

Globale Optimierung

Prof. Dr. Oliver Stein

Sommersemester 2017

Karlsruher Institut für Technologie

Kapitel 1

Einführungen

1.1 Grundbegriffe

Zwei grundlegende Begriffe in der Optimierung sind der **optimale Punkt** und der **optimale Wert**. Als Beispiel kann man fragen, wer in einer Gruppe von Personen die meisten Münzen bei sich trägt. Der optimale Wert ist die gefundene größtmögliche Anzahl von Münzen. Er ist eindeutig bestimmt. Im Gegensatz dazu können durchaus mehrere Personen diese Anzahl von Münzen bei sich tragen. Demnach ist der Punkt (hier: die Person), an dem der optimale Wert angenommen wird, nicht notwendigerweise eindeutig. Dies gilt analog bei jedem Optimierungsproblem.

Beispiel (Projektion auf eine Menge).

Sei $M \subseteq \mathbb{R}^n$ und $z \in \mathbb{R}^n$

$$P_M : \min_{x \in \mathbb{R}^n} \|x - z\|_2 \text{ s.t. } x \in M.$$

Nach Satz 1.2.10 ist dieses Projektionsproblem für jede nicht-leere und kompakte Menge $M \subseteq \mathbb{R}^n$ lösbar, denn $\|\cdot\|_2$ ist eine stetige Funktion.

Beachte: Projektion unter euklidischer Norm = orthogonale Projektion

Falls M eine Hyperebene ist, lässt sich das Problem P_M explizit lösen. Dazu sei

$$M = \{x \in \mathbb{R}^n : a^T x = b\},$$

d.h. M sei durch eine lineare Gleichung beschrieben. Dann wird das Optimierungsproblem P zu

$$P_M : \min_{x \in \mathbb{R}^n} \|x - z\|_2 \text{ s.t. } a^T x = b$$

damit ist der eindeutige optimale Punkt \bar{x} und der dazugehörige optimale Wert v gegeben durch

$$\bar{x} = z - \frac{a^T z - b}{a^T a} \cdot a, \quad v = \left| \frac{a^T z - b}{a^T a} \cdot a \right|_2 = \frac{|a^T z - b|}{\|a\|_2}$$

Definition (Minimalpunkte und Minimalwerte).

Gegeben sei eine Menge von zulässigen Punkte $M \subseteq \mathbb{R}^n$ und eine Zielfunktion $f: M \rightarrow \mathbb{R}$.

a) $\bar{x} \in M$ heißt **lokaler Minimalpunkt** von f auf M , falls eine Umgebung U von \bar{x} existiert mit

$$\forall x \in U \cap M : \quad f(x) \geq f(\bar{x}).$$

b) $\bar{x} \in M$ heißt **globaler Minimalpunkt** von f auf M , falls man in a) $U = \mathbb{R}^n$ wählen kann.

c) Ein lokaler oder globaler Minimalpunkt heißt **strikt**, falls in a) bzw. b) für $x \neq \bar{x}$ sogar die strikte Ungleichung „ $>$ “ gilt.

d) Zu jedem globalen Minimalpunkt \bar{x} heißt $f(\bar{x})$ ($= v = \min_{x \in M} f(x)$) **globaler**, und zu jedem lokalen Minimalpunkt \bar{x} heißt $f(\bar{x})$ **lokaler Minimalwert**.

Strikte globale Minimalpunkte sind eindeutig, und strikte lokale Minimalpunkte sind lokal eindeutig.

Beispiel (1.1.5, 1.2.26, 1.2.31 Zentrum einer Punktwolke).

Gegeben seien Punkte $x^1, x^2, \dots, x^m \in \mathbb{R}^n$. Gesucht ist ein Punkt $z \in \mathbb{R}^n$ „im Zentrum der Punkte“. Benutzt man die 2-Norm so führt dies auf das Optimierungsproblem

$$P : \quad \min_{z \in \mathbb{R}^n} \left\| \left(\|z - x^1\|_2, \dots, \|z - x^m\|_2 \right)^T \right\|_2,$$

ohne Nebenbedingungen. Als Optimalpunkt berechnet man das arithmetische Mittel der Punkte

$$\bar{z} = \frac{1}{m} \sum_{i=1}^m x^i.$$

Ein Problem dieses Ansatzes besteht darin, dass Ausreißer bei wachsender Punkteanzahl m vernachlässigt werden. Eine Möglichkeit dieses Problem zu umgehen, könnte

$$\|x\|_\infty = \max_{j=1, \dots, n} |x_j|$$

als äußere Norm zu verwenden.

Dieses Problem ist koerziv, denn es gilt

$$f(z) = \sqrt{\sum_{i=1}^m \|z - x^i\|_2^2} \geq \sqrt{\|z - x^1\|_2^2} = \|z - x^1\| \geq \left| \|z\| - \|x^1\|_2 \right| \xrightarrow{\|z\| \rightarrow \infty} +\infty$$

Somit ist dieses Beispiel nach 1.2.26 und Korollar 1.2.30 lösbar.

Definition (Epigraph). Für $M \subseteq \mathbb{R}^n$ und $f: M \rightarrow \mathbb{R}$ heißt die Menge

$$\text{epi}(f, M) = \{(x, \alpha) \in M \times \mathbb{R} \mid f(x) \leq \alpha\}$$

Epigraph von f auf M . Der Epigraph besteht aus dem Graphen von f auf M sowie allen darüber liegenden Punkten.

Die Suche nach einem Minimalpunkt von f auf M ist äquivalent zur Suche nach einem (x, α) mit minimaler α -Komponente auf dem Epigraphen von f (vgl. Übung 1.3.7).

Definition (Epigraph-Umformulierung). Das Problem

$$P : \quad \min_x f(x) \text{ s.t. } x \in M$$

ist in diesem Sinne also äquivalent zu

$$P_{\text{epi}} : \quad \min_{(x, \alpha)} \alpha \text{ s.t. } f(x) \leq \alpha, \quad x \in M.$$

Beispiel (Epigraph-Umformulierung der Punktwolke).

Im vorliegenden Fall des Zentrums einer Punktwolke mit Maximumsnorm als äußerer Norm gilt

$$P : \quad \min_{z \in \mathbb{R}^n} \left\| \left(\|z - x^1\|_2, \dots, \|z - x^m\|_2 \right)^T \right\|_2,$$

also $M = \mathbb{R}^n$ und

$$f(z) = \max_{i=1, \dots, m} \|z - x^i\|_2.$$

Die Epigraph-Umformulierung lautet dann

$$P_{\text{epi}} : \quad \min_{(z, \alpha)} \alpha \text{ s.t. } f(z) \leq \alpha,$$

wobei die Nebenbedingung von P_{epi} ausgeschrieben lautet:

$$\max_{i=1,\dots,m} \|z - x^i\|_2 \leq \alpha;$$

bzw. das Maximum ausnutzend:

$$P_{epi} : \min_{(z,\alpha)} \alpha \text{ s.t. } \|z - x^i\|_2 \leq \alpha, \quad i = 1, \dots, m$$

d.h. man sucht einen Kreis mit Mittelpunkt z und minimalem Radius α zu finden, der alle Punkte x^1, \dots, x^m enthält.

Beispiel (1.1.6, 1.2.27, Clusteranalyse). Anstatt wie bei der Punktwolke nur nach einem Mittelpunkt zu suchen, gehen wir nun von k Zentren aus. Das Problem ergibt sich dann zu

$$P : \min_{z^1, \dots, z^k \in \mathbb{R}^n} \left\| \left(\min_{\ell=1, \dots, k} \|z^\ell - x^1\|, \dots, \min_{\ell=1, \dots, k} \|z^\ell - x^m\| \right)^T \right\|$$

Als äußere Norm benutzt man häufig die ℓ_1 -Norm, d.h.

$$P_1 : \min_{z^1, \dots, z^k \in \mathbb{R}^n} \sum_{i=1}^m \min_{\ell=1, \dots, k} \|z^\ell - x^i\|$$

oder auch die ℓ_∞ -Norm mit

$$P_\infty : \min_{z^1, \dots, z^k \in \mathbb{R}^n} \max_{i=1, \dots, m} \min_{\ell=1, \dots, k} \|z^\ell - x^i\|.$$

Beachte: der Vektor der Entscheidungsvariable im Problem hat in jedem Fall die Dimension $n \cdot k$, so dass in praktisch relevanten Anwendungen häufig hochdimensionale Probleme auftreten. Dieses Problem ist nicht koerziv.

1.2 Lösbarkeit

Definition (Infimum). Wir bezeichnen mit $\alpha \in \mathbb{R}$ die **untere Schranke** für f auf M , falls

$$\forall x \in M : \quad \alpha \leq f(x)$$

gilt. Das **Infimum** von f auf M ist die größte untere Schranke von f auf M , es gilt also $v = \inf_{x \in M} f(x)$ falls

- $\forall x \in M : v \leq f(x)$ (d.h. v ist selbst untere Schranke) und
- für alle unteren Schranken α von f auf M als kleinste obere Schranke definiert.

Analog wird das **Supremum** $\sup_{x \in M} f(x)$ von f auf M als kleinste obere Schranke definiert.

Falls f auf M nicht nach unten beschränkt ist, setzt man formal

$$\inf_{x \in M} f(x) = -\infty,$$

und für das Infimum über die leere Menge definiert man formal

$$\inf_{x \in \emptyset} f(x) = +\infty.$$

Definition (1.2.3, Lösbarkeit). Das Minimierungsproblem P heißt **lösbar**, falls ein $\bar{x} \in M$ mit $\inf_{x \in M} f(x) = f(\bar{x})$ existiert.

Um anzudeuten, dass das Infimum angenommen wird, schreiben wir $\min_{x \in M} f(x)$ anstelle von $\inf_{x \in M} f(x)$.

Satz (1.2.5). Das Minimierungsproblem P ist genau dann lösbar, wenn es einen globalen Minimalpunkt besitzt.

1.3 Arten von Unlösbarkeit

Definition (1.2.6). Wir bezeichnen mit $\text{pr}(z, M)$ die Menge der **orthogonalen Projektionen** eines Punktes $z \in \mathbb{R}^n$ auf $M \subseteq \mathbb{R}^n$ und mit

$$\text{pr}(Z, M) = \left\{ \text{pr}(z, M) \mid z \in M \right\}$$

die orthogonale Projektion einer Menge $Z \subseteq \mathbb{R}^n$ auf M .

Definition (1.2.7, Parallelprojektion). Für $Z \subseteq \mathbb{R}^n = \mathbb{R}^k \times \mathbb{R}^\ell$ heißt

$$\text{pr}_a(Z) = \left\{ a \in \mathbb{R}^k \mid \exists b \in \mathbb{R}^\ell: (a, b) \in Z \right\}.$$

Parallelprojektion von Z in die Menge \mathbb{R}^k .

Folgerung (Parallelprojektion der Epigraphen). Die Parallelprojektion von $\text{epi}(f, M)$ auf den „ α -Raum“ \mathbb{R} lautet also

$$\text{pr}_\alpha \text{epi}(f, M) = \left\{ \alpha \in \mathbb{R} \mid \exists x \in M : f(x) \leq \alpha \right\}$$

Satz (1.2.8). Für jedes Optimierungsproblem P hat die Menge $\text{pr}_\alpha \text{epi}(f, M)$ genau eine der vier Gestalten

- a) $\text{pr}_\alpha \text{epi}(f, M) = \emptyset$
- b) $\text{pr}_\alpha \text{epi}(f, M) = \mathbb{R}$
- c) $\text{pr}_\alpha \text{epi}(f, M) = (v, +\infty)$ mit $v \in \mathbb{R}$
- d) $\text{pr}_\alpha \text{epi}(f, M) = [v, +\infty)$ mit $v \in \mathbb{R}$

Ferner ist der Fall a) zu $M = \emptyset$

Beispiel (Distanz eines Punktes von einer Menge). Für eine Menge $M \subseteq \mathbb{R}^n$ und einen Punkt $z \in \mathbb{R}^n$ heißt

$$\text{dist}(z, M) := \inf_{x \in M} \|x - z\|_2$$

Distanz von z zu M . Wegen der Nichtnegativität der Norm ist Unlösbarkeit wegen Unbeschränktheit für dieses Problem nicht möglich. Ferner gilt

$$\text{dist}(z, \emptyset) = +\infty \text{ für jedes } z \in \mathbb{R}^n.$$

Beachte: die fehlende Abgeschlossenheit von M führt hier zu Unlösbarkeit.

1.4 Der Satz von Weierstraß

Eine Minimalvoraussetzung für die Lösbarkeit ist offensichtlich die Konsistenz von M , also $M \neq \emptyset$.

Satz (1.2.10, Satz von Weierstraß). *Die Menge $M \subseteq \mathbb{R}^n$ sei nicht-leer und kompakt, und die Funktion $f: M \rightarrow \mathbb{R}$ sei stetig. Dann besitzt f auf M einen globalen Minimalpunkt und einen globalen Maximalpunkt.*

Beachte: außer der Konsistenz von M keine dieser Bedingungen auch notwendig für Lösbarkeit ist. Dass die Bedingungen stärker als nötig sind, sieht man alleine schon daran, auch die Existenz eines globalen Maximalpunkts garantiert wird.

Definition (Untere Niveaumenge). Für $M \subseteq \mathbb{R}^n$, $f: M \rightarrow \mathbb{R}$ und $\alpha \in \overline{\mathbb{R}}$ heißt

$$\text{lev}_{\leq}^{\alpha}(f, \mathbb{R}^n) = \{x \in M \mid f(x) \leq \alpha\}$$

untere Niveaumenge von f auf M zum Niveau α . Im Falle $M = \mathbb{R}^n$ schreiben wir auch kurz

$$f_{\leq}^{\alpha} := \text{lev}_{\leq}^{\alpha}(f, \mathbb{R}^n) = \{x \in \mathbb{R}^n \mid f(x) \leq \alpha\}$$

Für jede reellwertige Funktion f gilt $\text{lev}_{\leq}^{+\infty}(f, M) = M$ und $\text{lev}_{\leq}^{-\infty}(f, M) = \emptyset$. In der Analysis wird gezeigt, dass für stetiges f die Mengen f_{\leq}^{α} für alle $\alpha \in \mathbb{R}$ abgeschlossen sind.

Beachte: $\text{lev}_{\leq}^{\alpha}(f, \mathbb{R}^n) \neq \text{epi}(f, M)$, bzw.

$$\{x \in M \mid f(x) \leq \alpha\} \neq \{(x, \alpha) \in M \times \mathbb{R} \mid f(x) \leq \alpha\}$$

Definition (Menge der Minimalpunkte). Wir führen die Menge der globalen Minimalpunkte

$$S = \{\bar{x} \in M \mid \forall x \in M : f(x) \geq f(\bar{x})\}$$

von P ein. Die Lösbarkeit von P ist dann zu $S \neq \emptyset$ äquivalent.

Lemma (1.2.13). *Es sei $v = \inf_{x \in M} f(x)$. Dann gilt $S = \text{lev}_{\leq}^v(f, M)$.*

Lemma (1.2.15). Für ein $\alpha \in \mathbb{R}$ sei $\text{lev}_{\leq}^{\alpha}(f, M) \neq \emptyset$. Dann gilt $S \subseteq \text{lev}_{\leq}^{\alpha}(f, M)$.

Satz (1.2.16, Verschärfter Satz von Weierstraß). Für eine Menge $M \subseteq \mathbb{R}^n$ sei $f: M \rightarrow \mathbb{R}$ stetig, und mit einem $\alpha \in \mathbb{R}$ sei $\text{lev}_{\leq}^{\alpha}(f, M)$ nicht-leer und kompakt. Dann ist auch S nicht-leer und kompakt.

Beispiel. Betrachtet werde das Problem

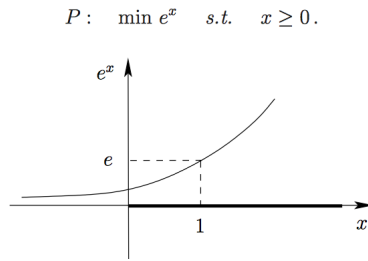


Abbildung 1.17: e^x mit $x \geq 0$

Hier ist $M = \{x \in \mathbb{R} \mid x \geq 0\}$ unbeschränkt, der Satz von Weierstraß (Satz 1.2.10) also nicht anwendbar. Aber beispielsweise mit $\alpha = e$ ist

$$\text{lev}_{\leq}^e(f, M) = \{x \in M \mid e^x \leq e\} = \{x \geq 0 \mid x \leq 1\} = [0, 1]$$

nicht-leer und kompakt. Folglich ist der Verschärfte Satz von Weierstraß (Satz 1.2.16) anwendbar und P daher lösbar.

Korollar (1.2.19, Verschärfter Satz von Weierstraß für unrestringierte Probleme). Sei $f: \mathbb{R}^n \rightarrow \mathbb{R}$ stetig und mit einem $\alpha \in \mathbb{R}$ sei f_{\leq}^{α} nicht-leer und kompakt. Dann ist auch S nicht-leer und kompakt.

Definition (1.2.23, Koerzitivität bei ∞). Gegeben sei eine Funktion $f: M \rightarrow \mathbb{R}$ mit $M \subseteq \mathbb{R}^n$. Falls für alle Folgen $(x^{\nu}) \subseteq M$ mit $\lim_{\nu} \|x^{\nu}\| \rightarrow +\infty$ auch

$$\lim_{\nu} f(x^{\nu}) = +\infty$$

gilt, dann heißt f koerziv bei ∞ auf M . Falls M abgeschlossen ist, heißt f kurz koerziv auf M .

Beispiel (1.2.28, Beispiel). Auf kompakten Mengen M ist jede Funktion f trivialerweise koerziv.

Lemma (1.2.29). Die Funktion $f: M \rightarrow \mathbb{R}$ sei koerziv bei ∞ auf der Menge $M \subseteq \mathbb{R}^n$. Dann sind die Mengen $\text{lev}_{\leq}^{\alpha}(f, M)$ für jedes Niveau $\alpha \in \mathbb{R}$ beschränkt.

Korollar (1.2.30). Es sei M nicht-leer und abgeschlossen und $f: M \rightarrow \mathbb{R}$ sei stetig und koerziv auf M . Dann ist S nicht-leer und kompakt.

Bemerkung (1.2.35). Auf der abgeschlossenen Menge $X \subseteq \mathbb{R}^n$ seien die Funktionen $f, g_i, i \in I$, stetig, die Menge

$$\{x \in X \mid g_i(x) \leq 0, i \in I\}$$

sei nicht-leer, und mindestens eine der Funktionen $f, g_i, i \in I$, sei koerziv auf X . Zeigen Sie, dass die Menge S der Optimalpunkte von f auf M dann nicht-leer und kompakt ist.

Definition (1.2.37, Koerzitivität). Gegeben seien eine (nicht notwendigerweise abgeschlossene) Menge $M \subseteq \mathbb{R}^n$ und eine Funktion $f: M \rightarrow \mathbb{R}$. Falls für alle Folgen $(x^{\nu}) \subseteq M$ mit $\lim_{\nu} \|x^{\nu}\| \rightarrow \infty$ und alle konvergenten Folgen $(x^{\nu}) \subseteq M$ mit $\lim_{\nu} x^{\nu} \notin M$ die Bedingung

$$\lim_{\nu} f(x^{\nu}) = +\infty$$

gilt, dann heißt f **koerziv** auf M .

Beispiel (1.2.36, 1.2.38). Gegeben seien N Beobachtungen $\hat{x}_1, \dots, \hat{x}_N \geq 0$ mit $\bar{x} = \frac{1}{N} \sum_{i=1}^N \hat{x}_i > 0$, die als Realisierungen stochastisch unabhängiger und mit Parameter $\lambda > 0$ exponentialverteilter Zufallsvariablen X_1, \dots, X_N aufgefasst werden. Gesucht ist der Parameter λ , der zu den Beobachtungen „am besten passt“. Dazu kann man mit Hilfe der Dichtefunktionen der einzelnen X_i ,

$$f(\lambda x_i) = \begin{cases} \lambda e^{-\lambda x_i}, & x_i \geq 0 \\ 0, & x_i < 0, \end{cases}$$

zunächst die gemeinsame Dichte aller Zufallsvariablen

$$L(\lambda, x) = \prod_{i=1}^N f(\lambda, x_i)$$

betrachten. Der Maximum-Likelihood-Schätzer bestimmt dann λ als optimalen Punkt des Problems

$$ML : \quad \max_{\lambda} L(\lambda, \tilde{x}) \text{ s.t. } \lambda > 0.$$

Die zulässige Menge $M = (0, \infty)$ dieses Problems ist offensichtlich nicht abgeschlossen. Es ist auch sinnlos, den Parameterwert $\lambda = 0$ künstlich hinzuzufügen, da $f(0, x)$ keine Wahrscheinlichkeitsdichte ist.

Wir werden im Folgenden sehen, wie die Lösbarkeit dieses Problems trotzdem garantiert werden und später auch einen globalen Maximalpunkt bestimmen. Dazu berechnen wir zunächst

$$L(\lambda, \hat{x}) = \prod_{i=1}^N \lambda e^{-\lambda \hat{x}_i} = \lambda^N e^{-\lambda N \bar{x}}, \ell(\lambda, \hat{x}) := \log(L(\lambda, \hat{x})) = N \log(\lambda) - \lambda N \bar{x}$$

Da die Funktion \log streng monoton wachsend auf dem Bildbereich $(0, \infty)$ von L ist, kann man Hilfe von Übung 1.3.5 zeigen, dass das Problem

$$ML_{\log} : \quad \max_{\lambda} \ell(\lambda, \hat{x}) \text{ s.t. } \lambda > 0$$

dieselben Optimalpunkte wie ML besitzt. Schließlich streichen wir mittels Übung 1.3.1a) die Konstante $N > 0$ aus der Zielfunktion und gehen zum äquivalenten Minimierungsproblem

$$P_{LM} : \quad \min_{\lambda} \lambda \bar{x} - \log(\lambda) \text{ s.t. } \lambda > 0$$

über.

Für $\bar{x} > 0$ ist die Zielfunktion $f(\lambda) = \lambda \bar{x} - \log(\lambda)$ des Problems P_{ML} koerziv auf $M = (0, \infty)$. Wir merken an, dass nur die Log-Likelihood-Funktion ℓ zu Koerzivitt fhrt, die Likelihood-Funktion L selbst aber nicht.

Das Problem P_{ML} und damit auch das Problem ML sind nach Korollar 1.2.40 lsbar.

Lemma (1.2.39). Die Funktion $f: M \rightarrow \mathbb{R}$ sei stetig und koerziv auf der Menge $M \subseteq \mathbb{R}^n$. Dann sind die Mengen $\text{lev}_{\leq}^{\alpha}(f, M)$ fr jedes Niveau $\alpha \in \mathbb{R}$ kompakt.

Korollar (1.2.40). Es sei M nicht-leer, und $f: M \rightarrow \mathbb{R}$ sei stetig und koerziv auf M . Dann ist S nicht-leer und kompakt.

Bemerkung (Rechenregeln).

- *Skalare und Vielfache (1.3.1)* $\alpha \geq 0 : \min_{x \in M} (\alpha f(x) + \beta) = \alpha (\min_{x \in M} f(x)) + \beta,$
 $\alpha < 0 : \min_{x \in M} (\alpha f(x) + \beta) = \alpha \left(\max_{x \in M} f(x) \right) + \beta, \quad \min_{x \in M} (f(x) + g(x)) \geq \min_{x \in M} f(x) + \min_{x \in M} g(x)$
- $\min_{x \in X} \max_{y \in Y} f(x, y) \geq \max_{y \in Y} \min_{x \in X} f(x, y)$
- *Monotone Transformation (1.35)* $M \subseteq \mathbb{R}^n, f: M \rightarrow Y$ und $\psi: Y \rightarrow \mathbb{R}$ eine streng monotone wachsende Funktion. Dann gilt

$$\min_{x \in M} \psi(f(x)) = \psi(\min_{x \in M} f(x))$$

- *Projektionsumformulierung (1.3.6): folgende Probleme sind äquivalent*

$$P : \min_{(x,y) \in \mathbb{R}^n \times \mathbb{R}^m} f(x) \text{ s.t. } (x, y) \in M$$

und

$$P_{proj} : \min_{x \in \mathbb{R}^n} f(x) \text{ s.t. } x \in \text{pr}_x M$$

- *Monotones Funktional (1.3.8): wir nennen $F: \mathbb{R}^k \rightarrow \mathbb{R}$ monoton, falls $x \leq y \Rightarrow F(x) \leq F(y)$.*

Bemerkung (1.3.7, Epigraph Umformulierung). Gegeben seien $M \subseteq \mathbb{R}^n$ und eine Funktion $f: M \rightarrow \mathbb{R}$. Dann sind die folgenden Probleme äquivalent:

$$P : \min_{x \in \mathbb{R}^n} f(x) \text{ s.t. } x \in M$$

und

$$P_{epi} : \min_{(x,\alpha) \in \mathbb{R}^n \times \mathbb{R}} \alpha \text{ s.t. } f(x) \leq \alpha, x \in M$$

Bemerkung (1.3.9, Verallgemeinerte Epigraph Umformulierung). Gegeben seien $X \subseteq \mathbb{R}^n$, Funktionen $f: X \rightarrow \mathbb{R}^k$ und $g: X \rightarrow \mathbb{R}^\ell$ sowie monotone Funktionen $F: \mathbb{R}^k \rightarrow \mathbb{R}$ und $G: \mathbb{R}^\ell \rightarrow \mathbb{R}$. Dann sind die folgenden Probleme äquivalent:

$$P : \min_{x \in \mathbb{R}^n} F(f(x)) \text{ s.t. } G(g(x)) \leq 0, x \in M$$

und

$$P_{epi} : \min_{(x,\alpha,\beta) \in \mathbb{R}^n \times \mathbb{R}^k \times \mathbb{R}^\ell} F(\alpha) \text{ s.t. } G(\beta) \leq 0, f(x) \leq \alpha, g(x) \leq \beta, x \in M$$

Kapitel 2

Konvexe Optimierungsprobleme

2.1 Konvexität

Definition (2.1.1, Konvexe Mengen und Funktionen).

a) Eine Menge $M \subseteq \mathbb{R}^n$ heißt konvex, falls $\forall x, y \in M, \lambda \in (0, 1): (1 - \lambda)x + \lambda y \in M$ gilt. Konvexe Mengen brauchen weder beschränkt noch abgeschlossen zu sein.

b) Für eine konvexe Menge $M \subseteq \mathbb{R}^n$ heißt eine Funktion $f: M \rightarrow \mathbb{R}$ konvex, falls

$$\forall x, y \in M, \lambda \in (0, 1): f((1 - \lambda)x + \lambda y) \leq (1 - \lambda)f(x) + \lambda f(y)$$

d.h. der Funktionsgraph von f verläuft unter jeder seiner Sekanten.

c) Für eine konvexe Menge $M \subseteq \mathbb{R}^n$ heißt eine Funktion $f: M \rightarrow \mathbb{R}$ strikt konvex, falls in b) für $x \neq y$ sogar die strikte Ungleichung gilt.

d) Für eine konvexe Menge $M \subseteq \mathbb{R}^n$ heißt eine Funktion $f: M \rightarrow \mathbb{R}$ gleichmäßig konvex (auf M), falls mit einer Konstanten $c > 0$ die Funktion $f(x) - \frac{c}{2}\|x\|_2^2$ konvex auf M ist.

e) Für eine konvexe Menge $M \subseteq \mathbb{R}^n$ heißt eine Funktion $f: M \rightarrow \mathbb{R}$ konkav, strikt konkav oder gleichmäßig konkav (auf M) falls $-f$ konvex, strikt konvex bzw. gleichmäßig konvex auf M ist.

Übung (2.1.2). Auf einer konvexen Menge $M \subseteq \mathbb{R}^n$ ist die Funktion $f: M \rightarrow \mathbb{R}$ genau dann konvex, wenn die Mengen $\text{epi}(f, M)$ konvex ist.

Definition (2.1.3, Konvexes Optimierungsproblem). Das Optimierungsproblem

$$P: \quad \min f(x) \text{ s.t. } x \in M$$

heißt **konvex**, falls M und $f: M \rightarrow \mathbb{R}$ konvex sind.

Da $M = \mathbb{R}^n$ eine konvexe Menge ist, sind unrestringierte Probleme genau dann konvex, wenn f konvex auf \mathbb{R}^n ist.

Satz (2.1.4). P sei konvex. Dann ist jeder lokale Minimalpunkt von P auch globaler Minimalpunkt von P .

Der Grund für diesen Effekt liegt darin, dass Konvexität eine globale Voraussetzung an P ist.

Übung (2.1.5). Die Menge $M \subseteq \mathbb{R}^n$ und die Funktion $f: M \rightarrow \mathbb{R}$ seien konvex. Dann ist $\text{lev}_{\leq}^{\alpha}(f, M)$ für jedes $\alpha \in \mathbb{R}$ konvex. Die Umkehrung dieser Aussage ist falsch.

Übung (2.1.6). Der Schnitt beliebig vieler konvexer Mengen ist konvex.

Korollar (2.1.8). Mit beliebigen Indexmengen I, J seien die Funktionen $g_i: \mathbb{R}^n \rightarrow \mathbb{R}$, $i \in I$, konvex, und die Funktionen $h_j: \mathbb{R}^n \rightarrow \mathbb{R}$, $j \in J$, linear. Dann ist die Menge

$$M = \{x \in \mathbb{R}^n \mid g_i(x) \leq 0, i \in I, h_j(x) = 0, j \in J\}$$

konvex.

Beispiel (2.1.9). Falls f, g_i auf \mathbb{R}^n konvex und h_j lineare Funktionen sind, dann ist

$$P: \quad \min f(x) \text{ s.t. } g_i(x) \leq 0, i \in I, h_j(x) = 0, j \in J$$

ein konvexes Optimierungsproblem.

Beispiel (2.1.10). Mit $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$ und einer (m, n) -Matrix $A = (a_1^T, \dots, a_m^T)^T$ mit $a_i \in \mathbb{R}^n$ ist

$$P: \quad \min c^T x \text{ s.t. } Ax \leq b$$

ein lineares Optimierungsproblem (die Ungleichungsnebenbedingungen können sowohl eine Nichtnegativitätsbedingung $x \geq 0$ enthalten als auch Gleichungen modellieren).

P ist auch ein konvexes Optimierungsproblem, denn mit den Setzungen $f(x) = c^T x$, und $g_i(x) = a_i^T x - b_i$ sind $f, g_i: \mathbb{R}^n \rightarrow \mathbb{R}$ linear und damit konvex auf \mathbb{R}^n .

Zum Beispiel setzt man für das lineare Optimierungsproblem

$$\min x_1 + x_2 \text{ s.t. } x \geq 0$$

$$f(x) = x_1 + x_2, \quad g_1(x) = -x_1 \text{ und } g_2(x) = -x_2.$$

2.2 C^1 -Charakterisierung

Bemerkung (Kettenregel). $D(f \circ g)(\bar{x}) = Df(g(\bar{x})) \cdot Dg(\bar{x})$

Satz (2.2.1, Linearisierung per Satz von Taylor im \mathbb{R}^n). Für eine nicht-leere, offene und konvexe Menge $U \subseteq \mathbb{R}^n$ sei die Funktion $f: \rightarrow \mathbb{R}$ differenzierbar an $x \in U$. Dann gilt für alle $y \in U$

$$f(y) = f(x) + \langle \nabla f(x), y - x \rangle + o(\|y - x\|),$$

wobei $o(\|y - x\|)$ einen Ausdruck der Form $w(y)\|y - x\|$ mit einer an x stetigen Funktion w mit $w(x) = 0$ bezeichnet.

Satz (2.2.2., C^1 -Charakterisierung von Konvexität). Auf einer konvexen Mengen $M \subseteq \mathbb{R}^n$ ist eine Funktion $f \in C^1(M, \mathbb{R})$ genau dann konvex, wenn folgendes gilt

$$\forall x, y \in M: \quad f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle$$

Die C^1 -Charakterisierung besagt, dass eine C^1 -Funktion genau dann konvex auf M ist, wenn ihr Graph über jeder seiner Tangentialebenen verläuft.

2.3 Lösbarkeit

Für die Lösbarkeit von P ist Konvexität alleine kein wesentlicher Vorteil: sowohl $f_1(x) = (x - 5)^2$ als auch $f_2(x) = e^x$ sind sogar strikt konvex, aber nur f_1 besitzt einen globalen Minimalpunkt auf \mathbb{R} . Im Folgenden werden wir sehen, dass der entscheidende Vorteil von f_1 gegenüber f_2 die gleichmäßige Konvexität ist.

Übung (2.3.1). Für eine konvexe Menge $M \subseteq \mathbb{R}^n$ sei $f: M \rightarrow \mathbb{R}$ gleichmäßig konvex. Dann ist f auch strikt konvex auf M .

Lemma (2.3.2). Für eine abgeschlossene und konvexe Menge $M \subseteq \mathbb{R}^n$ sei $f: M \rightarrow \mathbb{R}$ gleichmäßig konvex. Dann ist f auch

- a) koerziv auf M und
- b) stetig auf dem Inneren von M .

Satz (2.3.3). P sei konvex. Dann gelten die folgenden Aussage:

- a) Die Menge der Minimalpunkte ist konvex.
- b) Falls f strikt konvex auf M ist, dann besitzt S höchstens ein Element.
- c) Es sei M nicht-leer und abgeschlossen, und f sei gleichmäßig konvex und stetig auf M . Dann besitzt S genau ein Element (d.h. P ist eindeutig lösbar).

Angemerkt sei, dass die Stetigkeitsforderung an f in Satz 2.3.3c) nach Lemma 2.3.2b) unnötig ist, falls die Menge M mit ihrem Inneren übereinstimmt (z.B. für $M = \mathbb{R}^n$).

2.4 Optimalitätsbedingungen für unrestringierte Probleme

Wir betrachten im Folgenden das unrestringierte Problem

$$P : \min f(x).$$

Allgemeiner bezeichnet man auch Probleme mit offener zulässiger Menge M als unrestringiert. In der Tat lassen sich die im Folgenden besprochenen Optimalitätsbedingungen ohne weiteres auf diesen Fall übertragen, was wir aus Gründen der Übersichtlichkeit aber nicht explizit angeben werden.

Definition (2.4.1, Kritischer Punkt). Ein Punkt $\bar{x} \in \mathbb{R}^n$ heißt kritisch für $f \in C^1(\mathbb{R}^n, \mathbb{R})$, falls $\nabla f(\bar{x}) = 0$ gilt.

Satz (2.4.2, Fermat'sche Regel, notwendige Optimalitätsbedingung). *Der Punkt $\bar{x} \in \mathbb{R}^n$ sei lokal minimal für $f \in C^1(\mathbb{R}^n, \mathbb{R})$. Dann ist \bar{x} kritischer Punkt von f . Achtung dies ist lediglich notwendig und nicht hinreichend.*

Beispiel (2.4.3, Fortsetzung von 1.1.5 - Zentrum einer Punktwolke).

Gegeben seien Punkte $x^1, x^2, \dots, x^m \in \mathbb{R}^n$. Gesucht ist ein Punkt $z \in \mathbb{R}^n$ „im Zentrum der Punkte“. Benutzt man die 2-Norm so führt dies auf das Optimierungsproblem

$$P : \min_{z \in \mathbb{R}^n} \left\| \left(\|z - x^1\|_2, \dots, \|z - x^m\|_2 \right)^T \right\|_2,$$

ohne Nebenbedingungen. Als Optimalpunkt berechnet man das arithmetische Mittel der Punkte

$$\bar{z} = \frac{1}{m} \sum_{i=1}^m x^i.$$

Beweis hiervon folgt aus:

$$f(z) = \sqrt{\sum_{i=1}^m \|z - x^i\|_2^2} = \sqrt{mz^T z - 2z^T \sum_{i=1}^m x^i + \sum_{i=1}^m \|x^i\|_2^2}$$

so dass f nicht differenzierbar ist und Satz 2.4.2 nicht angewendet werden kann. Nach Übung 1.3.5 mit $\psi(y) = y^2$ und $Y = \{y \in \mathbb{R} \mid y \geq 0\}$ besitzt f aber dieselben Minimalpunkte wie die stetig differenzierbare Funktion

$$f^2(z) = mz^T z - 2z^T \sum_{i=1}^m x^i + \sum_{i=1}^m \|x^i\|_2^2$$

Kritische Punkte dieser Funktion sind genau die Lösungen der Gleichung

$$0 = \nabla \left(f^2(z) \right) = 2mz - 2 \sum_{i=1}^m x^i$$

also besitzt f^2 den eindeutigen kritischen Punkt

$$\bar{z} = \frac{1}{m} \sum_{i=1}^m x^i$$

Tatsächlich ist \bar{z} auch globaler Minimalpunkt von F^2 sowie von f , wie man durch folgende Argumente sieht: nach Beispiel 1.2.31 existiert zunächst ein globaler Minimalpunkt \tilde{z} von f und damit auch von f^2 auf \mathbb{R}^n . Aufgrund der Fermat'schen Regel (Satz 2.4.2) muss \tilde{z}

kritischer Punkt von f^2 ist aber das gerade berechnete \bar{z} , so dass nur $\tilde{z} = \bar{z}$ gelten kann. Also ist \bar{z} globaler Minimalpunkt von f auf \mathbb{R}^n .

In Kapitel 2.5 werden wir außerdem nachweisen, dass f^2 eine auf \mathbb{R}^n konvexe Funktion ist, was es ermöglichen wird, die globale Minimalität von \bar{z} alternativ zu beweisen, ohne zunächst die Lösbarkeit des zugrundeliegenden Optimierungsproblems zu zeigen.

Beispiel (2.4.4, Fortsetzung von 1.2.36). Die Zielfunktion $f(\lambda) = \lambda\bar{x} - \log(\lambda)$ mit $\bar{x} > 0$ des Problems P_{ML} aus Beispiel 1.2.36 erfüllt $f'(\lambda) = \bar{x} - \frac{1}{\lambda}$, besitzt als eindeutiges kritischen Punkt also $\bar{\lambda} = \frac{1}{\bar{x}}$.

Wie in Beispiel 2.4.3 lässt sich nun argumentieren, dass $\bar{\lambda} = \frac{1}{\bar{x}}$ globaler Minimalpunkt von P_{ML} und damit der gesuchte Maximum-Likelihood-Schätzer der Exponentialverteilung ist: nach Beispiel 1.2.41 existiert zunächst ein globaler Minimalpunkt $\tilde{\lambda}$ von P_{ML} . Da sich Optimierungsprobleme mit offenen zulässigen Mengen wie unrestringierte Optimierungsprobleme behandeln lassen, muss $\tilde{\lambda}$ nach der Fermat'schen Regel ein kritischer Punkt von f sein. Einzig kritischer Punkt von f ist aber das gerade berechnete $\bar{\lambda}$, woraus die Behauptung folgt.

In Kapitel 2.5 werden wir außerdem sehen, dass f auf der zulässigen Menge $(0, \infty)$ von P_{ML} konvex ist, was es wieder ermöglichen wird, die globale Minimalität von $\bar{\lambda}$ alternativ zu beweisen, ohne zunächst die Lösbarkeit von P_{ML} zu zeigen.

Satz (2.4.5, Hinreichende Optimalitätsbedingung). $f \in C^1(\mathbb{R}^n, \mathbb{R})$ sei konvex. Dann ist jeder kritische Punkt von f globaler Minimalpunkt von f .

Korollar (2.4.6, Charakterisierung globaler Minimalpunkte). $f \in C^1(\mathbb{R}^n, \mathbb{R})$ sei konvex. Dann sind die globalen Minimalpunkte genau die kritischen Punkte von f .

Zur Bestimmung globaler Minimalpunkte unrestringierter konvexer C^1 -Probleme genügt es also nicht nur, lokale Minimalpunkte zu suchen (wie schon in Satz 2.1.4 gesehen), sondern sogar nur kritische Punkte. Das globale Minimierungsproblem ist damit auf das Lösen der Gleichung $\nabla f(x) = 0$ zurückgeführt.

Insbesondere erhält man auch diese Aussage: falls f keinen kritischen Punkt besitzt, dann auch keinen globalen Minimalpunkt. Dazu muss allerdings bewiesen werden, dass f keinen kritischen Punkt besitzen kann (ein einfaches Beispiel hierfür ist $f(x) = e^x$).

2.5 C^2 -Charakterisierung von Konvexität

Satz (2.5.1, Quadratische Approximation per Satz von Taylor im \mathbb{R}^n). *Für eine nicht-leere, offene und konvexe Menge $U \subseteq \mathbb{R}^n$ sei die Funktion $f: U \rightarrow \mathbb{R}$ zweimal differenzierbar an $x \in U$. Dann gilt für alle $y \in U$*

$$f(y) = f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2} (y - x)^T D^2 f(x) (y - x) + o(\|y - x\|^2)$$

Der Fehlerterm lässt sich dabei auch mit Hilfe des Lagrange-Restgliedes angeben:

$$o(\|y - x\|^2) = \frac{1}{2} (y - x)^T D^2 f(\xi) (y - x) - \frac{1}{2} (y - x)^T D^2 f(x) (y - x),$$

wobei ξ ein nicht näher bekannter Punkt auf der Verbindungsstrecke zwischen x und y ist.

Definition. Eine symmetrische (n, n) -Matrix A heißt **positiv semidefinit** (kurz: $A \succeq 0$), wenn

$$\forall x \in \mathbb{R}^n : \quad x^T A x \geq 0$$

gilt, und **positiv definit** (kurz: $A \succ 0$), wenn die Ungleichung für alle $x \neq 0$ sogar strikt ist.

In der Linearen Algebra wird allerdings gezeigt, dass $A \succeq 0$ ($\succ 0$) genau dann gilt, wenn $\lambda \geq 0$ (> 0) für alle Eigenwerte λ von A gilt.

Satz (2.5.2, Notwendige Optimalitätsbedingung zweiter Ordnung). *Der Punkt \bar{x} sei lokaler Minimalpunkt von $f: \mathbb{R}^n \rightarrow \mathbb{R}$, und f sei an \bar{x} zweimal differenzierbar. Dann gilt $\nabla f(\bar{x}) = 0$ und $D^2 f(\bar{x}) \succeq 0$.*

2.6 C^2 -Charakterisierung von Konvexität

Satz (2.5.3, C^2 -Charakterisierung von Konvexität). *Auf einer konvexen Menge $M \subseteq \mathbb{R}^n$ sei die Funktion $f \in C^2(M, \mathbb{R})$ gegeben.*

a) *Falls $\forall x \in M: D^2 f(x) \succeq 0$ gilt, dann ist f auf M konvex.*

b) *Falls M außerdem offen ist, dann gilt auch die Umkehrung der Aussage in Teil a).*

Dass die Voraussetzung der Offenheit von M in b) nicht nur beweistechnischer Natur ist, sieht man an der C^2 -Funktion $f(x) = x_1^2 - x_2^2$, die nirgends eine positiv semidefinite Hessematrix besitzt, aber trotzdem auf der Menge $M = \mathbb{R} \times \{0\}$ konvex ist. M ist in diesem Beispiel natürlich nicht offen.

Die Voraussetzung der Offenheit von M in Satz 2.5.3b) lässt sich zur Volldimensionalität von M abschwächen, also im Wesentlichen zur Forderung, dass M innere Punkte besitzt. Eine weitergehende Abschwächung ist nicht möglich.

Beispiel (2.5.3). *Die Funktion $f(x) = (x - 5)^2$ erfüllt $f''(x) = 2 \geq 0$ für alle $x \in \mathbb{R}$ und ist damit konvex auf \mathbb{R} .*

Beispiel (2.5.5). *Die Funktion $f(x) = e^x$ erfüllt $f''(x) = e^x \geq 0$ für alle $x \in \mathbb{R}$ und ist damit konvex auf \mathbb{R} .*

Beispiel (2.5.6, Fortsetzung von 1.1.5). *Für den Gradienten der Funktion f^2 auf Beispiel 2.4.3 haben wir bereits die Darstellung*

$$\nabla (f^2(z)) = 2mz - 2 \sum_{i=1}^m x^i$$

hergeleitet, woraus

$$D^2 (f^2(z)) = 2mE \text{ (mit der } (n, n)\text{-Einheitsmatrix } E)$$

folgt. Damit besitzt $D^2 f^2(z)$ den n -fachen Eigenwert $2m \geq 0$. Hieraus folgt die Konvexität von f^2 auf \mathbb{R}^n .

Nach Korollar 2.4.6 ist jeder kritische Punkt von f^2 globaler Minimalpunkt von f^2 und damit auch von f , und der (einzige) kritische Punkt von f^2 ist das in Beispiel 2.4.3 berechnete arithmetische Mittel \bar{z} der Punktwolke. Damit ist alternativ zur Argumentation in Beispiel 2.4.3 die globale Minimalität von \bar{z} gezeigt, ohne zuvor die Lösbarkeit des Optimierungsproblems nachzuweisen.

Beispiel (2.5.7, Fortsetzung von 1.2.36). Die Funktion $f(\lambda) = \lambda\bar{x} - \log(\lambda)$ mit $\bar{x} > 0$ auf Beispiel 2.4.4 erfüllt $f'(\lambda) = \bar{x} - \frac{1}{\lambda}$ und $f''(\lambda) = \frac{1}{\lambda^2} \geq 0$ für alle $\lambda \in (0, \infty)$. Sie ist damit konvex auf $(0, \infty)$. Die Konvex von f korrespondiert zur Konkavität der Log-Likelihood-Funktion ℓ . Wir bemerken, dass die Likelihood-Funktion L selbst nicht konkav ist.

Der in Beispiel 2.4.4 berechnete kritische Punkt $\bar{\lambda} = \frac{1}{\bar{x}}$ von f ist nach Korollar 2.4.6 wieder globaler Minimalpunkt von P_{ML} und damit der gesuchte Maximum-Likelihood-Schätzer der Exponentialverteilung. Auch bei diesem Argument ist (alternativ zu dem in Beispiel 2.4.4) die Betrachtung der Lösbarkeit von P_{ML} unnötig.

Übung (2.5.8, Fortsetzung von Beispiel 1.1.6). Die Zielfunktion

$$f(z^1, \dots, z^k) = \sum_{i=1}^m \min_{\ell=1, \dots, k} \|z^\ell - x^i\| = \left\| \left(\|z^{\ell(1)} - x^1\|, \dots, \|z^{\ell(m)} - x^m\| \right)^T \right\|_1$$

mit $k \geq 2$ aus dem Problem P_1 der Clusteranalyse (Bsp. 1.1.6) ist nicht konvex auf \mathbb{R}^{nk} (wenn die Indizes $\ell(i)$, $i = 1, \dots, m$, allerdings a priori bekannt sind, dann lässt mit einem anderen Argument als der Benutzung von Satz 2.5.3 zeigen, dass f doch konvex ist).

Übung (2.5.9, Hinreichende Bedingung für strikte Konvexität). Auf einer konvexen Menge $M \subseteq \mathbb{R}^n$ sei die Funktion $f \in C^2(M, \mathbb{R})$ gegeben, und es gelte

$$\forall x \in M : \quad D^2 f(x) \succ 0$$

Dann ist f auf M strikt konvex.

Satz (2.5.10, C^2 -Charakterisierung von gleichmäßiger Konvexität). Auf einer konvexen Menge $M \subseteq \mathbb{R}^n$ sei die Funktion $f \in C^2(M, \mathbb{R})$ gegeben.

a) Falls mit einer Konstanten $c > 0$

$$\forall x \in M : \quad \lambda_{\min} \left(D^2 f(x) \right) \geq c$$

gilt, dann ist f auf M gleichmäßig konvex.

b) Falls M außerdem offen ist, dann gilt auch die Umkehrung der Aussage in Teil a)

Beispiel (2.5.11). Die Funktion $f(x) = (x - 5)^2$ erfüllt $f''(x) = 2 > 0$ für alle $x \in \mathbb{R}$ und ist damit nicht nur strikt, sondern sogar gleichmäßig konvex auf \mathbb{R} .

Beispiel (2.5.12). Die Funktion $f(x) = e^x$ erfüllt $f''(x) = e^x > 0$ für alle $x \in \mathbb{R}$ und ist damit strikt konvex auf \mathbb{R} . Allerdings gilt $\lim_{x \rightarrow -\infty} f''(x) = 0$, so dass sie nicht gleichmäßig konvex auf \mathbb{R} ist

Beispiel (2.5.13, Fortsetzung von 1.1.5). Für die Hessematrix des Quadrats der Funktion f aus Beispiel 1.1.5 haben wir in Beispiel 2.5.6 die Darstellung $D^2 f^2(z) = 2mE$ und damit den n -fachen Eigenwert $2m$ für jedes $z \in \mathbb{R}^n$ hergeleitet. Insbesondere gilt dann $\lambda_{\min}(D^2 f^2(z)) = 2m > 0$ für alle $z \in \mathbb{R}^n$, woraus die strikte und sogar gleichmäßige Konvexität von f^2 auf \mathbb{R}^n folgt.

Beispiel (2.5.14, Fortsetzung von 1.2.36). Die Funktion $f(\lambda) = \lambda \bar{x} - \log(\lambda)$ mit $\lambda x > 0$ aus Beispiel 1.2.36 erfüllt $f''(\lambda) = \frac{1}{\lambda^2} > 0$ für alle $\lambda \in (0, \infty)$, aber $\lim_{\lambda \rightarrow \infty} f''(\lambda) = 0$. Sie ist damit zwar strikt, aber nicht gleichmäßig konvex auf $(0, \infty)$.

Nach Beispiel 1.2.38 ist f trotzdem koerziv. Dies zeigt, dass die Bedingung für Koerzivität aus Lemma 2.3.2 nur hinreichend, aber nicht notwendig ist.

2.7 Konvexität und Monotonie der ersten Ableitung

Für $n = 1$, ein offenes Intervall $M \subseteq \mathbb{R}$ und eine Funktion $g \in C^1(M, \mathbb{R})$ ist aus der Analysis bekannt, dass g genau dann monoton wachsend auf M ist, wenn $g'(x) \geq 0$ für alle $x \in M$ gilt. Nach Satz 2.5.3 gilt für $n = 1$, ein offenes Intervall $M \subseteq \mathbb{R}$ und $f \in C^2(M, \mathbb{R})$ also, dass f genau dann konvex auf M ist, wenn f' auf M monoton wächst. In $n = 1$ ist die Monotonie von g äquivalent zur Gültigkeit der Ungleichung $(g(y) - g(x))(y - x) \geq 0$. Die motiviert für das Mehrdimensionale die folgende Definition.

Definition (2.5.15, Monotoner Operator). *Für eine nicht-leere konvexe Menge $M \subseteq \mathbb{R}^n$ heißt $g: M \rightarrow \mathbb{R}^n$ **monoton** auf M , falls folgendes gilt:*

$$\forall x, y \in M : \quad \langle g(y) - g(x), y - x \rangle \geq 0$$

Satz (2.5.16, Monotonie-Charakterisierung von Konvexität). *Auf einer konvexen Menge $M \subseteq \mathbb{R}^n$ ist eine Funktion $f \in C^1(M, \mathbb{R})$ genau dann konvex, wenn ∇f auf M monoton ist.*

2.8 Dualität