenden.

<u>Aber:</u> $\mathcal{T}_X(x^*) \subseteq \mathcal{T}_X^{lin}(x^*)$ $\Rightarrow$ $\boxed{\mathcal{T}_X^{lin}(x^*)^\circ \subseteq \mathcal{T}_X(x^*)^\circ}$

Es ist daher $\boxed{\text{i.A.}}$ nicht zu erwarten, dass $-\nabla f(x^*)$ auch in $\mathcal{T}_X^{lin}(x^*)^\circ$ liegt!

Gleichheit gilt unter einer constraint <u>qualification</u> $\boxed{\text{bei } x^*}$

Quizfrage

Wozu braucht man die Lagrange-Funktion

$$\mathcal{L}(x,\mu,\lambda) = f(x) + \mu^\top g(x) + \lambda^\top h(x) \text{ ?}$$

0 A zur Darstellung von $\mathcal{T}_X(x)^\circ$

1 B Man minimiert sie an Stelle der Zielfunktion.

0 C zur Darstellung von $\mathcal{T}_X^{\text{lin}}(x)^\circ$

Beweis mit Farkas-Lemma

7 D zur Darstellung der KKT-Bedingungen

$\mathcal{T}_X^{\text{lin}}(x)^\circ = \{$ LK der $\nabla h_j(x)$ und nicht-neg. LK der $\nabla g_i(x)$, aber nur der aktiven $\}$

$= \{ \sum_{j=1}^{p} \lambda_j \nabla h_j(x) + \sum_{i=1}^{m} \mu_i \nabla g_i(x),\ \mu_i \ge 0$ für $i \in \mathcal{A}(x)$ und $\mu_i = 0$ für $i \in \mathcal{I}(x) \}$

Zusammenfassung Optimalitätsbed.

In einem lokalen Minimum x^* gilt:

$-\nabla f(x^*) \in \mathcal{T}_X(x^*)^\circ = \mathcal{T}_X^{\text{lin}}(x^*)^\circ$

↑ unter Annahme einer CQ

$\Leftrightarrow$ die KKT-Bedingungen sind erfüllt

Algorithmen versuchen die zu erfüllen!

KKT-Bedingungen $\Leftrightarrow$ $\boxed{-\nabla f(x^*) \in \mathcal{T}_X^{\text{lin}}(x^*)^\circ}$

$$\left.\begin{aligned} \nabla_x \mathcal{L}(x^*, \underbrace{\lambda, \mu}_{\text{Lagrange-Multiplikatoren}}) &= \nabla f(x^*) + g'(x^*)^T \mu + h'(x^*)^T \lambda = 0 \\ \mu \ge 0, \quad g(x^*) &\le 0, \quad \mu^T g(x^*) = 0 \\ h(x^*) &= 0 \end{aligned}\right\}$$

Komplementarität

Quizfrage

Warum sind Aufgaben mit Ungleichungsnebenbedingungen algorithmisch schwieriger als Aufgaben mit ausschließlich Gleichungsnebenbedingungen?

0 **A** Die Aussage stimmt nicht, denn Gleichungen $h(x) = 0$ entsprechen ja sogar zwei Ungleichungen $h(x) \leqslant 0$ und $-h(x) \leqslant 0$.

0 **B** Der Tangentialkegel bei Aufgaben mit Ungleichungsnebenbedingungen ist komplizierter.

0 **C** Aufgaben mit Ungleichungen haben nicht unbedingt eine Lösung.

6 **D** Man weiß vorher nicht, welche Ungleichungen in der Lösung aktiv sein werden.

Quizfrage

LICQ $\Rightarrow$ MFCQ $\Rightarrow$ ACQ $\Rightarrow$ GCQ

Welche Constraint Qualifications sind für Aufgaben mit ausschließlich Box-Beschränkungen in jedem zulässigen Punkt erfüllt?

$$\begin{aligned} \text{Minimiere} \quad & f(x) \\ \text{unter} \quad & \underbrace{\ell}_{\in \mathbb{R}^n} \leqslant x \leqslant \underbrace{u}_{\in \mathbb{R}^n} \end{aligned}$$

0 A GCQ

4 B ACQ

0 C MFCQ, wenn $\ell < u$ gilt

0 D LICQ, wenn $\ell < u$ gilt

MFCQ: es ex. $d \in \mathbb{R}^n$ mit $\nabla g_i(x)^T d < 0$ für alle $i \in \mathcal{A}(x)$.

$g(x) = \begin{pmatrix} \ell - x \\ x - u \end{pmatrix} \Rightarrow$ „$\nabla g_i(x)^T d = \pm d_i$" Ist möglich, wenn $\ell < u$.

Quizfrage

Was versteht man unter einem projizierten Gradientenverfahren?

A Man verwendet als Suchrichtung d eine Abänderung des Gradienten, sodass $x + \alpha d$ für alle $\alpha \geqslant 0$ zulässig bleibt.

B Man verwendet den üblichen Gradienten als Suchrichtung d und projiziert $x + \alpha d$ innerhalb der Liniensuche auf $[\ell, u]$.

C Das ist ein normales Gradientenverfahren, wobei man Verletzungen von $\ell \leqslant x \leqslant u$ bestraft.

D Das ist ein normales Gradientenverfahren, wobei man abwechselnd auf die Mengen $\{x \in \mathbb{R}^n : \ell \leqslant x\}$ und $\{x \in \mathbb{R}^n : x \leqslant u\}$ projiziert.

Zeit für Ihre Frage

Was sind Ihre Fragen zu den Themen der Woche?

$\longrightarrow$ Benutzen Sie den Chat.

Fragen und Antworten 1

Fragen und Antworten 2