TECHNISCHE UNIVERSITÄT DRESDEN

Fakultät Mathematik Institut für Numerik, Professur für Numerik der 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`

Vortrag 2

KONVEXE FUNKTIONEN

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 $i \in \{1,\ldots,n\}$ 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

$$\begin{aligned} g(\lambda x + (1-\lambda)y) &= g(0.5x) = 0.5^4 \cdot x^4 = 1/16 x^4 \\ \lambda g(x) + (1-\lambda)g(y) &= 0.5g(x) = 0.5x^4 \end{aligned}$$

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

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

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:

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

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

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

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) $\square$