TECHNISCHE UNIVERSITÄT DRESDEN

Fakultät Mathematik Institut für Numerik, Professur für Numerik der linearen Optimierung

PROSEMINAR: NUMERIK

Numerische Verfahren zur Lösung unrestringierter Optimierungsaufgaben

Prof. Dr. Andreas Fischer

Sommersemester 2019

Autor : Eric Kunze
E-Mail : eric.kunze@mailbox.tu-dresden.de

Inhaltsverzeichnis

Vortrag 2
KONVEXE FUNKTIONEN

Eric Kunze

In diesem Vortrag beschäftigen wir uns mit der Klasse der konvexen Funktionen, deren Rolle in der Optimierung in späteren Vorträgen behandelt werden wird.
Zunächst benötigen wir eine gewisse Struktur auf den Definitionsbereichen von konvexen Funktionen. Diese wird durch *konvexe Mengen* realisiert.

Definition 2.1 (konvexe Menge)

Eine Menge $X \subseteq \mathbb{R}^n$ heißt **konvex**, wenn für alle $x, y \in X$ und alle $\lambda \in (0,1)$ auch

$$\lambda x + (1-\lambda)y \in X \tag{2.1}$$

Eine solche Linearkombination $\sum_{i=1}^n \mu_i x_i$ von Vektoren $x_i \in \mathbb{R}^n$ mit $\mu_i \in (0,1)$ für alle $i \in \{1, \dots, n\}$ und $\sum_{i=1}^n \mu_i = 1$ nennt man auch **Konvexkombination**

Anschaulich kann die Definition wie folgt gedeutet werden: Eine Menge ist genau dann konvex, wenn jede Verbindungsstrecke zweier Punkte der Menge wieder vollständig in der Menge liegt.

Ausgehend davon können wir nun eine neue Klasse von Funktionen definieren.

Definition 2.2 (konvexe Funktion)

Sei $X \subseteq \mathbb{R}^n$ eine konvexe Menge. Eine Funktion $f\colon X \to \mathbb{R}$ heißt

- **konvex** (auf X), wenn für alle $x, y \in X$ und alle $\lambda \in (0,1)$ gilt, dass
$$f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda) f(y) \tag{2.2}$$
- **strikt konvex** (auf X), wenn für alle $x, y \in X$ mit $x \neq y$ und alle $\lambda \in (0,1)$ gilt, dass
$$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda) f(y) \tag{2.3}$$
- **gleichmäßig konvex** (auf X), wenn es ein $\mu > 0$ gibt mit
$$f(\lambda x + (1-\lambda)y) + \mu\lambda(1-\lambda)\left\|x-y\right\|^2 \le \lambda f(x) + (1-\lambda) f(y) \tag{2.4}$$
für alle $x, y \in X$ und alle $\lambda \in (0,1)$.

Bemerkung 2.3

Der Faktor μ in (2.4) wird auch als **Modulus** bezeichnet. Man sagt dann, dass f *gleichmäßig konvex mit Modulus* μ ist.

Analog ließe sich auch die Eigenschaft Konkavität definieren, indem das Relationszeichen umgedreht wird. Jedoch ist f genau dann konkav, wenn $-f$ konvex ist, d.h. Konkavität lässt sich immer mittles Definition 2.2 nachweisen.

Im folgenden wollen wir auf die explizite Angabe der konvexen Menge verzichten, wenn dies aus dem Kontext klar wird.

Anschauliche Bedeutung: Ist f konvex, so liegt kein Punkt einer Verbindungsstrecke von zwei Punkten $(x, f(x)), (y, y(x)) \in \operatorname{graph}(f) \subseteq \mathbb{R}^{n+1}$ unterhalb des Graphen von f.

Aus Zeitgründen wird das folgende Lemma nur erwähnt.

Lemma

Sei $X \subseteq \mathbb{R}^n$ eine konvexe Menge und $f\colon X \to \mathbb{R}$. Dann gilt

$$f \text{ gleichmäßig konvex} \quad \Rightarrow \quad f \text{ strikt konvex} \quad \Rightarrow \quad f \text{ konvex}$$

Beweis. Dies folgt unmittelbar aus Definition 2.2. □

Insbesondere gilt im Allgemeinen keine weitere Implikation in obigem Lemma was durch Beispiele im Laufe des Vortrags widerlegt werden wird.

Beispiel 2.4

(i) Die Gerade $f(x) := x$ ist konvex, aber nicht strikt konvex.

(ii) Die Parabel $f(x) := x^2$ ist gleichmäßig konvex. Dagegen ist $g(x) := x^4$ strikt konvex, nicht jedoch gleichmäßig konvex.

Beweis. (i) Für $f(x) = x$ gilt in (2.2)

$$f(\lambda x + (1-\lambda)y) = \lambda x + (1-\lambda)y = \lambda \cdot f(x) + (1-\lambda)f(y)$$

Insbesondere gilt dann $\leq$. Die obige Gleichung zeigt aber auch, dass f nicht strikt konvex ist, da stets auch Gleichheit gilt. Dies ist auch ein Beispiel dafür, dass nicht jede konvexe Funktion auch strikt konvex ist.

(ii) Die gleichmäßige Konvexität von f folgt später einfach aus Lemma 2.5. Wir wollen hier insbesondere die gleichmäßige Konvexität von g widerlegen. Nehmen wir also an g sei gleichmäßig konvex. Dann gilt dies insbesondere auch für $y = 0$ und $\lambda = 1/2$. Somit ergibt sich

$$g(\lambda x + (1-\lambda)y) = g(0.5x) = 0.5^4 \cdot x^4 = 1/16 x^4$$
$$\lambda g(x) + (1-\lambda)g(y) = 0.5g(x) = 0.5x^4$$

Um dies zum Widerspruch zu führen, suchen wir alle $x \neq 0$, für die die Ungleichung (2.4) nicht gilt, d.h.

$$\frac{1}{16}x^4 + \frac{1}{4}\mu x^2 > \frac{1}{2}x^4 \quad \Rightarrow \quad \mu > \frac{7}{4}x^2 \quad \Rightarrow \quad |x| < \sqrt{\frac{4}{7}\mu}$$

Damit finden wir für alle $\mu > 0$ also ein $x \neq 0$, sodass die Ungleichung in der falschen Richtung erfüllt ist.
□

Eine besondere Stellung nehmen die quadratischen Funktionen ein, für die weitere Implikationen ausgehend vom Lemma oben gelten.

Lemma 2.5

Sei $f\colon \mathbb{R}^n \to \mathbb{R}$ eine quadratische Funktion mit

$$f(x) := \frac{1}{2}x^\top Qx + c^\top x + \gamma$$

mit einer symmetrischen Matrix $Q \in \mathbb{R}^{n\times n}$, einem Vektor $c \in \mathbb{R}^n$ und einer Konstante $\gamma \in \mathbb{R}$. Dann gilt:

(i) f ist konvex $\;\Leftrightarrow\;$ Q ist positiv semidefinit.

(ii) f ist strikt konvex $\;\Leftrightarrow\;$ f ist gleichmäßig konvex $\;\Leftrightarrow\;$ Q ist positiv definit.

Beweis. Der Beweis geht von der folgenden Beobachtung aus:

$$\begin{aligned}
\lambda f(x) + (1-\lambda) f(y) &= \frac{1}{2}\lambda x^\top Qx + \lambda c^\top x + \lambda\gamma + \frac{1}{2}(1-\lambda) y^\top Qy + (1-\lambda)c^\top y + (1-\lambda)\gamma \\
&= \frac{1}{2}\lambda x^\top Qx + \frac{1}{2}(1-\lambda) y^\top Qy + c^\top(\lambda x + (1-\lambda)y) + \gamma \\
&= f(\lambda x + (1-\lambda)y) \\
&\quad - \frac{1}{2}\left(\lambda x + (1-\lambda)y\right)^\top Q(\lambda x + (1-\lambda)y) + \\
&\quad \frac{1}{2}\lambda x^\top Qx + \frac{1}{2}(1-\lambda) y^\top Qy \\
&= f(\lambda x + (1-\lambda)y) + \underbrace{\frac{1}{2}\lambda(1-\lambda)}_{>0}\left(x-y\right)^\top Q(x-y)
\end{aligned}$$

Somit ist also f genau dann konvex, wenn $(x-y)^\top Q(x-y) \geq 0$ für alle $x, y \in \mathbb{R}^n$, was äquivalent zur positiven Semidefinitheit von Q ist. Analog ist f genau dann strikt konvex, wenn $(x-y)^\top Q(x-y) > 0$ für alle $x \neq y$, was wiederum bedeutet, dass Q positiv definit ist.

Da Q symmetrisch und positiv definit ist, existiert eine Orthonormalbasis $\mathcal{B} = (v_1, \ldots, v_n)$ von $\mathbb{R}^n$ aus Eigenvektoren von Q mit zugehörigen Eigenwerten θ_i $(i = 1, \ldots, n)$. Für jedes $z \in \mathbb{R}^n$ existieren dann α_i $(i = 1, \ldots, n)$, so dass $z = \sum_{i=1}^n \alpha_i \cdot v_i$. Sei $\theta_{\min}$ der kleinste (positive) Eigenwert von Q. Dann gilt

$$\begin{aligned}
\langle z\,,\, Qz\rangle &= \left\langle \sum_{i=1}^n \alpha_i \cdot v_i\,,\, A \cdot \sum_{i=1}^n \alpha_i \cdot v_i \right\rangle = \left\langle \sum_{i=1}^n \alpha_i \cdot v_i\,,\, \sum_{i=1}^n \alpha_i \cdot (Av_i) \right\rangle = \left\langle \sum_{i=1}^n \alpha_i \cdot v_i\,,\, \sum_{i=1}^n \theta_i \cdot \alpha_i \cdot v_i \right\rangle \\
&= \sum_{i=1}^n \theta_i \cdot \alpha_i^2 \geq \theta_{\min} \cdot \sum_{i=1}^n \alpha_i^2 = \theta_{\min} \cdot \langle z\,,\, z\rangle = \theta_{\min} \cdot z^\top z = \theta_{\min} \cdot \|z\|
\end{aligned}$$

Mit $z = x - y$ folgt daraus nun

$$(x-y)^\top Q(x-y) \geq \theta_{\min} \cdot \|x-y\|^2$$

Somit ist f gleichmäßig konvex mit Modulus $\mu = 2\theta_{\min}$. □

Satz 2.6

Seien $X \subseteq \mathbb{R}^n$ eine offene und konvexe Menge sowie $f\colon X \to \mathbb{R}$ stetig differenzierbar. Dann gelten:

(i) f ist genau dann konvex (auf X), wenn für alle $x, y \in X$

$$f(x) - f(y) \geq \nabla f(y)^\top (x-y) \tag{2.5}$$

gilt

(ii) f ist genau dann strikt konvex (auf X), wenn für alle $x, y \in X$ mit $x \neq y$

$$f(x) - f(y) > \nabla f(y)^\top (x-y) \tag{2.6}$$

gilt

(iii) f ist genau dann gleichmäßig konvex (auf X), wenn es ein $\mu > 0$ gibt mit

$$f(x) - f(y) \geq \nabla f(y)^\top (x-y) + \mu \|x-y\|^2 \tag{2.7}$$

für alle $x, y \in X$.

Beweis. Wir zeigen zuerst die Rückrichtungen. Dabei wollen wir insbesondere die gleichmäßige Konvexität betrachten, die beiden anderen Teile folgen dann aus dieser.

Gelte also (2.7). Seien $x, y \in X$ und $\lambda \in (0,1)$ beliebig. Setzen wir nun als z als Konvexkombination $z := \lambda x + (1-\lambda)y \in X$. Wegen (2.7) gilt dann

$$f(x) - f(z) \geq \nabla f(z)^\top (x-z) + \mu \|x-z\|^2 \tag{2.8}$$

und

$$f(y) - f(z) \geq \nabla f(z)^\top (y-z) + \mu \|y-z\|^2 \tag{2.9}$$

Nun multiplizieren wir (2.8) mit λ und (2.9) mit $(1-\lambda)$. Anschließend addieren wir beide Ungleichungen und erhalten für die linke Seite

$$\lambda f(x) - \lambda f(z) + (1-\lambda) f(y) - (1-\lambda) f(z) = \lambda f(x) + (1-\lambda) f(y) - f(z)$$

bzw. für den ersten Teil der rechten Seite wegen $z = \lambda x + (1-\lambda) y$

$$\begin{aligned}
\lambda \nabla f(z)^\top (x-z) + (1-\lambda) \nabla f(z)^\top (y-z) &= \nabla f(z)^\top \left(\lambda(x-z) + (1-\lambda)(y-z)\right) \\
&= \nabla f(z)^\top \left(\lambda x - \lambda z + y - z - \lambda y + \lambda z\right) \\
&\overset{\text{Def. } z}{=} \nabla f(z)^\top \left(\lambda(x-y) + y - \lambda x - (1-\lambda) y\right) \\
&= \nabla f(z)^\top \left(\lambda x - \lambda y + y - (1-\lambda) y\right) \\
&= \nabla f(z)^\top \cdot \left(y \underbrace{(-\lambda + 1 - 1 + \lambda)}_{=0} \right)
\end{aligned}$$

Beachten wir nun noch

$$x - z = x - \lambda x - (1-\lambda) y = (1-\lambda)(x-y) \quad \text{und} \quad y - z = y - \lambda x - (1-\lambda) y = \lambda(y-x)$$

dann gilt für die "Norm-Terme" der rechten Seite

$$\begin{aligned}
\lambda\mu \|x-z\|^2 + +\mu \|y-z\|^2 &= \mu \left(\lambda \|(1-\lambda)(x-y)\|^2 + (1-\lambda) \|\lambda(y-x)\|^2\right) \\
&= \mu \left(\lambda(1-\lambda)^2 \|(x-y)\|^2 + (1-\lambda)\lambda^2 \|(x-y)\|^2\right) \\
&= \mu \|x-y\|^2 \left(\lambda(1-\lambda)(1-\lambda+\lambda)\right) \\
&= \lambda(1-\lambda)\mu \|x-y\|^2
\end{aligned}$$

Somit gilt für die Ungleichung mit $z = \lambda x + (1-\lambda) y$ schließlich wieder

$$\lambda f(x) + (1-\lambda) f(y) - f(\lambda x + (1-\lambda) y) \geq \lambda(1-\lambda)\mu \|x-y\|^2$$

und damit ist f gleichmäßig konvex. Analog zeigt man nun auch, dass aus (2.5) bzw. (2.6) die (strikte) Konvexität folgt. Somit sind alle Rückrichtungen gezeigt.

Für die Hinrichtungen setzen wir nun f als gleichmäßig konvex voraus. Dann existiert ein $\mu > 0$, sodass für alle $x, y \in X$ und alle $\lambda \in (0,1)$

$$\begin{aligned}
f(y + \lambda(x-y)) &= f(\lambda x + (1-\lambda) y) \\
&\leq \lambda f(x) + (1-\lambda) f(y) - \mu\lambda(1-\lambda) \|x-y\|^2
\end{aligned}$$

gilt. Division mit λ ergibt schließlich

$$\frac{f(y + \lambda(x-y)) - f(y)}{\lambda} \leq f(x) - f(y) - \mu(1-\lambda) \|x-y\|^2$$

Da f nun stetig differenzierbar ist folgt für $\lambda \searrow 0$

$$\nabla f(y)^\top (x-y) = \lim_{\lambda \searrow 0} \frac{f(y + \lambda(x-y)) - f(y)}{\lambda} \leq f(x) - f(y) - \mu \|x-y\|^2 \tag{2.10}$$

Dies entspricht gerade der Aussage in (2.7). Mit $\mu = 0$ ergibt sich ebenso die Aussage in (2.5).

Um (2.6) zu zeigen, müssen wir anders argumentieren, da im Übergang zur Grenze die strikte Relation verloren ginge. Sei nun also f strikt konvex und $x, y \in X$ mit $x \neq y$. Insbesondere ist f dann konvex, d.h. es gilt (2.5). Definieren wir nun

$$z := \frac{1}{2}(x+y) = \frac{1}{2}x + \left(1 - \frac{1}{2}\right) y$$

ergibt sich wegen $2(z-y) = 2 \cdot (1/2x + 1/2y - y) = 2 \cdot (1/2x - 1/2y) = x - y$

$$\nabla f(y)^\top (x-y) = 2 \nabla f(y)^\top (z-y) \overset{(2.5)}{\leq} 2(f(z) - f(y)) \tag{2.11}$$

Jedoch ist $x \neq y$. Aus der strikten Konvexität von f folgt nun

$$f(z) = f\left(\frac{1}{2}x + \left(1 - \frac{1}{2}\right)\right) < \frac{1}{2} f(x) + \frac{1}{2} f(y) \tag{2.12}$$

Aus (2.11) und (2.12) folgt damit $\nabla f(y)^\top (x-y) < f(x) - f(y)$ also gerade (2.6) $\square$

Vortrag 3

KONVEXITÄT & MONOTONIE

Friedemann Krannich

3.1 Wiederholung

Wir erinnern uns an die folgenden Sätze und Definitionen:

Definition 3.1 (Konvexität)
Sei $X \subseteq \mathbb{R}^n$ konvex, $f : X \to \mathbb{R}$ eine Funktion.

(1) f ist konvex (auf X), falls $\forall x, y \in X \; \forall \lambda \in (0,1)$:

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

(2) f ist strikt konvex (auf X): nutze in 1. $<$ statt $\leq$

(3) f ist gleichmäßig konvex (auf X), falls $\exists \mu > 0 : \; \forall x, y \in X \; \forall \lambda \in (0,1)$:

$$f(\lambda x + (1-\lambda)y) + \mu\lambda(1-\lambda)\|x-y\| \leq \lambda f(x) + (1-\lambda) f(y)$$

Satz 3.2 (Zusammenhang Konvexität und Ableitung einer Funktion)
Sei $X \subseteq \mathbb{R}^n$ offen und konvex, $f : X \to \mathbb{R}$ stetig diffbar

(1) f ist konvex (auf X) $\Leftrightarrow \forall x, y \in X : f(x) - f(y) \geq f'(y)^T(x-y)$

(2) f ist strikt konvex (auf X) $\Leftrightarrow \forall x \neq y \in X : f(x) - f(y) > f'(y)^T(x-y)$

(3) f ist gleichmäßig konvex (auf X) $\Leftrightarrow \exists \mu > 0 : \forall x, y \in X$:
$f(x) - f(y) \geq f'(y)^T(x-y) + \mu\|x-y\|^2$

3.2 Monotonie einer Funktion

Definition 3.3 (Monotonie)
Sei $X \subseteq \mathbb{R}^n$. Eine Funktion $f : X \to \mathbb{R}^n$ heißt

(1) monoton (auf X), falls $\forall x, y \in X$:

$$(x-y)^T(f(x) - f(y)) \geq 0$$

(2) strikt monoton (auf X), falls $\forall x, y \in X$ mit $x \neq y$:

$$(x-y)^T(f(x) - f(y)) > 0$$

(3) gleichmäßig monoton (auf X) falls $\exists \mu > 0 : \forall x, y \in X$:

$$(x-y)^T (f(x) - f(y)) \geq \mu \|x-y\|^2$$

Bemerkung 3.4

f gleichmäßig monoton $\Rightarrow$ f strikt monoton
f strikt monoton $\Rightarrow$ f monoton

> **Satz 3.5 (Zusammenhang Monotonie und Konvexität einer Funktion)**
> $X \subseteq \mathbb{R}^{\mathrm{n}}$ offen und konvex, Funktion f : $X \to \mathbb{R}$ stetig differenzierbar
>
> (1) f konvex $\Leftrightarrow$ f' monoton
>
> (2) f strikt konvex $\Leftrightarrow$ f' strikt monoton
>
> (3) f gleichmäßig konvex $\Leftrightarrow$ f' gleichmäßig monoton

Beweis. zu 1. und 2.:
"$\Rightarrow$" : Sei f konvex. Satz 1.2 1. liefert:

$$\forall x, y \in X : f(x) - f(y) \geq f'(y)^T (x-y) \tag{3.1}$$

sowie

$$\forall x, y \in X : f(y) - f(x) \geq f'(x)^T (y-x) \tag{3.2}$$

Addition von (**??**) und (**??**) liefert

$$0 \geq (f'(y) - f'(x))^T (x-y)$$

äquivalent zu

$$0 \geq -(f'(x) - f'(y))^T (x-y)$$

was wiederum

$$(f'(x) - f'(y))^T (x-y) \geq 0$$

impliziert, also ist f' monoton. Für strenge Konvexität bzw. strenge Monotonie funktioniert der Beweis analog, nutze $>$ statt $\geq$.
"$\Leftarrow$" : Seien $x, y \in X$ beliebig aber fest. Der Mittelwertsatz liefert:

$$\exists \theta \in (0,1) : f(x) - f(y) = f'(\xi)^T (x-y) \text{ mit } \xi = y + \theta(x-y) \in X \tag{3.3}$$

Da f' monoton ist folgt mit der Definition von ξ

$$0 \leq (\xi - y)^T (f'(\xi) - f'(y)) = \theta (x-y)^T (f'(\xi) - f'(y)) \tag{3.4}$$

(**??**) und (**??**) zusammen liefern

$$\begin{aligned} f(x) - f(y) &= f'(\xi)^T (x-y) - f'(y)^T (x-y) + f'(y)^T (x-y) \\ &= \theta^{-1} (f'(\xi) - f'(y))^T \theta (x-y) + f'(y)^T (x-y) \geq f'(y)^T (x-y) \end{aligned}$$

mit Satz 1.2 1. folgt dann die Konvexität von f.
Der Beweis, dass strikte Monotonie von f' strikte Konvexität von f impliziert erfolgt analog unter Nutzung von $>$ statt $\geq$.
Zu 3.: "$\Rightarrow$" : f gleichmäßig konvex. Satz 1.2 3. liefert $\exists \mu > 0 : \forall x, y \in X$:

$$f(x) - f(y) \geq f'(y)^T (x-y) + \mu \|x-y\|^2$$

sowie

$$f(y) - f(x) \geq f'(x)^T (y-x) + \mu \|x-y\|^2$$

Addition der beiden Ungleichungen liefert:

$$\begin{aligned} 0 &\geq f'(y)^T (x-y) + f'(x)^T (y-x) + 2\mu \|x-y\|^2 \\ \Rightarrow 0 &\geq f'(y)^T (x-y) - f'(x)^T (x-y) + 2\mu \|x-y\|^2 \\ \Rightarrow 0 &\geq (f'(y) - f'(x)^T (x-y) + 2\mu \|x-y\|^2 \\ \Rightarrow (x-y)(f'(x) - f'(y))^T &\geq 2\mu \|x-y\|^2 \end{aligned}$$

$\Rightarrow$ f' ist gleichmäßig monoton

"$\Leftarrow$" : f' gleichmäßig monoton, d.h. $\exists \mu > 0 : \forall x, y \in X$:

$$(x-y)^T(f'(x) - f'(y)) \geq \mu \|x-y\|^2$$

Sei nun $x, y \in X$ fest, $m \in \mathbb{N}$ beliebig aber fest.
Definiere $t_k := \frac{k}{m+1} \quad k \in \{0, ..., m+1\}$.
Nach dem Mittelwertsatz existiert ein $\theta_k \in (t_k, t_{k+1})$ mit $\xi_k = y + \theta_k(x-y)$ mit

$$f(y + t_{k+1}(x-y)) - f(y + t_k(x-y)) = (t_{k+1} - t_k) f'(\xi_k)^T (x-y)$$

$$\Rightarrow f(x) - f(y) = \sum_{k=0}^{m} [f(y + t_{k+1}(x-y)) - f(y + t_k(x-y))]$$

$$= \sum_{k=0}^{m} (t_{k+1} - t_k) f'(\xi_k)^T (x-y) = f'(y)^T (x-y) + \sum_{k=0}^{m} (t_{k+1} - t_k)(f'(\xi_k) - f'(y))^T (x-y)$$

$$= f'(y)^T(x-y) + \sum_{k=0}^{m} \frac{t_{k+1} - t_k}{\theta_k} (f'(\xi_k) - f'(y))^T (\xi_k - y) \text{ nach der Definition von } \xi_k$$

$$\geq f'(y)^T(x-y) + \mu \sum_{k=0}^{m} \frac{t_{k+1} - t_k}{\theta_k} \|\xi_k - y\|^2 \text{ da } f' \text{ gleichmäßig monoton ist}$$

$$= f'(y)^T(x-y) + \mu \|x-y\|^2 \sum_{k=0}^{m} \theta_k (t_{k+1} - t_k) \text{ nach der Definition von } \xi_k.$$

Da $\theta_k \in (t_k, t_{k+1})$ folgt

$$\sum_{k=0}^{m} \theta_k (t_{k+1} - t_k) \geq \sum_{k=0}^{m} t_k (t_{k+1} - t_k)$$

$$= \sum_{k=0}^{m} \frac{k}{m+1} \left(\frac{k+1}{m+1} - \frac{k}{m+1} \right) = \frac{1}{(m+1)^2} \sum_{k=0}^{m} k = \frac{1}{2} \frac{m}{m+1}$$

$$\Rightarrow f(x) - f(y) \geq f'(y)^T (x-y) + \frac{1}{2} \mu \frac{m}{m+1} \|x-y\|^2$$

Mit $m \to \infty$ folgt

$$f(x) - f(y) \geq f'(y)^T (x-y) + \frac{1}{2} \mu \|x-y\|^2$$

Mit Satz 1.2 3. folgt, dass f gleichmäßig konvex ist. $\square$

Vortrag 4
KONVEXITÄT & DEFINITHEIT

Klemens Fritzsche

4.1 Wiederholung

Definition 4.1
Sei $X \subseteq \mathbb{R}^n$ eine konvexe Menge. Eine Funktion $f\colon X \to \mathbb{R}$ heißt

(a) **konvex** (auf X), wenn für alle $x, y \in X$ und alle $\lambda \in (0,1)$ gilt:

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y);$$

(b) **strikt konvex** (auf X), wenn für alle $x, y \in X$ mit $x \neq y$ und alle $\lambda \in (0,1)$ gilt:

$$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda) f(y);$$

(c) **gleichmäßig konvex** (auf X), wenn es ein $\mu > 0$ gibt mit

$$f(\lambda x + (1-\lambda)y) + \mu\lambda(1-\lambda)\left\|x-y\right\|^2 \leq \lambda f(x) + (1-\lambda) f(y).$$

für alle $x, y \in X$ und alle $\lambda \in (0,1)$. (Man bezeichnet f dann auch als **gleichmäßig konvex mit Modulus** μ.)

Satz 4.2
Seien $X \subseteq \mathbb{R}^n$ eine offene und konvexe Menge und $f\colon X \to \mathbb{R}$ stetig differenzierbar. Dann gelten:

(a) f konvex (auf X) $\iff \forall x, y \in X : f(x) - f(y) \geq \nabla f(y)^\top (x-y)$.
(b) f strikt konvex (auf X) $\iff \forall x, y \in X, x \neq y : f(x) - f(y) > \nabla f(y)^\top (x-y)$.
(c) f gleichmäßig konvex (auf X) $\iff \exists \mu > 0 : f(x) - f(y) \geq \nabla f(y)^\top (x-y) + \mu \left\|x-y\right\|^2 \; \forall x, y \in X$

Definition 4.3
Sei $X \subseteq \mathbb{R}^n$ eine gegebene Menge. Eine Funktion $F\colon X \to \mathbb{R}^n$ heißt

(a) **monoton** (auf X), wenn

$$(x-y)^\top (F(x) - F(y)) \geq 0$$

für alle $x, y \in X$ gilt;

(b) **strikt monoton** (auf X), wenn

$$(x-y)^\top (F(x) - F(y)) > 0$$

für alle $x, y \in X$ mit $x \neq y$ gilt;

(c) **gleichmäßig monoton** (auf X), wenn es ein $\mu > 0$ gibt mit

$$(x-y)^\top (F(x) - F(y)) \geq \mu \left\|x-y\right\|^2$$

für alle $x, y \in X$. (Man bezeichnet F dann auch als **gleichmäßig monoton mit Modulus** μ.)

Satz 4.4
Sei $X \subseteq \mathbb{R}^n$ eine offene und konvexe Menge und $f\colon X \to \mathbb{R}$ stetig differenzierbar. Dann gelten:

(a) f konvex $\Longleftrightarrow \nabla f$ monoton
(b) f strikt konvex $\Longleftrightarrow \nabla f$ strikt monoton
(c) f gleichmäßig konvex $\Longleftrightarrow \nabla f$ gleichmäßig monoton.

Satz 4.5 (Mittelwertsatz in Integralform)
Sei $F\colon \mathbb{R}^n \to \mathbb{R}^m$ stetig differenzierbar, sowie $x, y \in \mathbb{R}^n$ gegeben. Dann gilt

$$F(x) = F(y) + \int_0^1 F'(y + \tau(x-y))(x-y) \; \mathrm{d}\tau.$$

Hier nutzen wir $F(x) = \nabla f(x)$, d.h. es gilt

$$\nabla f(x) = \nabla f(y) + \int_0^1 \nabla^2 f(y + \tau(x-y))(x-y) \; \mathrm{d}\tau. \tag{4.1}$$

4.2 Fortsetzung

Wir zeigen zuerst ein Resultat über zweimal stetig differenzierbare (strikt, gleichmäßig) konvexe Funktionen unter Zuhilfenahme von Satz 4.4:

Satz 4.6
Sei $X \subseteq \mathbb{R}^n$ eine offene und konvexe Menge und $f\colon X \to \mathbb{R}^n$ zweimal stetig differenzierbar. Dann gilt

(a) f konvex (auf X) $\Longleftrightarrow \nabla^2 f(x)$ positiv semidefinit für alle $x \in X$
(b) $\nabla^2 f(x)$ positiv definit für alle $x \in X \Longrightarrow f$ strikt konvex (auf X)
(c) f gleichmäßig konvex (auf X) $\Longleftrightarrow \nabla^2 f(x)$ gleichmäßig positiv definit auf X, d.h., wenn es ein $\mu > 0$ gibt mit

$$d^\top \nabla^2 f(x) d \geq \mu \|d\|^2 \tag{4.2}$$

für alle $x \in X$ und für alle $d \in \mathbb{R}^n$.

Beweis. (c) **(⇒)** Sei f gleichmäßig konvex. Wegen Satz 4.4 (c) ist ∇f gleichmäßig monoton, d.h. es existiert ein $\mu > 0$ so, dass für alle $x, y \in X$ gilt

$$(x-y)^\top (\nabla f(x) - \nabla f(y)) \geq \mu \|x-y\|^2 . \tag{4.3}$$

Da ∇f stetig differenzierbar ist folgt mit einer geeigneten Konstanten $\mu > 0$ also

$$\begin{aligned} d^\top \nabla^2 f(x) d &= d^\top \lim_{t\to 0} \frac{\nabla f(x+td) - \nabla f(x)}{t} \\ &= \lim_{t\to 0} \frac{td^\top (\nabla f(x+td) - \nabla f(x))}{t^2} \\ &\overset{(4.3)}{\geq} \lim_{t\to 0} \frac{1}{t^2} \mu \|td\|^2 = \lim_{t\to 0} \mu \|d\|^2 \\ &= \mu \|d\|^2 \end{aligned}$$

für alle $x \in X$ und alle $d \in \mathbb{R}^n$, d.h., $\nabla^2 f(x)$ ist gleichmäßig positiv definit (auf X).

(⇐) Angenommen es gilt (4.2). Aus dem Mittelwertsatz in der Integralform 4.5 und der Monotonie des

Integrals ergibt sich

$$
\begin{aligned}
(x-y)^\top(\nabla f(x)-\nabla f(y)) &\overset{(4.1)}{=} \int_0^1 (x-y)^\top \nabla^2 f(y+\tau(x-y))(x-y)\mathrm{d}\tau \\
&\overset{(4.2)}{\geq} \mu \int_0^1 \|x-y\|^2 \,\mathrm{d}\tau \\
&= \mu \|x-y\|^2 , \qquad (4.4)
\end{aligned}
$$

d.h., nach Definition 4.3 (c) ∇f ist gleichmäßig monoton auf X. Wegen Satz 4.4 (c) ist f selbst daher gleichmäßig konvex auf X.

(a) Folgt aus (c) wenn man $\mu = 0$ setzt.

(b) Sei $\nabla^2 f(z)$ positiv definit für alle $z \in X$. Dann ist $\theta(\tau) := (x-y)^\top \nabla^2 f(y+\tau(x-y))(x-y) > 0$ für alle $\tau \in [0,1]$ und alle $x, y \in X$ mit $x \neq y$. Folglich ist

$$(x-y)^\top(\nabla f(x)-\nabla f(y)) \overset{(4.1)}{=} \int_0^1 \theta(\tau)\mathrm{d}\tau > 0 \qquad (4.5)$$

für alle $x, y \in X$ mit $x \neq y$, vgl. (4.4). Also ist ∇f nach Definition 4.3 (b) strikt monoton und somit f selbst strikt konvex aufgrund von Satz 4.4 (b). □

Die Aussage von Satz 4.6 (b) ist i.A. nur von hinreichendem Charakter, denn die Funktion $f(x) := x^4$ ist strikt konvex, aber $\nabla^2 f(0) = 0$ ist nur positiv semidefinit auf $X = [0,1] \subseteq \mathbb{R}$:

- Aus $f(x) = x^4$ folgt $\nabla f(x) = 4x^3$ und $\nabla^2 f(x) = 12x^2$. Also gilt $\nabla^2 f(0) = 0$ und f kann nicht positiv definit auf X sein.
- $f(x) = x^4$ ist strikt konvex genau dann, wenn $\nabla f(x)$ strikt monoton ist, d.h., wenn für alle $x, y \in X$ mit $x \neq y$ gilt

  $$(x-y)(\nabla f(x) - \nabla f(y)) > 0.$$

  o.B.d.A. können wir $x > y$ voraussetzen. Eingesetzt liefert dies

  $$\underbrace{(x-y)}_{>0}(4x^3 - 4y^3) > 0$$

  was wegen $x^3 > y^3$ erfüllt ist.

Dass Levelmengen von gleichmäßig konvexen Funktionen stets kompakt sind, zeigen wir im folgenden

Lemma 4.7

Sei $f\colon \mathbb{R}^n \to \mathbb{R}$ stetig differenzierbar, $x^0 \in \mathbb{R}^n$ beliebig, die Levelmenge

$$\mathcal{L}(x^0) := \left\{ x \in \mathbb{R}^n \mid f(x) \leq f(x^0) \right\} \qquad (4.6)$$

konvex und f gleichmäßig konvex auf $\mathcal{L}(x^0)$. Dann ist die Menge $\mathcal{L}(x^0)$ kompakt.

Beweis. Die Levelmenge $\mathcal{L}(x^0)$ ist nichtleer, da $x^0 \in \mathcal{L}(x^0)$. Die Funktion f ist gleichmäßig konvex auf $\mathcal{L}(x^0)$, d.h., es existiert ein geeignetes $\mu > 0$ mit $\lambda := \frac{1}{2}$ sodass für alle $x \in \mathcal{L}(x^0)$ und $y = x^0$ gilt (nach Einsetzen in die Definition 4.1 (c))

$$
\begin{aligned}
& f(\tfrac{1}{2}x + \tfrac{1}{2}x^0) + \mu\tfrac{1}{2}\cdot\tfrac{1}{2}\left\|x - x^0\right\|^2 \leq \tfrac{1}{2}f(x) + \tfrac{1}{2}f(x^0) \\
\Leftrightarrow \quad & f(\tfrac{1}{2}(x + x^0)) + \mu\tfrac{1}{4}\left\|x - x^0\right\|^2 \leq \tfrac{1}{2}\left(f(x) + f(x^0)\right)
\end{aligned}
$$

bzw. umgestellt

$$\begin{aligned}
\tfrac{1}{4}\mu \left\|x - x^0\right\|^2 &\leq \tfrac{1}{2}(f(x) + f(x^0)) - f(\tfrac{1}{2}(x + x^0)) \\
&= \tfrac{1}{2} \underbrace{(f(x) - f(x^0))}_{\leq 0,\text{ da } x \in \mathcal{L}(x^0)} - \big(f(\tfrac{1}{2}(x + x^0)) - f(x^0)\big) \\
&\leq -\big(f(\tfrac{1}{2}(x + x^0)) - f(x^0)\big) \\
&\overset{4.2(a)}{\leq} -\tfrac{1}{2}\nabla f(x^0)^\top (x - x^0) \\
&\leq \tfrac{1}{2} \left\|\nabla f(x^0)^\top (x - x^0)\right\| \\
&\leq \tfrac{1}{2} \left\|\nabla f(x^0)\right\| \left\|(x - x^0)\right\|
\end{aligned}$$

und daraus folgt

$$\left\|x - x^0\right\| \leq c := \frac{2\left\|\nabla f(x^0)\right\|}{\mu}$$

für alle $x \in \mathcal{L}(x^0)$. Also ist $\mathcal{L}(x^0)$ beschränkt. Aus Stetigkeitsgründen ist die Levelmenge $\mathcal{L}(x^0)$ aber auch abgeschlossen. Nach dem nach dem Satz von Heine-Borel ist $\mathcal{L}(x^0)$ also kompakt. □

Es ist hierbei egal ob man f als gleichmäßig konvex auf dem gesamten $\mathbb{R}^n$ oder auch nur auf einer konvexen Menge $X \subseteq \mathbb{R}^n$ voraussetzt: Es folgt automatisch, dass die Levelmenge $\mathcal{L}(x^0)$ konvex ist. Man könnte in diesem Fall also auf die explizit geforderte Voraussetzung der Konvexität von $\mathcal{L}(x^0)$ im Lemma 4.7 verzichten.

Vortrag 6

ALLGEMEINES ABSTIEGSVERFAHREN

Gesucht ist ein Verfahren zur Lösung des Problems $\min f(x)$, $x \in \mathbb{R}^n$ mit $f\colon \mathbb{R}^n \to \mathbb{R}$.

Idee: Man ermittelt zu einem Punkt $x \in \mathbb{R}^n$ eine Richtung $d \in \mathbb{R}^n$, in die $f(x)$ absteigt und verkleinert in dieser $f(x)$ hinreichend.

Definition 6.1 (Abstiegsrichtung)

Sei $f\colon \mathbb{R}^n \to \mathbb{R}$ eine stetig differenzierbare Funktion, $x \in \mathbb{R}^n$. Ein Vektor $d \in \mathbb{R}^n$ heißt Abstiegsrichtung von f in x, wenn es ein $\bar{t} > 0$ gibt mit $f(x + td) < f(x)$ für alle $t \in (0, \bar{t})$.

Lemma 6.2

Sei $f\colon \mathbb{R}^n \to \mathbb{R}$ stetig differenzierbar, $x \in \mathbb{R}^n$, $d \in \mathbb{R}^n$ mit $\nabla f(x)^\top d < 0$.
$\Rightarrow$ d ist eine Abstiegsrichtung

Beweis. Da f ist stetig differenzierbar ist, folgt für die Richtungsableitung $f'(x; d)$

$$f'(x; d) = \lim_{t \to 0} \frac{f(x + td) - f(x)}{t} = \nabla f(x) d^\top < 0$$

Somit ist $\frac{f(x+td)-f(x)}{t} < 0$ für hinreichend kleine t. □

Bemerkung 6.3

Lemma 6.2 ist ein hinreichendes aber kein notwendiges Kriterium für eine Abstiegsrichtung. Ist beispielsweise $x \in \mathbb{R}^n$ ein striktes lokales Maximum eine beliebigen Funktion f, so ist nach Definition 6.1 jedes $0 \neq d \in \mathbb{R}^n$ eine Abstiegsrichtung, jedoch ist die Bedingung von 6.2 nicht erfüllt, da $\nabla f(x) = 0$.

Beispiel 6.4

Sei $f\colon \mathbb{R}^n \to \mathbb{R}$ stetig differenzierbar, $x \in \mathbb{R}^n$ mit $\nabla f(x) \neq 0$. Dann ist $d = -\nabla f(x)$ nach Lemma 6.2 eine Abstiegsrichtung. Allgemeiner ist auch $d = -B \cdot \nabla f(x)$ mit einer symmetrischen, positiv definiten Matrix $B \in \mathbb{R}^{n \times n}$ eine Abstiegsrichtung.

Algorithmus 6.5

Schritt 1. Wähle $x^0 \in \mathbb{R}^n$ und setze $k := 0$.
Schritt 2. Wenn x^k einem Abbruchkriterium genügt: Stop.
Schritt 3. Bestimme d^k von f in x.
Schritt 4. Bestimme $t^k > 0$ mit $f(x^k + t^k d^k) < f(x^k)$.
Schritt 5. Setze $x^{k+1} = x^k + t^k d^k$, $k \to k + 1$, gehe zu Schritt 1.

Implizit nehmen wir im Folgenden an, dass dieser Algorithmus eine unendliche Folge $\left\{x^k\right\}_{k \in \mathbb{N}}$ generiert.

Definition 6.6

Sei $f\colon \mathbb{R}^n \to \mathbb{R}$ stetig differenzierbar, $x \in \mathbb{R}^n$, $d \in \mathbb{R}^n$ eine Abstiegsrichtung von f in x.

- Eine Abbildung $T\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathcal{P}(\mathbb{R}_+)$ heißt **Schrittweitenstrategie**. Diese heißt wohldefiniert, wenn für $(x, d) \in \mathbb{R}^n \times \mathbb{R}^n$ mit $\nabla f(x)^\top d < 0$ gilt, dass $T(x, d) \neq \emptyset$.
- T heißt **effizient**, falls es eine von x und d unabhängige Konstante $\theta > 0$ gibt mit

$$f(x + +td) \leq f(x) - \theta \left(\frac{\nabla f(x)^\top d}{\|d\|} \right) \qquad \text{für alle } t \in T(x, d)$$

Eine Schrittweite t heißt effizient, wenn sie mit einer effizienten Schrittweitenstrategie erzeugt wurde.

Satz 6.7

Sei $f\colon \mathbb{R}^n \to \mathbb{R}$ stetig differenzierbar und $\{x^k\}_{k\in\mathbb{N}}$ eine Folge, die mit Algorithmus 6.5 erzeugt wurde. Außerdem gelte:

1. Es existiert eine Konstante $c > 0$ mit
$$-\frac{\nabla f(x^k)^\top d^k}{\|\nabla f(x^k)\| \cdot \|d^k\|} \geq c \text{für alle} f k \in \mathbb{N} \tag{6.0}$$
2. Die Schrittweiten t^k seien effizient für alle $k \in \mathbb{N}$.

Dann ist jeder Häufungspunkt der Folge $\{x^k\}_{k\in\mathbb{N}}$ ein stationärer Punkt von f.

Beweis. Da alle t^k effizient sind, folgt die Existenz eines $\theta > 0$ mit

$$f(x^{k+1}) = f(x^k + t^k d^k) \leq f(x^k) - \theta \left(\frac{\nabla f(x^k)^\top d^k}{\|d^k\|} \right)^2 \quad \text{für alle } k \in \mathbb{N}$$

Mit der Winkelbedingung folgt nun

$$f(x^{k+1}) \leq f(x^k) - \theta c^2 \left\|\nabla f(x^k)\right\|^2$$

Sei x^* Häufungspunkt von $\{x^k\}_{k\in\mathbb{N}}$. Es ist klar, das $\{f(x^k)\}_{k\in\mathbb{N}}$ monoton fallend ist und zumindest eine Teilfolge gegen $f(x^*)$ konvergiert. Damit konvergiert dann auch $f(x^k) \to f(x^*)$. Insbesondere gilt $f(x^{k+1}) - f(x^k) \to 0$ und $\left\|\nabla f(x^k)\right\| \to 0$. Somit ist jeder Häufungspunkt von $\{x^k\}_{k\in\mathbb{N}}$ ein stationärer Punkt von f. □

Beispiel 6.8

Sei $f\colon \mathbb{R}^n \to \mathbb{R}$ eine quadratische Funktion, d.h. $f(x) := 1/2 x^\top Qx + c^\top x + \gamma$ mit $Q \in \mathbb{R}^{n\times n}$ symmetrisch und positiv definit. Seien $x \in \mathbb{R}^n$ und $d \in \mathbb{R}^n$ eine Abstiegsrichtung, die beliebig gegeben sind. Dann liefert

$$t_{\min} = -\frac{\nabla f(x)^\top d}{d^\top Q d}$$

den stärksten Abstieg.

Beweis. Sei $t \in \mathbb{R}$. Es ist $f(x+td) = f(x) + t\nabla f(x)^\top d + 1/2 t^2 \cdot d^\top Qd$. Definieren wir $\varphi(t) := f(x+td)$, dann ist $\varphi'(t_{\min}) = 0$. Daraus folgt nun

$$0 = \varphi'(t_{\min}) = \nabla f(x)^\top d + t_{\min} d^\top Q d \qquad \Rightarrow \qquad t_{\min} = -\frac{\nabla f(x)^\top d}{d^\top Q d}$$

Nun kann man noch zeigen, dass $t_{\min}$ effizient ist:

$$\begin{aligned} f(x + t_{\min} d) &= f(x) + t_{\min} \nabla f(x)^\top d + \frac{1}{2} t_{\min}^2 d^\top Q d \\ &= f(x) - \frac{1}{2} t_{\min}^2 d^\top Q d \\ &= f(x) - \frac{\left(\nabla f(x)^\top d\right)^2}{2 d^\top Q d} \end{aligned}$$

Damit gilt offensichtlich, dass ein θ exisitert mit $\frac{1}{2d^\top Qd} \geq \frac{\theta}{\|d\|^2}$ □

Vortrag 7

FORTSETZUNG ABSTIEGSVERFAHREN

Satz 7.1

Seien $x_0 \in \mathbb{R}^n$, $f\colon \mathbb{R}^n \to \mathbb{R}$ stetig differenzierbar, die Levelmenge $Ł(x^0) := \{x \in \mathbb{R}^n : f(x) \leq f(x^0)\}$ konvex und f gleichmäßig konvex auf $Ł(x^0)$. Sei $\{x^k\}$ eine durch den Algorithmus 6.5 (Abstiegsverfahren) erzeugte Folge, sodass

(1) Es ist $\sum_{k=0}^{\infty} \delta_k = \infty$ wobei

$$\delta_k := \left(\frac{\nabla f(x^k)^\top d^k}{\|\nabla f(x^k)\| \, \|d^k\|} \right)^2 \qquad \text{(Zoutendijk-Bedingung)}$$

(2) Die Schrittweiten $t_k > 0$ sind effizient für alle $k \in \mathbb{N}$.

Dann konvergiert die Folge $\{x^k\}$ gegen das eindeutig bestimmte globale Minimum von f.

Beweis. Sei $k \in \mathbb{N}$. Wegen $x^0 \in \mathcal{L}(x^0)$ gilt $\mathcal{L}(x^0) \neq \emptyset$ und mit Lemma 4.7 folgt, dass $\mathcal{L}(x^0)$ kompakt. Da jedes globale Minimum von f azf dem $\mathbb{R}^n$ notwendig in der Levelmenge $\mathcal{L}(x^0)$ liegen muss, besitzt f nach Theorem 5.4 genau ein globales Minimum x^*. Da f gleichmäßig konvex auf $\mathcal{L}(x^0)$ ist, existiert ein $\mu > 0$, sodass für alle

$$\forall x, y \in \mathcal{L}(x^0) : f(x) - f(y) \geq \nabla f(y)^\top (x - y) + \mu \|x - y\|^2 \tag{7.1}$$

Aus der trivialen Ungleichung

$$0 \leq \left\| \sqrt{\frac{\mu}{2}}(x^* - x^k) + \sqrt{\frac{1}{2\mu}} \nabla f(x^k) \right\|^2$$

folgt nach kurzer Rechnung

$$-\frac{1}{2\mu} \left\|\nabla f(x^k)\right\|^2 \leq \frac{\mu}{2} \left\|x^* - x^k\right\|^2 + \nabla f(x^k)^\top (x^* - x^k) \leq f(x^*) - f(x^k)$$

Daraus folgt nun

$$-\left\|\nabla f(x^k)\right\|^2 \leq 2\mu(f(x^*) - f(x^k)) \tag{7.2}$$

Aus der Effizienz der Schrittweiten t_k existiert ein $\theta > 0$ mit

$$\begin{aligned} f(x^{k+1}) &= f(x^k + t_k d^k) \\ &\leq f(x^k) - \theta \left(\frac{\nabla f(x^k) d^k}{\|d^k\|} \right)^2 \cdot \frac{\left\|\nabla f(x^k)\right\|}{\nabla f(x^k)} \\ &= f(x^k) - \theta \delta_k \cdot \left\|\nabla f(x^k)\right\|^2 \\ &\overset{(7.2)}{\leq} f(x^k) - 2\mu\theta(f(^*) - f(x^k)) \end{aligned} \tag{7.3}$$

Also ist

$$0 \leq f(x^{k+1}) - f(x^k) \overset{(7.2)}{\leq} f(x^k) - 2\delta_k \theta(f(x^k) - f(x^*)) - f(x^*) = (1 - 2\mu\theta\delta_k)(f(x^k) - f(x^*))$$

Durch $k+1$-fache Anwendung dieser Ungleichung sowie unter AUsnutzung der bekannten Ungleichung $\exp(x) \geq$

$1 + x$ für alle $x \in \mathbb{R}$. Für $x = -2\mu\theta\delta_k$ ergibt sich

$$\begin{aligned}
0 &\leq f(x^{k+1}) - f(x^*) \\
&\leq \prod_{j=0}^{k} (1 - 2\mu\theta\delta_j)(f(x^0) - f(x^*)) \\
&\leq \prod_{j=0}^{k} \exp(-2\mu\theta\delta_j)(f(x^0) - f(x^*)) \\
&= \exp\left(-2\mu\theta \sum_{j=0}^{k} \delta_j\right)(f(x^0) - f(x^*))(6)
\end{aligned} \tag{7.4}$$

Wegen $\sum_{j=0}^{k} \delta_j \overset{k\to\infty}{\longrightarrow} \infty$ ergibt sich hieraus die Konvergenz von $\left\{f(x^k)\right\}$ gegen $f(x^*)$. Aus Lemma 5.6 folgt $0 \leq \mu \left\|x^k - x^*\right\|^1 \leq f(x^k) - f(x^*)$ für alle $k \in \mathbb{N}$ und damit folgt auch die Konvergenz $x^k \to x^*$. $\square$

Bemerkung 7.2

Die Winkelbedingung aus Satz 6.7 ist hinreichend für die Zoutendijk-Bedingung, denn

$$\exists c > 0 \;\; \forall k \in \mathbb{N}\colon -\frac{\nabla f(x^k) d^k}{\|\nabla f(x^k)\| \cdot \|d^k\|} \geq c \quad \Rightarrow \quad \forall k \in \mathbb{N}\colon \delta_k \geq c^2 \geq 0 \quad \Rightarrow \sum_{j=0}^{\infty} \delta_j \geq \sum_{j=0}^{\infty} c^2 = \infty$$

Bemerkung 7.3

Die beiden Konvergenzsätze Satz 6.7 und 7.1 sind von fundamentaler Bedeutung zum Nachweis der globalen Konvergenz verschiedener Abstiegsverfahren.

Folgerung 7.4

Seien $x_0 \in \mathbb{R}^n$, $f\colon \mathbb{R}^n \to \mathbb{R}$ stetig differenzierbar, die Levelmenge $\mathcal{L}(x^0) := \left\{x \in \mathbb{R}^n\colon f(x) \leq f(x^0)\right\}$ konvex und f gleichmäßig konvex auf $\mathcal{L}(x^0)$. Sei $\left\{x^k\right\}$ eine durch den Algorithmus 6.5 (Abstiegsverfahren) erzeugte Folge, sodass

(1) $\exists \delta > 0 \;\; \forall k \in \mathbb{N}\colon \delta_k \geq \delta$ wobei
$$\delta_k := \left(\frac{\nabla f(x^k)^\top d^k}{\|\nabla f(x^k)\| \, \|d^k\|}\right)^2$$

(2) Die Schrittweiten $t_k > 0$ sind effizient für alle $k \in \mathbb{N}$.

Dann konvergiert die Folge $\left\{x^k\right\}$ gegen das eindeutig bestimmte globale Minimum x^* von f und es existieren Konstanten $c \geq 0$ und $q \in (0, 1)$ mit $\left\|x^k - x^*\right\| \leq cq^k$ für alle $k \in \mathbb{N}$ (d.h. $\left\{x^k\right\}$ konvergiert R-linear gegen x^*).

Beweis.

$$\forall k \in \mathbb{N}\colon \delta_k \geq \delta \;\; \Rightarrow \;\; \forall k \in \mathbb{N}\colon \sum_{j=0}^{k} \delta_j \geq \delta(k+1) \;\; \Rightarrow \sum_{j=0}^{k} \delta_j \overset{k\to\infty}{\longrightarrow} \infty$$

Mit Satz 7.1 folgt die Konvergenz von $\left\{x^k\right\}$ gegen x^*. Weiterhin gilt für alle $k \in \mathbb{N}$ mit Lemma 5.6

$$\begin{aligned}
\mu \left\|x^k - x^*\right\|^2 &\leq f(x^k) - f(x^*) \\
&\leq \exp\left(-2\mu\theta \sum_{j=0}^{k-1} \delta_j\right)(f(x^0) - f(x^*)) \\
&\leq \exp\left(-2\mu\theta\delta \cdot k\right)(f(x^0) - f(x^*))
\end{aligned}$$

$$\Rightarrow \quad \left\|x^k - x^*\right\| \leq \sqrt{\frac{(f(x^0) - f(x^*))}{\mu} \cdot \exp(-2\mu\theta\delta k)} = \underbrace{\sqrt{\frac{(f(x^0) - f(x^*))}{\mu}}}_{=:c} \cdot \underbrace{\exp(-\mu\theta\delta)}_{=:q}^{k}$$

Vortrag 8
SCHRITTWEITENSTRATEGIEN

Lars Ortscheidt

Das bisher kennengelernte Abstiegsverfahren hat in Bezug auf die Abstiegsrichtung d und die Schrittweite t große Freiheitsgrade. Da die Schrittweite $t_{\min}$ mit $f(x+t_{\min}d) = \min_{t\geq 0} f(x+td)$ im Allgemeinen nicht in endlich vielen Schritten berechnet werden kann, werden nun zwei Schrittweitenregeln vorgestellt, welche "realisierbar" sind, d.h. sie brauchen nur endlich viele Schritte zur Berechnung.

8.1 Armijo-Regel

Definition 8.1

Seien $f : \mathbb{R}^n \to \mathbb{R}$ stetig differenzierbar und $\sigma, \beta \in (0,1)$ fest für das ganze Abstiegsverfahren. Zu $x, d \in \mathbb{R}^n$ mit $\nabla f(x)^\top d < 0$ bestimme $t := \max_{\ell \in \mathbb{N}_0} \beta^\ell$, so dass gilt

$$f(x+td) \leq f(x) + \sigma t \nabla f(x)^\top d. \tag{8.1}$$

Dieses Verfahren heißt **Armijo-Regel**.

Setzt man $\varphi(t) := f(x+td)$ lautet (8.1)

$$\varphi(t) \leq \varphi(0) + \sigma t \varphi'(0),$$

die sogenannte Armijo-Goldstein-Bedingung.

Bemerkung 8.2

- Zur Bestimmung von t wird (8.1) also für $t = \beta^\ell$, $\ell = 0, 1, 2, \ldots$ überprüft und bei der ersten Gültigkeit abgebrochen.
- $T(x,d)$ besitzt bei der Armijo-Regel höchstens ein Element.

Satz 8.3

Seien $f : \mathbb{R}^n \to \mathbb{R}$ stetig differenzierbar mit $\sigma, \beta \in (0,1)$, fest. Dann existiert für alle $x, d \in \mathbb{R}^n$ mit $\nabla f(x)^\top d < 0$ ein endliches $\ell \in \mathbb{N}$ mit

$$f(x+\beta^\ell d) \leq f(x) + \sigma \beta^\ell \nabla f(x)^\top d,$$

d.h. die Armijo-Regel ist wohldefinitionniert.

Beweis. Angenommen für alle $\ell \in \mathbb{N}$ gilt

$$f(x+\beta^\ell d) > f(x) + \sigma \beta^\ell \nabla f(x)^\top d$$

und somit auch

$$\frac{f(x+\beta^\ell d) - f(x)}{\beta^\ell} > \sigma \nabla f(x)^\top d.$$

Dann folgt mit $\ell \to \infty$ wegen der Differenzierbarkeit von f

$$\nabla f(x)^\top d > \sigma \nabla f(x)^\top d$$

und weil $\sigma \in (0,1)$ ergibt sich

$$\nabla f(x)^\top d \geq 0,$$

was im Widerspruch zur Voraussetzung des Satzes steht. □

8.2 Wolfe-Powell-Schrittweitenstrategie

Definition 8.4

Seien $f : \mathbb{R}^n \to \mathbb{R}$ stetig differenzierbar und $\sigma \in (0, \frac{1}{2})$, $\rho \in [\sigma, 1)$ fest vorgegeben. Für $x, d \in \mathbb{R}^n$ mit $\nabla f(x)^\top d < 0$ bestimme man ein $t > 0$ mit

$$f(x + td) \leq f(x) + \sigma t \nabla f(x)^\top d \tag{8.2}$$

und

$$\nabla f(x + td)^\top d \geq \rho \nabla f(x)^\top d. \tag{8.3}$$

(8.2) und (8.3) heißen dann die **Wolfe-Powell-Bedingungen**.
Setzt man $\varphi(t) := f(x + td)$, so lauten (8.2) und (8.3)

$$\varphi(t) \leq \varphi(0) + \sigma t \varphi'(0)$$

und

$$\varphi'(t) \geq \rho \varphi'(0).$$

Satz 8.5

Seien $f : \mathbb{R}^n \to \mathbb{R}$ stetig differenzierbar und $\sigma \in (0, \frac{1}{2})$, $\rho \in [\sigma, 1)$, $x^0 \in \mathbb{R}^n$ fest. Zu $x \in \mathcal{L}(x^0) := \{z \in \mathbb{R}^n | f(z) \leq f(x^0)\}$ und $d \in \mathbb{R}^n$ mit $\nabla f(x)^\top d < 0$ sei

$$T_{WP}(x, d) := \{t > 0 | (8.2) \text{ und } (8.3) \text{ gelten}\}$$

die Menge der Wolfe-Powell-Schrittweiten in x in Richtung d. Dann gelten:

(a) Ist f nach unten beschränkt, so ist $T_{WP}(x, d) \neq \emptyset$, d.h. die Wolfe-Powell-Strategie ist wohldefinitionniert.

(b) Ist außerdem der Gradient ∇f auf der Levelmenge $\mathcal{L}(x^0)$ Lipschitz-stetig, so existiert eine Konstante $\theta > 0$ (unabhängig von x und d) mit

$$f(x + td) \leq f(x) - \theta \left(\frac{\nabla f(x)^\top d}{\|d\|} \right)^2$$

für alle $t \in T_{WP}(x, d)$, d.h. die Wolfe-Powell-Schrittweitenstrategie ist effizient.

Beweis. Zu (a): Setze

$$\begin{aligned} \varphi(t) &:= f(x + td), \\ \psi(t) &:= f(x + td) - f(x) - \sigma t \nabla f(x)^\top d = \varphi(t) - \varphi(0) - \sigma t \varphi'(0) \end{aligned}$$

und somit

$$\begin{aligned} \varphi'(t) &= \nabla f(x + td)^\top d, \\ \psi'(t) &= \varphi'(t) - \sigma \varphi'(0). \end{aligned}$$

So lauten (8.2) und (8.3)

$$\psi(t) \leq 0 \tag{8.4}$$

$$\varphi'(t) \geq \rho \varphi'(0). \tag{8.5}$$

Offenbar gilt $\psi(0) = 0, \lim_{t \to \infty} \psi(t) = \infty$ (da f beschränkt), $\psi'(0) < 0$, damit $\psi'(t) < 0$ für $t \in [0, t_0), t_0 > 0$. Daher existiert $t^* > 0$ minimal mit $\psi(t^*) = 0, \psi(t) < 0$ für $t \in (0, t^*)$, somit erfüllt t^* (8.4). Da $\psi'(t^*) \geq 0$ folgt $\varphi'(t^*) \geq \sigma \varphi'(0)$, womit wegen $\rho \geq \sigma > 0$ und $\varphi'(0) < 0$ (8.5) für t^* folgt, also $t^* \in T_{WP}(x, d)$.

Zu (b): Sei $t \in T_{WP}(x,d)$ gegeben. Dann ist $f(x+td) \le f(x)$ und somit insbesondere $x + td \in \mathcal{L}(x^0)$. Aus der Wolfe-Powell-Regel folgt zunächst

$$\begin{aligned} & \rho \nabla f(x)^\top d - \nabla f(x)^\top d \le \nabla f(x+td)^\top d - \nabla f(x)^\top d \\ \iff \quad & (\rho - 1)\nabla f(x)^\top d \le (\nabla f(x+td) - \nabla f(x))^\top d. \end{aligned}$$

Mit der Cauchy-Schwarzschen Ungleichung und vorausgesetzten Lipschitz-Stetigkeit von $\nabla f(x)$ auf $\mathcal{L}(x^0)$ folgt mit einer geeigneten Konstanten $\ell > 0$:

$$(\rho - 1)\nabla f(x)^\top d \le \|\nabla f(x+td) - \nabla f(x)\| \, \|d\| \le Lt\|d\|^2$$

Hieraus folgt

$$t \ge \frac{(\rho-1)\nabla f(x)^\top d}{\ell \|d\|^2}$$

und daher

$$\begin{aligned} f(x+td) &\overset{(8.2)}{\le} f(x) + \sigma t \nabla f(x)^\top d \\ &\le f(x) + \sigma \frac{(\rho-1)\nabla f(x)^\top d}{\ell\|d\|^2} \nabla f(x)^\top d \\ &\le f(x) - \theta \left(\frac{\nabla f(x)^\top d}{\|d\|} \right)^2 \end{aligned}$$

mit

$$\theta := \frac{(1-\rho)\sigma}{\ell}$$

Damit ist die Behauptung bewiesen. □

Abschließend zwei hinreichende Bedingungen dafür, dass der Gradient ∇f auf der Levelmenge $\mathcal{L}(x^0)$ Lipschitz-stetig ist.

Lemma 8.6

Seien $f : \mathbb{R}^n \to \mathbb{R}$ zweimal stetig differenzierbar und $x^0 \in \mathbb{R}^n$. Ist eine der folgenden Bedingungen erfüllt:

(a) $\|\nabla^2 f(x)\|$ ist beschränkt auf einer konvexen Obermenge X der Levelmenge $\mathcal{L}(x^0)$,

(b) die Levelmenge $\mathcal{L}(x^0)$ ist kompakt,

so ist der Gradient ∇f Lipschitz-stetig auf $\mathcal{L}(x^0)$.

Beweis. Sei Bedingung (b) erfüllt. Definiere

$$[\mathcal{L}(x^0)] := \{ax + (1-a)y \mid a \in [0,1], x, y \in \mathcal{L}(x^0)\},$$

offenbar ist diese Menge eine konvexe Obermenge von $\mathcal{L}(x^0)$. Betrachte die Funktion

$$f \colon \begin{cases} [0,1] \times \mathcal{L}(x^0) \times \mathcal{L}(x^0) & \to \quad [\mathcal{L}(x^0)] \\ (a, x, y) & \mapsto \quad ax + (1-a)y \end{cases}$$

Offenbar ist f stetig und surjektiv mit kompaktem Definitionsbereich. Da das Bild von kompakten Mengen unter stetigen Funktionen wieder kompakt ist, folgt $[\mathcal{L}(x^0)]$ kompakt, d.h. es es existiert eine konvexe und kompakte Obermenge $X \subseteq \mathbb{R}^n$ mit $\mathcal{L}(x^0) \subseteq X$. Aus Stetigkeitsgründen existiert dann eine Konstante L mit

$$\|\nabla^2 f(x)\| \le L \qquad \text{für alle } x \in X, \tag{8.6}$$

d.h. die Bedingung (a) ist erfüllt.

Sei nun die Bedingung (a) erfüllt. Dann existiert eine Zahl $L > 0$ mit (8.6) . Es gilt

$$\int_0^1 \nabla^2 f(y + \tau(x-y))(x-y)\, d\tau = [\nabla f(y + \tau(x-y))]_{\tau=0}^1 = \nabla f(x) - \nabla f(y)$$

für alle $x, y \in X$. Wegen $y + \tau(x-y) \in X$ ist daher

$$\begin{aligned}\|\nabla f(x) - \nabla f(y)\| &\leq \int_0^1 \|\nabla^2 f(y + \tau(x-y))\| \, d\tau \, \|x-y\| \\ &\leq \int_0^1 L \, d\tau \, \|x-y\| \\ &= L\|x-y\|\end{aligned}$$

für alle $x, y \in X$, was zu zeigen war. □

Vortrag 9

SCHRITTWEITENALGORITHMEN

Michael Kunert

Erinnerung 9.1

Sei $f\colon \mathbb{R}^n \to \mathbb{R}$ stetig differenzierbar und $\sigma \in (0, \frac{1}{2})$ sowie $\rho \in [0, 1)$ gegeben. Zu $x, d \in \mathbb{R}^n$ mit $\nabla f(x)^\top d < 0$ bestimme man ein $t > 0$ mit

$$f(x + td) \leq f(x) + \sigma t \nabla f(x)^\top d \tag{9.1}$$

$$\nabla f(x + td)^\top d \geq \rho \nabla f(x)^\top d \tag{9.2}$$

Zur Vereinfachung im Algorithmus setzen wir im Folgenden stets

$$\varphi(t) := f(x + td)$$
$$\psi(t) := \varphi(t) - \varphi(0) - \sigma t \varphi'(0)$$

Bemerkung 9.2

Die Wolfe-Powellbedingungen (9.1) und (9.2) sind damit

$$\psi(t) \leq 0 \quad \text{und} \quad \varphi'(t) \geq \rho \varphi'(0)$$

Beweis. Zum einen ist

$$\begin{aligned} & \psi(t) = \varphi(t) - \varphi(0) - \sigma t \varphi'(0) \leq 0 \\ \Leftrightarrow\ & f(x + td) - f(x) - \sigma t \nabla f(x)^\top d \leq 0 \\ \Leftrightarrow\ & f(x + td) \leq f(x) + \sigma t \nabla f(x)^\top d \end{aligned}$$

und außerdem

$$\varphi'(t) \geq \rho \varphi'(0) \quad \Leftrightarrow \quad \nabla f(x + td)^\top d \geq \rho \nabla f(x)^\top d$$

Lemma 9.3

Seien $\sigma < \rho$ und $\varphi'(0) < 0$. Ist $[a, b]$ mit $0 \leq a \leq b$ ein Intervall mit den Eigenschaften $\psi(a) \leq 0$, $\psi(b) \geq 0$ und $\psi'(a) < 0$, so enthält das Intervall $[a, b]$ einen Punkt $\bar{t}$ mit $\psi(\bar{t}) < 0$ und $\psi'(\bar{t}) = 0$. $\bar{t}$ ist ein innerer Punkt eines Intervalls I, sodass für alle $t \in I$ gilt:

$$\psi(t) \leq 0 \quad \text{und} \quad \varphi'(t) \geq \rho \varphi'(0) \tag{9.3}$$

Beweis. Sei $\bar{t}$ ein gloables Minimum von ψ auf $[a, b]$ (Satz von Weierstraß). Wegen (9.3) ist $\bar{t}$ ein innerer Punkt von $[a, b]$. Außerdem ist $\psi(\bar{t})$ muss gelten. Wegen $\sigma < \rho$ folgt die Existenz von I, sodass für alle $t \in I$ gilt

$$\left.\begin{array}{rcl} \psi(t) & \leq & 0 \quad \text{und} \quad \psi'(t) \geq (\rho - \sigma)\varphi'(0) \\ \psi'(t) & = & \varphi'(t) - \sigma \varphi'(0) \end{array}\right\} \Rightarrow \ \psi(t) \leq 0 \quad \text{und} \quad \varphi'(t) \geq \rho \varphi'(0) \ \forall\, t \in I$$

Algorithmus 9.4
Gegeben seien $x \in \mathbb{R}^n$ und $d \in \mathbb{R}^n$ mit $\nabla f(x)^\top d < 0$.

Phase A.

A.0 Wähle $t_0 > 0$, $\gamma > 1$ und setze $i := 0$.
A.1 Ist $\psi(t_i) \geq 0$, so setze $a := 0$ und $b := t$ und gehe zu (B.0).
Ist $\psi(t_i) < 0$, $\psi'(t_i) \geq \rho\varphi'(0)$, so setze $t := t_i$ und breche ab. `STOP 1`.
Ist $\psi(t_i) < 0$, $\psi'(t_i) < \rho\varphi'(0)$, so setze $t_{i+1} := \gamma t$, $i = i + 1$ und gehe zu (A.1)

Phase B.

B.0 Wähle $\tau_1, \tau_2 \in (0, \frac{1}{2}]$, setze $j := 0$ und setze $a_0 := a$ sowie $b_0 := b$.
B.1 Wähle $t_j \in [a_j + \tau_1(b_j - a_j), b_j + \tau_2(b_j - a_j)]$.
B.2 Ist $\psi(t_j) \geq 0$, so setze $a_{j+1} = a_j$, $b_{j+1} = t_j$, $j = j + 1$ und gehe zu (B.1).
Ist $\psi(t_j) < 0$, $\psi'(t_i) \geq \rho\varphi'(0)$, so setze $t := t_j$ und breche ab. `STOP 2`.
Ist $\psi(t_i) < 0$, $\psi'(t_i) < \rho\varphi'(0)$, so setze $a_{j+1} := t_j$, $b_{j+1} := b_j$, $j = j + 1$ und gehe zu (B.1).

Wenn f nach unten beschränkt ist und eine Schranke $\underline{f}$ bekannt ist, dann gilt

$$\psi(t) \leq 0 \quad \Leftrightarrow \quad \sigma t\varphi'(0) \leq \varphi(t) - \varphi(0) \quad \Rightarrow \quad t \leq \frac{\underline{f} - \varphi(0)}{\sigma\varphi'(0)}$$

also ist $t_0 \in (0, t)$.

Satz 9.5
Sei $f\colon \mathbb{R}^n \to \mathbb{R}$ stetig differenzierbar und nach unten beschränkt. Des Weiteren seien $\sigma \in (0, \frac{1}{2})$ und $\rho \in (\sigma, 1)$. Dann bricht Algorithmus 9.4 nach endlich vielen Schritten ab.

Beweis. Phase A: Bricht Algorithmus 9.4 bei `STOP 1` ab, dann sind die Wolfe-Powell-Bedingungen erfüllt. Bei der Übergabe nach (B.0) hat $[a, b]$ offenbar die Eigenschaften (9.3) und $\varphi'(a) < \sigma\varphi'(0)$. Angenommen Phase A würde nicht abbrechen. Dann ist für $t_i = \gamma^i \cdot t_0$ aufgrund der Fallvoraussetzung $\psi(t_i) < 0$, also $\varphi(t_i) < \varphi(0) + \sigma t_i\varphi'(0)$. Das ist wegen $\gamma > 1$, $\varphi'(0) < 0$ und der Beschränkung von f nicht möglich.

Phase B: Wenn Phase B bei `STOP 2` abbricht, sind die Wolfe-Powell-Bedingungen erfüllt. Zeige nun, dass für alle Intervalle $[a_j, b_j]$ die Gleichung (9.3) erfüllt und $\varphi'(a_j) < \rho\varphi'(0)$ mittels Induktion. Für $j = 0$ ist die Aussage erfüllt. Habe nun $[a_j, b_j]$ die geforderten Eigenschaften. Falls $\varphi(t_j) \geq 0$ gilt $\psi(a_{j+1}) \leq 0$, $\psi(b_{j+1}) \geq 0$ und $\psi'(a_{j+1}) < 0$. Falls $\psi(t_j) < 0$, dann ist $\psi(a_{j+1}) \leq 0$, $\psi(b_{j+1}) \geq 0$ und $\psi'(a_{j+1}) = \psi'(t_j) < \rho\varphi'(0) < 0$. In beiden Fällen sind die Eigenschaften also erfüllt.
Es bleibt zu zeigen, dass auch auch Phase B abbricht. Die Intervalllängen $|b_j - a_j|$ ziehen sich auf einen Punkt t^* zusammen. Nach Lemma 9.2 gibt es jedem $[a_j, b_j]$ ein $t_j \in (a_j, b_j)$ mit $\psi(t_j) < 0$ und $\psi'(t_j) = 0$. Wegen $t_j \to t$ für $j \to \infty$ folgt $\psi'(t^*) = 0$, also $\varphi(t^*) = \sigma\varphi'(0)$. Das steht jedoch im Widerspruch zu $\varphi'(a_j) < \rho\varphi'(0)$ und daraus resultierend $\varphi'(t^*) \leq \rho\varphi'(0)$. $\square$