Subgradientenverfahren: Theorie und Anwendung

Fakultät für Informatik und Mathematik

Bachelorarbeit

zur Erlangung des akademischen Grades
Bachelor of Science

vorgelegt von

Maximiliano Fontana
geboren am 13.11.1990 in München
Matrikelnummer 102868

im September 2024

Prüfer Prof. Dr. Tomas Sauer

Zusammenfassung

In dieser Arbeit wird das Subgradientenverfahren eingeführt, das eine zentrale Methode zur Optimierung konvexer, aber nicht notwendigerweise überall differenzierbarer Funktionen darstellt. Zur Einführung des Verfahrens wird die zugrundeliegende Theorie konvexer Mengen behandelt, die die sogenannte Konvexitätsbedingung erfüllen. Diese Bedingung spielt eine zentrale Rolle bei der Untersuchung konvexer Funktionen. Es werden sowohl Funktionen betrachtet, die überall differenzierbar sind, als auch solche, die es nicht sind. Für Letztere wird gezeigt, dass sie trotz fehlender Differenzierbarkeit in allen weiteren Punkten richtungsdifferenzierbar und konvex sind. Diese Erkenntnisse führen zur Definition des Subgradienten bzw. des Subdifferentials. Diese ersetzen an der gegebenen Stelle den Gradienten. Das Subgradientenverfahren wird daraufhin als Methode vorgestellt, um das Optimum konvexer Funktionen zu bestimmen. Abschließend wird die Wahl der Schrittweite sowie die Konvergenz des Verfahrens ausführlich diskutiert.

Inhaltsverzeichnis

1 Einleitung **1**

2 Konvexe Mengen **3**
2.1 Grundlagen 3
2.2 Eigenschaften von konvexen Mengen 7

3 Konvexe Funktionen **13**
3.1 Grundlegende Definitionen 13
3.2 Charakterisierung konvexer Funktionen 19

4 Der Subgradient **23**
4.1 Einstieg 23
4.2 Das Subdifferential 25
4.3 Die Subdifferenzierbarkeit 27

5 Das Subgradientenverfahren **29**
5.1 Konvexe Optimierung 29
5.2 Das Lagrange-Multiplikatorenverfahren 32
5.3 Subgradientenverfahren 36
5.4 Konvergenzanalyse 37

6 Diskussion **43**

Literaturverzeichnis **45**

Eigenständigkeitserklärung **47**

Abbildungsverzeichnis

2.1 Eine konvexe Menge 3
2.2 Eine nicht konvexe Menge 3
2.3 Punkt z außerhalb zweier konvexer Mengen 8
2.4 Eine polyedrisch konvexe Menge im $\mathbb{R}^2$ 9
2.5 Konvexe Hülle von Punkten im $\mathbb{R}^2$ 11
2.6 Standard-Simplex Δ_3 in $\mathbb{R}^{3+1}$ 12

3.1 Epigraph der Funktion $f(x) = x^2$ 14
3.2 Nicht Konvexität der Funktion $g(x)$ 17

4.1 Graph der Betragsfunktion 23

5.1 Minimalpunkt der Funktion $f(x, y) = x^2 + y^2$ 34
5.2 Andere Pespektive der Abbildung 5.1 34
5.3 Graph der Funktion $f(x, y) = |x| + |y|$ 35

Tabellenverzeichnis

5.1 Werte des Subgradientenverfahrens von $f(x) = |x|$ 38

6.1 Subgradientenverfahren: Werte von $f(x) = |x|$ bei $x_0 = 10$ 44

1 Einleitung

Beim Betrachten von Funktionen, die in einer Variablen differenzierbar sind, kann festgestellt werden, dass es unendlich viele differenzierbare Funktionen gibt [1]. Um zu überprüfen, ob eine Funktion $f\colon I \to \mathbb{R}$ an einer Stelle $x \in I$ mit $I \subseteq \mathbb{R}$ differenzierbar ist, wird der Differentialquotient verwendet

$$\frac{d}{dx} f(x) := \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}. \tag{1.1}$$

Die daraus resultierenden Differentiationsregeln, wie die Produktregel

$$(f(x)g(x))' = f'(x)g(x) + f(x)g'(x),$$

sind grundlegende Werkzeuge der Analysis [2]. Für den $\mathbb{R}^n$ mit $n \in \mathbb{N}$ gelten ähnliche Differentiationsregeln, die es ermöglichen, die Konzepte des Differentialquotienten und die Differenzierbarkeit auf Funktionen mehrerer Variablen zu erweitern.

Eine naheliegende Frage ist, ob alle Funktionen differenzierbar sind. Anhand des Paradebeispiels der Betragsfunktion

$$f(x) = |x|,$$

findet sich ein Gegenargument für diese Annahme. Wenn der Differentialquotient (1.1) verwendet und die Funktion an der Stelle $x = 0$ untersucht wird, erhält man

$$\lim_{h \to 0} \frac{|0+h| - |0|}{h} = \lim_{h \to 0} \frac{|h|}{h}.$$

Dabei ist zu erkennen, dass für den rechten und linken Grenzwert an der Stelle $x = 0$ unterschiedliche Werte gelten,

$$\lim_{h \to 0^+} \frac{-h}{h} = -1 \quad \text{und} \quad \lim_{h \to 0^-} \frac{h}{h} = 1.$$

Nach der Definition von Differenzierbarkeit müssen die Grenzwerte übereinstimmen. Damit ist die Funktion an der Stelle $x = 0$ nicht differenzierbar.
Ein weiteres Beispiel ist die Funktion

$$g(x) = \begin{cases} (x-2)^2 + 1 & \text{wenn } x \geq 2 \\ 3(-x+2) + 1 & \text{wenn } x < 2 \end{cases}, \tag{1.2}$$

um die Differenzierbarkeit von $g(x)$ an der Stelle $x = 2$ zu überprüfen.
Zunächst werden die linken und rechten Grenzwerte des Differenzenquotienten berechnet.

Für den linken Grenzwert gilt

$$\lim_{h \to 0^-} \frac{g(2+h) - g(2)}{h}.$$

Nach Voraussetzung gilt für $x < 2$, dass $g(x) = 3(-x+2)$. Durch Einsetzen in den Differenzenquotienten erhält man

$$\begin{aligned} g(2+h) &= 3(-(2+h)+2)+1 \\ &= 3(-2-h+2)+1 \\ &= 3(-h)+1 \\ &= -3h+1. \end{aligned}$$

Damit folgt

$$\begin{aligned} \lim_{h \to 0^-} \frac{-3h+1-1}{h} &= \lim_{h \to 0^-} \frac{-3h}{h} \\ &= \lim_{h \to 0^-} -3 \\ &= -3. \end{aligned}$$

Ähnliches Vorgehen gilt für den rechten Grenzwert und man erhält

$$\begin{aligned} \lim_{h \to 0^+} \frac{h^2+1-1}{h} &= \lim_{h \to 0^+} \frac{h^2}{h} \\ &= \lim_{h \to 0^+} h \\ &= 0. \end{aligned}$$

Erkennbar ist, dass der linke Grenzwert -3 und der rechte Grenzwert 0 unterschiedlich sind, wodurch die Funktion $f(x)$ an der Stelle $x = 2$ nicht differenzierbar ist.
Um jedoch Funktionen wie die Betragsfunktion oder Funktionen wie in Beispiel (1.2) dennoch zu "differenzieren", kann der Subgradient herangezogen werden. Subgradienten verallgemeinern das Konzept des Gradienten und ermöglichen die Definition von Richtungsableitungen auch für nicht-glatte [1] Funktionen.
Subgradienten sind besonders in der konvexen Optimierung nützlich, da viele Optimierungsprobleme nicht differenzierbar sind. Das Subgradientenverfahren ermöglicht die iterative Annäherung an ein Optimum, selbst wenn die Zielfunktion an einer gewissen Stelle nicht differenzierbar ist und dadurch Ecken entstehen.

In dieser Arbeit werden die Grundlagen der Subgradienten und die zugrundeliegende Theorie genauer untersucht und ihre Anwendung in der mathematischen Optimierung erläutert.

Diese Arbeit stützt sich auf Definitionen, Sätze, Lemmata, etc. aus der Lektüre [3].

[1] Die Definition von glatten Funktionen beschreibt unendlich oft (stetig) differenzierbare Funktionen.

2 Konvexe Mengen

2.1 Grundlagen

Definition 2.1.1 ((abgeschlossene) Strecke)**.**
Sei $S \subseteq \mathbb{R}^n$, dann wird S als (abgeschlossene) Strecke bezeichnet, falls diese durch

$$S := [x, y] := \{x + \lambda(y - x) | \lambda \in [0, 1]\}$$

mit disjunkten Endpunkten $x, y \in \mathbb{R}^n$ parametrisiert werden kann.

Definition 2.1.2 (konvexe Menge)**.**
Sei $M \subseteq \mathbb{R}^n$, $n \in \mathbb{N}$ mit der Eigenschaft

$$(x + \lambda(y - x)) \in M \ (\textit{Konvexitätsbedingung})$$

für alle $x, y \in M$ und $\lambda \in (0, 1)$, dann heißt die Menge M konvex.

Bemerkung 2.1.3.
Sei $A \subseteq \mathbb{R}^n$ konvex und seien $a, b \in A$ beliebig. Gemäß Definition 2.1.2 ist die Strecke $[a, b]$ ebenfalls in A enthalten, was in Abbildung 2.1 dargestellt ist. Abbildung 2.2 hingegen veranschaulicht, dass die Strecke $[c, d]$ nicht vollständig in der Menge liegt und es sich daher um eine nicht konvexe Menge handelt.

Es ist zu beachten, dass Abbildungen dazu dienen, das Verständnis zu vertiefen und visuelle Intuition zu gewinnen, sie aber allein nicht als Beweis für mathematische Aussagen ausreichen.

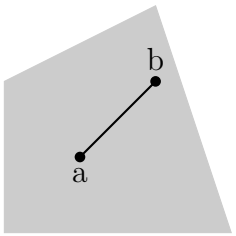

Abbildung 2.1 Eine konvexe Menge

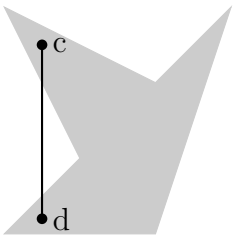

Abbildung 2.2 Eine nicht konvexe Menge

Definition 2.1.4 (Hyperebene und Halbräume)**.**
Eine Hyperebene im $\mathbb{R}^n$ ist eine Menge der Form

$$H := \{x \in \mathbb{R}^n | \langle a, x \rangle = b\} \text{ mit } a \in \mathbb{R}^n \setminus \{0\} \text{ und } b \in \mathbb{R}.$$

Eine Hyperebene H trennt $\mathbb{R}^n$ in zwei (abgeschlossene) Halbräume

$$H_{\leq b} := \{x \in \mathbb{R}^n | \langle a, x \rangle \leq b\} \tag{2.1}$$

und

$$H_{\geq b} := \{x \in \mathbb{R}^n | \langle a, x\rangle \geq b\}, \tag{2.2}$$

die sich in H schneiden.
Die Mengen

$$H_{<b} := \{x \in \mathbb{R}^n | \langle a, x\rangle < b\}, \tag{2.3}$$

$$H_{>b} := \{x \in \mathbb{R}^n | \langle a, x\rangle > b\} \tag{2.4}$$

werden als offene Halbräume bezeichnet.

Lemma 2.1.5.
Jeder Halbraum $H \subset \mathbb{R}^n$ ist konvex.

Beweis. Zu zeigen ist, dass H konvex ist. Somit ist zu zeigen, dass für alle $x, y \in H$ und alle $\lambda \in (0, 1)$ die Strecke $\lambda x + (1 - \lambda)y$ ebenso in H liegt. Sei $x, y \in H$ beliebig, $\lambda \in (0, 1)$ und $k, n \in \mathbb{N}$.
Dann gilt

$$\langle a, x\rangle \leq b \text{ und } \langle a, y\rangle \leq b \text{ mit } a \in \mathbb{R}^n \setminus \{0\} \text{ und } b \in \mathbb{R}. \tag{2.5}$$

Sei z der Punkt $z = \lambda x + (1 - \lambda)y$ mit $\lambda \in (0, 1)$. Damit H konvex ist, muss gezeigt werden, dass $z \in H$, also dass

$$\langle a, z\rangle = \langle a, \lambda x + (1 - \lambda)y\rangle \leq b.$$

Nach Definition des Skalarproduktes gilt

$$\begin{aligned}
\langle a, z\rangle &= \sum_{k=1}^{n} a_k z_k \\
&= \sum_{k=1}^{n} a_k(\lambda x_k + (1 - \lambda)y_k) \\
&= \sum_{k=1}^{n} \lambda a_k x_k + (1 - \lambda)a_k y_k \\
&= \sum_{k=1}^{n} \lambda a_k x_k + \sum_{k=1}^{n}(1 - \lambda)a_k y_k \\
&= \lambda \sum_{k=1}^{n} a_k x_k + (1 - \lambda) \sum_{k=1}^{n} a_k y_k \\
&= \lambda\langle a, x\rangle + (1 - \lambda)\langle a, y\rangle.
\end{aligned} \tag{2.6}$$

Aufgrund von (2.5) resultiert

$$\begin{aligned}
\lambda\langle a, x\rangle + (1 - \lambda)\langle a, y\rangle &\leq \lambda b + (1 - \lambda)b \\
&= \lambda b + b - \lambda b \\
&= b.
\end{aligned}$$

Damit wurde gezeigt, dass

$$\langle a, z\rangle \leq b$$

und damit ist $z \in H$.

Durch die beliebige Wahl von $x, y \in H$ und $\lambda \in (0, 1)$, folgt, dass H konvex ist. □

Bemerkung 2.1.6.
Der Beweis zum Lemma 2.1.5 behandelt nur den (abgeschlossen) Halbraum

$$H_{\leq b} := \{x \in \mathbb{R}^n \mid \langle x, a\rangle \leq b\} \text{ mit } a \in \mathbb{R}^n \setminus \{0\} \text{ und } b \in \mathbb{R}.$$

Dieser Beweis lässt sich auch auf die anderen Halbräume mit den restlichen Vergleichszeichen anwenden. Außerdem wird in (2.6) die Linearität des Skalarproduktes deutlich.

Notation

Das Skalarprodukt kann für alle $x, y \in \mathbb{R}^n$ wie folgt geschrieben werden

$$\langle x, y\rangle = \sum_{k=1}^{n} x_k y_k = x^\top y.$$

Es soll festgestellt werden, ob der abgeschlossene Halbraum (2.1) auch abgeschlossen ist. Zunächst wird gezeigt, dass $\mathbb{R}^n$, ausgestattet mit einer Metrik, ein metrischer Raum ist. Somit kann mittels Folgenkriterium für Abgeschlossenheit [4] gezeigt werden, dass der abgeschlossene Halbraum abgeschlossen ist.

Lemma 2.1.7.
Die Menge $\mathbb{R}^n$ mit einer Metrik d ist ein metrischer Raum.

Beweis. Es soll gezeigt werden, dass für $\mathbb{R}^n$ die Metrik-Eigenschaften gelten.
Sei hierzu

$$d(x, y) := \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

eine Metrik auf $\mathbb{R}^n$ und $x, y \in \mathbb{R}^n$. Genauer gesagt, handelt es sich hierbei um die euklidische Metrik. Sei also die euklidische Metrik auf $\mathbb{R}^n$ und $x, y, z \in \mathbb{R}^n$ gegeben. Als Erstes wird gezeigt, dass die positive Definitheit gilt. Es gilt $d(x, y)$ ist genau dann gleich Null, wenn die beiden Vektoren identisch sind. Das bedeutet, wenn $x = y$, dann ist

$$d(x, y) = d(x, x) = \sqrt{\sum_{i=1}^{n}(x_i - x_i)^2} = \sqrt{\sum_{i=1}^{n} 0} = 0.$$

Für $x \neq y$ gilt, dass mindestens ein Term in der Summe ungleich Null sein muss. Wäre dies nicht erfüllt, ergäbe sich der oben genannte Fall $x = y$. Daher ist $d(x, y) > 0$.

Weiter ist die Metrik symmetrisch, denn es gilt

$$d(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} = \sqrt{\sum_{i=1}^{n}(y_i - x_i)^2} = d(y, x).$$

Letztlich ist zu zeigen, dass die Dreiecksungleichung erfüllt ist,

$$d(x, z) \leq d(x, y) + d(y, z).$$

Die Minkowski-Ungleichung [5] besagt

$$\sqrt{\sum_{i=1}^{n}(a_i+b_i)^2} \leq \sqrt{\sum_{i=1}^{n}a_i^2} + \sqrt{\sum_{i=1}^{n}b_i^2}.$$

Seien $a_i = x_i - y_i$ und $b_i = y_i - z_i$.
Dann gilt für die quadrierte linke Seite der Ungleichung

$$\sum_{i=1}^{n}((x_i-y_i)+(y_i-z_i))^2 = \sum_{i=1}^{n}(x_i-z_i)^2.$$

Die Minkowski-Ungleichung liefert

$$\sqrt{\sum_{i=1}^{n}(x_i-z_i)^2} \leq \sqrt{\sum_{i=1}^{n}(x_i-y_i)^2} + \sqrt{\sum_{i=1}^{n}(y_i-z_i)^2}.$$

Dadurch gilt die Dreiecksungleichung

$$d(x,z) \leq d(x,y) + d(y,z).$$

Die Menge $\mathbb{R}^n$ mit der Metrik d erfüllt die oben genannten Eigenschaften und ist damit ein metrischer Raum. □

Korollar 2.1.8.
Der abgeschlossene Halbraum (2.1) ist abgeschlossen.

Beweis. Um zu zeigen, dass ein abgeschlossener Halbraum abgeschlossen ist, wird das Folgenkriterium für Abgeschlossenheit verwendet. Eine Menge A in einem metrischen Raum ist genau dann abgeschlossen, wenn jede konvergente Folge $(x_n)_{n\in\mathbb{N}}$ in A den Grenzwert $\lim\limits_{n\to\infty} x_n$ ebenfalls in A hat.
Nun wird ein Halbraum $H_{\leq b}$ betrachtet, der durch die Ungleichung

$$\langle a, x\rangle = \sum_{i=1}^{n} a_i x_i \leq b$$

dargestellt wird. Angenommen, $(a_n, x_n)_{n\in\mathbb{N}}$ ist eine konvergente Folge in $H_{\leq b}$. Das bedeutet, dass

$$(a_n, x_n) \to (a^*, x^*) \text{ für } n \to \infty.$$

Sei (a^*, x^*) der Grenzwert.
Durch die Tatsache, dass $(a_n, x_n)_{n\in\mathbb{N}}$ gegen (a^*, x^*) konvergiert, konvergieren auch die Komponenten a_n und x_n gegen a^* und x^*, d.h.,

$$a_n \to a^* \quad \text{und} \quad x_n \to x^* \text{ für } n \to \infty.$$

Da jedes $(a_n, x_n)_{n\in\mathbb{N}}$ in $H_{\leq b}$ liegt, gilt für alle $n \in \mathbb{N}$

$$\langle a_n, x_n\rangle \leq b.$$

Das Skalarprodukt $\langle \cdot, \cdot \rangle$ ist eine stetig definierte Funktion und damit gilt

$$\lim_{n\to\infty} \langle a_n, x_n \rangle = \langle a^*, x^* \rangle.$$

Aus der Stetigkeit des Grenzwertes und dadurch, dass $\langle a_n, x_n \rangle \leq b$ für alle $n \in \mathbb{N}$ ist, folgt

$$\langle a^*, x^* \rangle \leq b.$$

Da $\langle a^*, x^* \rangle \leq b$, liegt der Grenzwert (a^*, x^*) ebenfalls im Halbraum $H_{\leq b}$. Das bedeutet, dass $H_{\leq b}$ abgeschlossen ist, da jede konvergente Folge in $H_{\leq b}$ ihren Grenzwert ebenfalls in $H_{\leq b}$ hat.
Damit wurde gezeigt, dass der Halbraum $H_{\leq b}$, der durch die Ungleichung $\langle a, x \rangle \leq b$ definiert ist, abgeschlossen ist. □

Bemerkung 2.1.9.
Aus dem Korollar 2.1.8 zeigt sich, dass mit einem ähnlichen Argument und umgekehrten Vergleichszeichen die Menge (2.2) ebenfalls abgeschlossen ist und das jeweilige Komplement (2.3) und (2.4) somit offen ist.

Bemerkung 2.1.10.
Halbräume geben wichtige Beispiele für konvexe Mengen, denn sie haben eine einfache Struktur und sind ohne großen Aufwand zu erkennen.

Um das Verständnis für konvexe Mengen weiter zu vertiefen, werden im Folgenden Beispiele von konvexen Mengen betrachtet.

Beispiele 2.1.11.

i.) Ein Intervall bzw. eine Strecke auf der reellen Zahlengerade ist konvex.
Das Intervall $[a, b]$ mit $a, b \in \mathbb{R}$ ist beliebig gewählt und konvex und $a \leq b$, da für alle $x, y \in [a, b]$ und $\lambda \in [0, 1]$, ist $\lambda x + (1 - \lambda)y$ ebenso in $[a, b]$

ii.) Die Strecke $S \subseteq \mathbb{R}^n$ ist konvex, denn sie enthält die Verbindungsstrecke zwischen Anfangs- und Endpunkt

iii.) Ein Würfel im dreidimensionalen Raum ist konvex, da die Strecke zwischen zwei beliebigen Punkten innerhalb des Würfels ebenfalls vollständig innerhalb des Würfels verläuft. Die Flächen und Kanten des Würfels dienen dabei als Begrenzungen von Strecken. Dies entspricht genau der Definition von Konvexität

2.2 Eigenschaften von konvexen Mengen

Im Folgenden werden einige, für das Verständis entscheidende, Eigenschaften von konvexen Mengen diskutiert und bewiesen. Die Beweise werden dabei systematisch und präzise durchgeführt, um fundierte Kenntnisse für die Eigenschaften der konvexen Mengen zu vermitteln.

Satz 2.2.1.
Seien $C_1, C_2, \ldots \subset \mathbb{R}^n$ beliebige konvexe Mengen, dann ist $\bigcap_{i=1}^{\infty} C_i$ wieder konvex.

Beweis. Wähle $x, y \in \tilde{C} := \bigcap_{i=1}^{\infty} C_i$ und sei $\lambda \in [0,1]$.
Es soll die Eigenschaft aus Definiton 2.1.2 gezeigt werden, d. h.

$$(1-\lambda)x + \lambda y \in \tilde{C}. \tag{2.7}$$

Für alle $i \in \{1, \ldots, \infty\}$ sind $x, y \in C_i$, aufgrund der Definiton des Schnittes.
Weiter ist die Menge C_i konvex. Dies gilt für alle $i \in \{1, \ldots, \infty\}$ nach Voraussetzung.
Daher gilt

$$(1-\lambda)x + \lambda y \in C_i, \ \forall i \in \{1, \ldots, \infty\}.$$

Es wird $\tilde{\lambda} := (1-\lambda)x + \lambda y$ gesetzt.
Damit zeigt sich, dass bei Bildung der Vereinigung über alle $i \in \{1, \ldots, \infty\}$ das Element $\tilde{\lambda}$ in jedem C_i enthalten ist, d. h.

$$\tilde{\lambda} \in \tilde{C}.$$

Dadurch wurde (2.7) nachgewiesen und damit auch Satz 2.2.1 gezeigt. □

Bemerkung 2.2.2.
Sich nun die Frage zu stellen, ob die unendliche oder die endliche Vereinigung von konvexen Mengen wieder konvex ist, ist eine plausible Überlegung.
Im Allgemeinen trifft diese Aussage jedoch nicht zu.

Es lässt sich leicht ein Beispiel konstruieren, indem zwei beliebige konvexe Mengen $C_1, C_2 \subset \mathbb{R}^n$ gewählt werden, die nicht überlappen, d. h. $C_1 \cap C_2 = \emptyset$, Abbildung 2.3 stellt dies dar.

Angenommen, $x \in C_1$ und $y \in C_2$. In diesem Fall wird deutlich, dass die *Konvexitätsbedingung* gemäß Definition 2.1.2 nicht erfüllt ist. Das heißt, wenn ein Punkt $z = \lambda x + (1-\lambda)y$ mit $\lambda \in [0,1]$ betrachtet wird, so liegt z im Allgemeinen nicht in $C_1 \cup C_2$, da z nicht notwendigerweise in C_1 oder in C_2 liegt (vgl. Abb. 2.3).

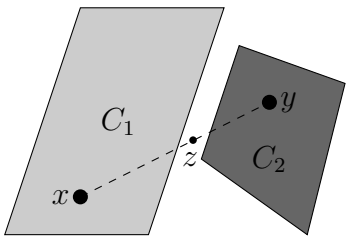

Abbildung 2.3 Punkt z außerhalb zweier konvexer Mengen

Korollar 2.2.3.
Sei I eine beliebige Indexmenge. Weiter sei $y_i \in \mathbb{R}^n$ und $\beta_i \in \mathbb{R}$ für $i \in I$.
Dann ist die Menge

$$K := \{x \in \mathbb{R}^n \mid \langle x, y_i \rangle \leq \beta_i, \ \textit{für alle } i \in I\}$$

konvex.

Beweis. Seien $x_1, x_2 \in K$. Dann gilt für alle $i \in I$

$$\langle x_1, y_i \rangle \leq \beta_i,$$
$$\langle x_2, y_i \rangle \leq \beta_i.$$

Es soll gezeigt werden, dass auch $\lambda x_1 + (1-\lambda)x_2 \in K$ für alle $\lambda \in [0,1]$. Sei also $\lambda \in [0,1]$ beliebig. Dann gilt für alle $i \in I$ und durch die Linearität des Skalarproduktes

$$\begin{aligned}\langle \lambda x_1 + (1-\lambda)x_2, y_i \rangle &= \lambda \langle x_1, y_i \rangle + (1-\lambda)\langle x_2, y_i \rangle \\ &\leq \lambda\beta_i + (1-\lambda)\beta_i \\ &= \lambda\beta_i + \beta_i - \lambda\beta_i \\ &= \beta_i.\end{aligned}$$

Somit ist K nach Satz 2.2.1 als Schnitt konvexer Mengen selbst konvex.

Angenommen die Menge K ist leer, dann ist sie konvex. Ist K also leer, dann existieren keine Punkte in K. Demnach ist die Aussage des Korollars trivialerweise erfüllt, da die Bedingung $\lambda x_1 + (1-\lambda)x_2 \in K$ für alle $x_1, x_2 \in K$ zutrifft. □

Bemerkung 2.2.4.
Die Aussage des Korollars 2.2.3 bleibt weiterhin gültig, auch wenn die Ungleichung mit anderen Vergleichszeichen vertauscht wird. Eine weitere Möglichkeit das Korollar 2.2.3 zu beweisen ist, ein beliebiges $i \in I$ festzuhalten und aufgrund der Struktur ist dies ein Halbraum. Wie bereits in Lemma 2.1.5 bewiesen, sind Halbräume konvex und mit Korollar 2.2.1 gilt dies auch für den unendlichen Schnitt von konvexen Mengen, wie im Korollar dargestellt.

Definition 2.2.5 (Polyedrisch konvex)**.**
Eine Menge im $\mathbb{R}^n$ heißt polyedrisch konvex, wenn sie der Schnitt einer endlichen Anzahl von abgeschlossenen Halbräumen ist.

Beispiel 2.2.6.

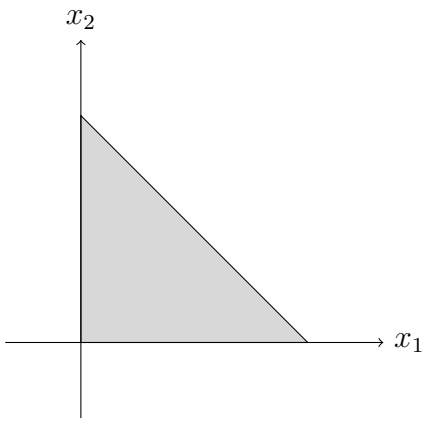

Abbildung 2.4 Eine polyedrisch konvexe Menge im $\mathbb{R}^2$

Die Abbildung 2.4 zeigt ein Dreieck in der Ebene, das durch die Bedingungen $x_1 \geq 0$, $x_2 \geq 0$ und $x_1 + x_2 \leq 1$ definiert ist.
Betrachtet wird folgende Menge

$$\{x \in \mathbb{R}^n \mid x_1 \geq 0, x_2 \geq 0, x_1 + x_2 \leq 1\}. \tag{2.8}$$

Diese Menge ist der Schnitt der abgeschlossenen Halbräume

$$\{x \in \mathbb{R}^n \mid x_1 \geq 0\},$$
$$\{x \in \mathbb{R}^n \mid x_2 \geq 0\},$$
$$\{x \in \mathbb{R}^n \mid x_1 + x_2 \leq 1\},$$

wodurch ist die Menge (2.8) polyedrisch konvex ist.

Definition 2.2.7 (Konvexkombination)**.**
Die Linearkombination mit $x \in \mathbb{R}^n$ ist gegeben durch

$$x = \lambda_1 x_1 + \lambda_2 x_2 + \ldots + \lambda_n x_n.$$

Dies wird als Konvexkombination bezeichnet, wenn die Koeffizienten $\lambda_i \in [0, 1]$ für alle $i \in \{1, \ldots, n\}$ und

$$\sum_{i=1}^{n} \lambda_i = 1$$

gilt.

Die Definition 2.2.7 ist ein neues Werkzeug, um gegebene Mengen in $\mathbb{R}^n$ auf Konvexität zu überprüfen. Diese Charakterisierung ermöglicht es, die Konvexität einer Menge anhand der Eigenschaft, alle konvexen Kombinationen ihrer Elemente zu enthalten, zu bestimmen. Dies wird im folgenden Satz nochmals deutlich.

Satz 2.2.8.
Sei $A \subseteq \mathbb{R}^n$. A ist genau dann konvex, wenn es alle konvexen Kombinationen seiner Elemente enthält.

Beweis. Sei dazu K konvex. Dann gilt für $\lambda \in [0, 1]$ und $a, b \in K$

$$\lambda a + (1 - \lambda) b \in A.$$

Für die Koeffizienten gilt

$$\lambda + (1 - \lambda) = 1.$$

Setze zur besseren Sichtbarkeit $\lambda_1 := \lambda$ und $\lambda_2 := (1 - \lambda)$.
Somit ist

$$\lambda_1 a + \lambda_2 b \in K.$$

Dies zeigt beide Richtungen für zwei Vektoren. Die Konvexität von A besagt, dass A abgeschlossen ist unter der Bildung von konvexen Kombinationen mit $m = 2$.
Um die Aussage für $m > 2$ zu zeigen, wird ein Induktionsbeweis durchgeführt.
Sei $m > 2$.
Angenommen sei eine konvexe Kombination $a = \lambda_1 a_1 + \ldots + \lambda_m a_m$ mit $a_i \in K$ für alle $i \in \{1, \ldots, m\}$ und gewählt sei λ_1 (hier kann ein beliebiges $i \in \{1, \ldots, m\}$ gewählt werden).
Für die Koeffizienten der konvexen Kombination a gilt dann

$$\begin{aligned} 1 &= \lambda_1 + \lambda_2 + \ldots + \lambda_m \\ \Leftrightarrow 1 - \lambda_1 &= \lambda_2 + \ldots + \lambda_m \\ \Leftrightarrow 1 &= \frac{\lambda_2}{1 - \lambda_1} + \ldots + \frac{\lambda_m}{1 - \lambda_1}. \end{aligned}$$

Definiere $\alpha_i := \frac{\lambda_i}{1-\lambda_1}$ für alle $i \in \{2, 3, \ldots, m\}$ und erhalte

$$1 = \alpha_2 + \ldots + \alpha_m. \tag{2.9}$$

Sei also

$$b = \alpha_2 a_2 + \ldots + \alpha_m a_m.$$

Durch (2.9) ist b eine konvexe Kombination von $m-1$ Vektoren in K, und daher ist $b \in K$ per Induktionsannahme.
Somit ist

$$a = (1 - \lambda_1)b + \lambda_1 a_1.$$

Die Schlussfolgerung ist, dass $a \in K$. Damit ist die Aussage gezeigt. □

Definition 2.2.9 (Konvexe Hülle)**.**
Die konvexe Hülle einer Menge $S \subseteq \mathbb{R}^n$, bezeichnet als $\text{conv}(S)$, ist definiert als die kleinste konvexe Menge, die S enthält. Formal ist die konvexe Hülle von S gegeben durch

$$\text{conv}(S) = \bigcap \{K \subseteq \mathbb{R}^n \mid S \subseteq K, K \text{ ist konvex}\}.$$

Lemma 2.2.10.
Die Menge conv(S) ist konvex.

Beweis. Das Lemma 2.2.10 ist nach Satz 2.2.1 konvex. □

Die folgenden Beispiele sollen dabei unterstützen, sich ein konkreteres Bild der konvexen Hülle machen zu können.

Beispiele 2.2.11.

i.) Die konvexe Hülle zweier Punkte $a, b \in \mathbb{R}$ ist ihre Strecke

$$\text{conv}(\{a, b\}) = [a, b] := \{a + \lambda(b - a) | \lambda \in [0, 1]\}$$

ii.) In der Abbildung 2.5 zeigt sich die konvexe Hülle $\text{conv}(\{X, Y, Z, H, G\})$

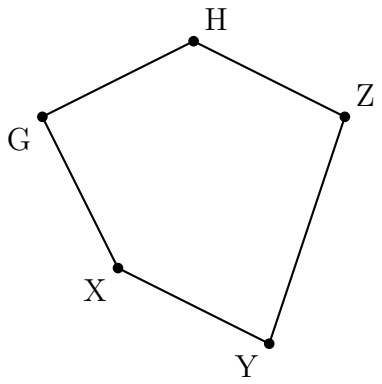

Abbildung 2.5 Konvexe Hülle von Punkten im $\mathbb{R}^2$

iii.) Die konvexe Hülle von endlich vielen Punkten wird als Polytop bezeichnet. Ein elementares Beispiel für ein Polytop ist der d-dimensionale Standard-Simplex in $\mathbb{R}^{d+1}$

$$\Delta_d = \left\{ (\alpha_1, \ldots, \alpha_{d+1}) \in \mathbb{R}^{d+1} \mid \sum_{i=1}^{d+1} \alpha_i = 1 \text{ und } \alpha_i \geq 0 \text{ für } i = 1, \ldots, d+1 \right\}$$

Dies ist die konvexe Hülle der $d+1$ Standardbasisvektoren in $\mathbb{R}^{d+1}$ (vgl. Abb. 2.6)

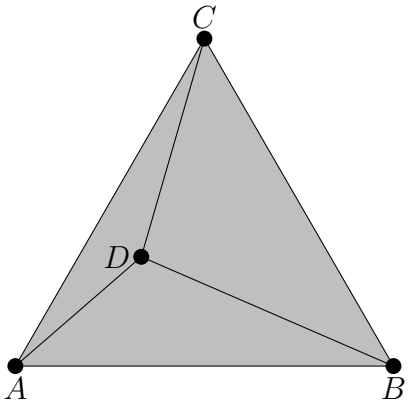

Abbildung 2.6 Standard-Simplex Δ_3 in $\mathbb{R}^{3+1}$

iv.) Ein (kompakter) Polyeder ist eine geometrische Form, die durch gerade Linien begrenzt wird. Nicht nur kann er als der Schnitt von endlich vielen Halbräumen beschrieben werden, alternativ können Polyeder auch aus Konvexkombinationen endlich vieler Punkte und nicht-negativer Linearkombinationen endlich vieler Richtungen erzeugt werden (vgl. [6]). Dadurch, dass Halbräume konvex sind (siehe Lemma 2.1.5), sind insbesondere Polyeder konvex. Die Halbräume werden durch je eine lineare Ungleichung definiert, so kann ein Polyeder P durch die Bedingungen

$$P = \{x \in \mathbb{R}^d \mid \ell_1(x) \leq \beta_1, \ldots, \ell_k(x) \leq \beta_k\}$$

beschrieben werden, wobei $\ell_1, \ldots, \ell_k$ Linearformen und $\beta_1, \ldots, \beta_k$ reelle Zahlen sind

3 Konvexe Funktionen

Konvexe Mengen und konvexe Funktionen sind eng miteinander verbunden. Nachdem im vorherigen Kapitel bereits konvexe Mengen analysiert wurden, wird sich in diesem Abschnitt konvexen Funktionen zugewandt. Eine konvexe Funktion ist eine Funktion, bei der die Strecke zwischen zwei Punkten des Graphen der Funktion in der Menge über dem Graphen liegt. Dies wird im Folgenden genauer erläutert.

Dazu wird die Zwei-Punkt-Kompaktifizierung genutzt, die wie folgt definiert ist

$$\overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\} \cup \{-\infty\}.$$

Entsprechend sei

$$\mathcal{F} := \mathcal{F}(X, \overline{\mathbb{R}}),$$

wobei $X \subseteq \mathbb{R}^n$, d.h, für $f \in \mathcal{F}$ gilt $f \colon X \to \overline{\mathbb{R}}$.

3.1 Grundlegende Definitionen

Um konvexe Funktionen präzise zu definieren, ist es notwendig, eine Beschreibung der Menge oberhalb der Funktion einzuführen. Diese Menge, bekannt als Epigraph der Funktion, liefert eine Grundlage für die Definition der Konvexität und ermöglicht es, die Eigenschaften der Funktion zu untersuchen.

Definition 3.1.1 (Epigraph)**.**
Sei $f \in \mathcal{F}$. Die Menge

$$\operatorname{epi}(f) := \{(x, \mu) \in X \times \overline{\mathbb{R}} \mid f(x) \leq \mu\}$$

wird als Epigraph bezeichnet.

Definition 3.1.2 (Effektiver Definitionsbereich)**.**
Die durch

$$\operatorname{dom}(f) = \{x \in X \mid f(x) < +\infty\}$$

definierte Teilmenge von X heißt effektiver Definitionsbereich der Funktion f.

Beispiel 3.1.3.
Der Epigraph der Funktion $f(x) = x^2$ ist in der Abbildung 3.1 mit grauer Schraffierung dargestellt. Die Schraffierung markiert den Bereich oberhalb der Funktionskurve, der durch die Funktion begrenzt wird.

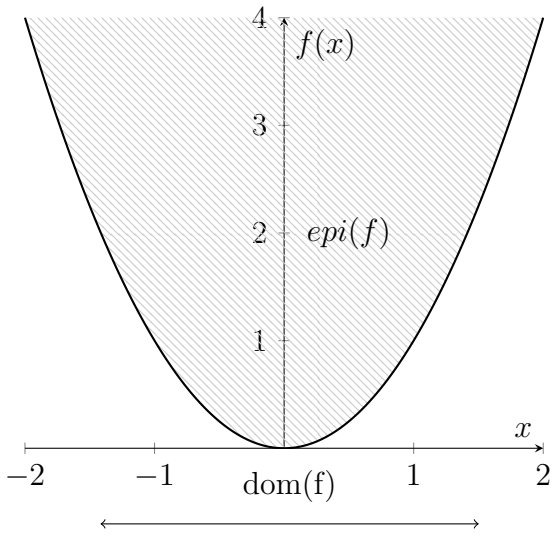

Abbildung 3.1 Epigraph der Funktion $f(x) = x^2$

Definition 3.1.4 (Konvexe Funktion)**.**
Eine Funktion f aus der Menge der Funktionen $\mathcal{F}$ ist konvex, wenn ihr Epigraph konvex ist.

Beispiel 3.1.5.
Die Behauptung ist, dass die Funktion $f(x) = x^2$, die in der Abbildung 3.1 dargestellt ist, konvex ist. Dies soll anhand der Definition 3.1.4 nachgewiesen werden.

Beweis. Es muss gezeigt werden, dass der Epigraph von f, definiert als

$$\text{epi}(f) = \{(x, \mu) \in X \times \overline{\mathbb{R}} \mid x^2 \leq \mu\},$$

konvex ist.
Seien $(x_1, \mu_1), (x_2, \mu_2) \in \text{epi}(f)$. Dann gilt

$$\mu_1 \geq f(x_1) = x_1^2 \quad \text{und} \quad \mu_2 \geq f(x_2) = x_2^2.$$

Für $\lambda \in [0, 1]$ wird die konvexe Kombination

$$\lambda \begin{pmatrix} x_1 \\ \mu_1 \end{pmatrix} + (1 - \lambda) \begin{pmatrix} x_2 \\ \mu_2 \end{pmatrix} = \begin{pmatrix} \lambda x_1 + (1 - \lambda) x_2 \\ \lambda \mu_1 + (1 - \lambda) \mu_2 \end{pmatrix}$$

betrachtet. Gezeigt werden soll, dass

$$\begin{pmatrix} \lambda x_1 + (1 - \lambda) x_2 \\ \lambda \mu_1 + (1 - \lambda) \mu_2 \end{pmatrix} \in \text{epi}(f).$$

Das bedeutet, es muss überprüft werden, ob

$$f(\lambda x_1 + (1 - \lambda) x_2) = (\lambda x_1 + (1 - \lambda) x_2)^2 \leq \lambda \mu_1 + (1 - \lambda) \mu_2.$$

Allgemein kann an dieser Stelle festgehalten werden, dass immer die Gültigkeit von

$$f(\lambda x_1 + (1 - \lambda) x_2) \leq \lambda \mu_1 + (1 - \lambda) \mu_2 \tag{3.1}$$

gezeigt werden muss.

Durch Ausmultiplizieren der linken Seite der Ungleichung erhält man

$$(\lambda x_1 + (1-\lambda)x_2)^2 = \lambda^2 x_1^2 + 2\lambda(1-\lambda)x_1x_2 + (1-\lambda)^2 x_2^2.$$

Es wird festgehalten, dass nach Voraussetzung $x_1^2 \leq \mu_1$ und $x_2^2 \leq \mu_2$ gilt und man

$$\lambda x_1^2 + (1-\lambda)x_2^2 \leq \lambda\mu_1 + (1-\lambda)\mu_2$$

erhält. Das Ziel ist, zu zeigen, dass

$$\lambda^2 x_1^2 + 2\lambda(1-\lambda)x_1x_2 + (1-\lambda)^2 x_2^2 \leq \lambda x_1^2 + (1-\lambda)x_2^2. \tag{3.2}$$

Die Ungleichung ist erfüllt, denn beim Subtrahieren der linken Seite von der rechten Seite folgt

$$\begin{aligned}
0 &\leq x_1^2 + (1-\lambda)x_2^2 - \lambda^2 x_1^2 + 2\lambda(1-\lambda)x_1x_2 + (1-\lambda)^2 x_2^2 \\
&= \lambda x_1^2 - \lambda^2 x_1^2 + (1-\lambda)x_2^2 - (1-\lambda)^2 x_2^2 - 2\lambda(1-\lambda)x_1x_2 \\
&= \lambda(1-\lambda)x_1^2 + (1-\lambda)(1-(1-\lambda))x_2^2 - 2\lambda(1-\lambda)x_1x_2 \\
&= \lambda(1-\lambda)x_1^2 + (1-\lambda)\lambda x_2^2 - 2\lambda(1-\lambda)x_1x_2 \\
&= \lambda(1-\lambda)(x_1^2 + x_2^2 - 2x_1x_2) \\
&= \lambda(1-\lambda)(x_1 - x_2)^2.
\end{aligned} \tag{3.3}$$

Damit die Ungleichung (3.2) erfüllt und die Aussage somit gültig ist, muss der Ausdruck (3.3) größer oder gleich Null sein.
Dies ergibt sich aus der Überlegung, dass λ im Intervall $[0,1]$ liegt, weshalb $1-\lambda \geq 0$ gilt. Zusätzlich ist das Quadrieren einer reellen Zahl stets nicht-negativ, also $(x_1 - x_2)^2 \geq 0$.
Somit ist bewiesen, dass

$$\lambda(1-\lambda)(x_1 - x_2)^2 \geq 0.$$

Diese Ungleichung zeigt, dass (3.2) tatsächlich erfüllt ist und daher gilt

$$\begin{pmatrix} \lambda x_1 + (1-\lambda)x_2 \\ \lambda\mu_1 + (1-\lambda)\mu_2 \end{pmatrix} \in \operatorname{epi}(f),$$

was die Konvexität des Epigraphen beweist. Somit ist die Funktion $f(x) = x^2$ konvex. □

Eine nicht konvexe Funktion ist z. B. die Funktion $g(x) = -x^3 + 2x^2 - 1$.
Nun soll gezeigt werden, dass

$$\operatorname{epi}(g) = \{(x,\mu) \in X \times \overline{\mathbb{R}} \mid -x^3 + 2x^2 - 1 \leq \mu\}$$

nicht konvex ist.

Beweis. Angenommen, $\operatorname{epi}(g)$ ist konvex.
Es wird $x_1 = 0$, $x_2 = \frac{3}{2}$ und $\lambda = \frac{1}{10}$ gewählt und dies in die Definition der Konvexität eingesetzt.
Für die linke Seite der Ungleichung gilt

$$g\left(\frac{1}{10}0 + \left(1 - \frac{1}{10}\right)\frac{3}{2}\right) = g\left(\frac{27}{20}\right).$$

Durch Berechnung von $g\left(\frac{27}{20}\right)$ ergibt sich

$$\begin{aligned} g\left(\frac{27}{20}\right) &= -\left(\frac{27}{20}\right)^3 + 2\left(\frac{27}{20}\right)^2 - 1 \\ &= -\frac{19683}{8000} + \frac{1458}{400} - 1 \\ &= -2.460375 + 3.645 - 1 \\ &= 0.184625. \end{aligned}$$

Für die rechte Seite gilt

$$\lambda g(x_1) + (1-\lambda) g(x_2) = \frac{1}{10} g(0) + \frac{9}{10} g\left(\frac{3}{2}\right).$$

Berechnet wird $g(0)$ und $g\left(\frac{3}{2}\right)$

$$g(0) = -0^3 + 20^2 - 1 = -1, \tag{3.4}$$

$$\begin{aligned} g\left(\frac{3}{2}\right) &= -\left(\frac{3}{2}\right)^3 + 2\left(\frac{3}{2}\right)^2 - 1 \\ &= -\frac{27}{8} + 2\frac{9}{4} - 1 \\ &= -\frac{27}{8} + \frac{18}{4} - 1 \\ &= -\frac{27}{8} + \frac{36}{8} - 1 \\ &= \frac{9}{8} - 1 \\ &= -\frac{1}{8}. \end{aligned} \tag{3.5}$$

In die rechte Seite der Ungleichung werden die Resultate von (3.4) und (3.5) eingesetzt,

$$\begin{aligned} \frac{1}{10} g(0) + \frac{9}{10} g\left(\frac{3}{2}\right) &= \frac{1}{10}(-1) + \frac{9}{10}\left(-\frac{1}{8}\right) \\ &= -\frac{1}{10} - \frac{9}{80} \\ &= -\frac{8}{80} - \frac{9}{80} \\ &= -\frac{17}{80} \\ &= -0.2125. \end{aligned}$$

Ein Vergleich der beiden Seiten zeigt

$$0.184625 \not\leq -0.2125.$$

Damit ist der Epigraph und ebenso die Funktion $g(x)$ nicht konvex. □

In der Abbildung 3.2 wird verdeutlicht, dass bei der Wahl der Punkte und einem $\lambda \in [0,1]$ die Strecke nicht vollständig im Epigraphen der Funktion $g(x)$ enthalten ist.

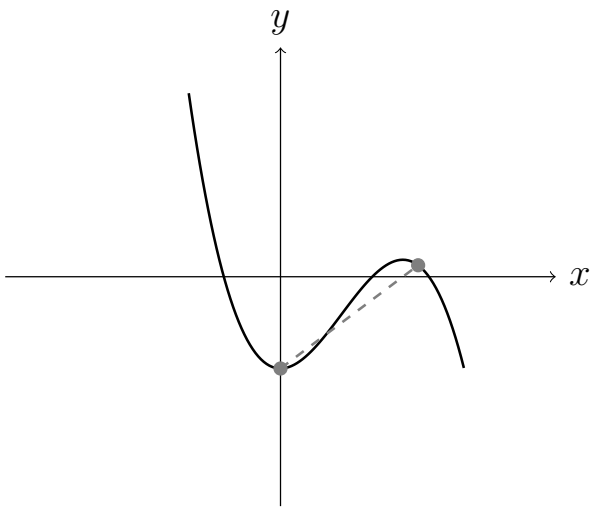

Abbildung 3.2 Nicht Konvexität der Funktion $g(x)$

Damit liegt ein Teil der Strecke im sogennanten *Hypographen*. Diese Tatsache führt zu weiteren Definitionen.

Definition 3.1.6 (Hypograph)**.**
Sei $f \in \mathcal{F}$. Die Menge

$$\text{hypo}(f) := \{(x, \mu) \in X \times \overline{\mathbb{R}} \mid f(x) \geq \mu\}$$

wird als Hypograph bezeichnet.

Definition 3.1.7 (Konkave Funktion)**.**
Eine Funktion f aus der Menge der Funktionen $\mathcal{F}$ ist konkav, wenn ihr Hypograph konvex ist oder wenn der Epigraph von $-f$ konvex ist.

Bemerkung 3.1.8.
Es existieren keine konkaven Mengen. Nur durch die Nicht-Konvexität einer Menge, ist diese nicht gleich konkav.

Beispiele 3.1.9.
Durch das Betrachten der folgenden Funktionen wird das Bild von konkaven und konvexen Funktionen verdeutlicht und gegenübergestellt.

i.) $f(x) = \exp(ax)$ ist eine konvexe Funktion auf $\mathbb{R}$, für alle $a \in \mathbb{R}$

ii.) $g(x) = x^a$ ist eine konvexe Funktion auf $\mathbb{R}_{>0}$, wenn $a \geq 1$ oder $a \leq 0$. Wenn $a \in [0, 1]$ ist $g(x)$ konkav

iii.) $h(x) = \log(x)$ ist konkav auf $\mathbb{R}_{>0}$

Diese Funktionen können auf ähnliche Weise auf Konvexität bzw. Konkavität geprüft werden, wie im gezeigten Beispiel 3.1.5.

Bemerkung 3.1.10.
Die Konvexität einer Funktion f ist äquivalent zur Konvexität der Einschränkung von f auf den effektiven Definitionsbereich $\text{dom}(f)$. Die Erweiterung von Funktionen auf den gesamten Raum $\mathbb{R}^n$ bietet Vorteile [7], wie

i.) Einheitliche Definition der Funktion ohne spezielle Bedingungen für den effektiven Definitionsbereich

ii.) Konsistente Analyse der Funktion überall im Raum $\mathbb{R}^n$

iii.) Klarheit und Konsistenz in der Definition und Analyse von Funktionen

iv.) Vermeidung von Spezialfällen, die beim Beschränken auf den effektiven Definitionsbereich auftreten könnten

Diese Erweiterung vereinfacht die Verarbeitung und Analyse konvexer Funktionen.

Bemerkung 3.1.11.
Durch die Wahl dieses Ansatzes wird der allgemeine arithmetische Umgang mit $\pm\infty$ vorausgesetzt.

Definition 3.1.12.
Sei $f\colon \mathbb{R}^n \to \mathbb{R}$ eine Funktion. Diese heißt positiv homogen, wenn für alle $x \in \mathbb{R}^n$

$$f(\lambda x) = \lambda f(x)$$

mit $\lambda > 0$ gilt.

Satz 3.1.13.
Eine Funktion $f\colon \mathbb{R}^n \to (-\infty, +\infty]$, die positiv homogen ist, ist konvex genau dann, wenn

$$f(x_1 + x_2) \leq f(x_1) + f(x_2)$$

für alle $x_1, x_2 \in \mathbb{R}^n$.

Beweis. Ein detaillierter Beweis ist zu finden in [8]. □

Definition 3.1.14 ((un-)echte konvexe Funktion)**.**
Sei $f \in \mathcal{F}$, wenn f eine konvexe Funktion ist, dann heißt f echt, wenn $\operatorname{epi}(f) \neq \emptyset$,

$$f(x) > -\infty \text{ für alle } x \in X$$

und ein $x_0 \in X$ existiert, sodass

$$f(x_0) < +\infty.$$

Außerdem ist $\operatorname{dom}(f) \neq \emptyset$. Konvexe Funktionen, die nicht echt sind, heißen unechte konvexe Funktionen.

Bemerkung 3.1.15.
Das Studienobjekt sind die echten konvexen Funktionen. Jedoch entstehen aus echten Funktionen in vielen natürlichen Situationen unechte Funktionen und es ist praktikabler, sie zuzulassen, als sie mühsam von der Betrachtung auszuschließen. Um auch hierführ ein kurzes Beispiel zu liefern, das nicht lediglich identisch mit $\pm\infty$ ist, kann die Funktion f auf $\mathbb{R}$ definiert betrachtet werden,

$$f(x) = \begin{cases} -\infty, & \text{wenn } |x| < 1, \\ 0, & \text{wenn } |x| = 1, \text{ (vgl. [9])}. \\ -\infty, & \text{wenn } |x| > 0 \end{cases}$$

3.2 Charakterisierung konvexer Funktionen

Wie durch das vorangegangene Beispiel 3.1.5 deutlich wird, gestaltet sich die Ermittlung von konvexen Funktionen anhand des Epigraphen recht rechenintensiv. Daher sollen nun gewisse Sätze eingeführt werden, um diese Analyse zu beschleunigen.

Satz 3.2.1.
Sei $C \subseteq \mathbb{R}^n$ und f wie folgt definiert

$$f\colon C \to (-\infty, \infty].$$

Dann ist f konvex genau dann, wenn

$$f(\lambda y_1 + (1-\lambda)y_2) \leq \lambda f(y_1) + (1-\lambda)f(y_2) \tag{3.6}$$

für alle $y_1, y_2 \in C$ und $\lambda \in (0,1)$ gilt[2].
Die Funktion f heißt strikt konvex genau dann, wenn

$$f(\lambda y_1 + (1-\lambda)y_2) < \lambda f(y_1) + (1-\lambda)f(y_2) \tag{3.7}$$

für alle $y_1, y_2 \in C$ und $\lambda \in (0,1)$ gilt.

Beweis. Gezeigt wird (3.6), da analoges Vorgehen bei (3.7).
Um zu zeigen, dass die Funktion $f : C \to (-\infty, \infty]$ konvex ist genau dann, wenn die Ungleichung

$$f(\lambda y_1 + (1-\lambda)y_2) \leq \lambda f(y_1) + (1-\lambda)f(y_2)$$

für alle $y_1, y_2 \in C$ und $\lambda \in (0,1)$ gilt, wird der Beweis in zwei Teile aufgeteilt.

Angenommen, f ist konvex. Ziel ist es, dass die Ungleichung (3.6) erfüllt ist. Dementsprechend seien $y_1, y_2 \in C$ und $\lambda \in (0,1)$. Aufgrund der Konvexität von f gilt für

$$(y_1, f(y_1)) \in \operatorname{epi}(f) \text{ und } (y_2, f(y_2)) \in \operatorname{epi}(f)$$

mit (3.1), wobei $y_1 = x_1, y_2 = x_2, f(y_1) = \mu_1$ und $f(y_2) = \mu_2$ gesetzt wird. Daraus folgt, dass

$$f(\lambda y_1 + (1-\lambda)y_2) \leq \lambda f(y_1) + (1-\lambda)f(y_2)$$

gilt.
Angenommen, die Ungleichung (3.6) für alle $y_1, y_2 \in C$ und $\lambda \in (0,1)$ gilt. Weiter soll gezeigt werden, dass f konvex ist. Seien $y_1, y_2 \in C$, $\lambda \in (0,1)$ und die Ungleichung

$$f(\lambda y_1 + (1-\lambda)y_2) \leq \lambda f(y_1) + (1-\lambda)f(y_2)$$

gegeben. Damit ist nach Definition 3.1.1 der Epigraph der Funktion f konvex und nach Definition 3.1.4 ist folglich auch f konvex.

Somit ist gezeigt, dass die Funktion $f : C \to (-\infty, \infty]$ genau dann konvex ist, wenn die Ungleichung (3.6) für alle $y_1, y_2 \in C$ und $\lambda \in (0,1)$ erfüllt ist. □

[2]Die Fälle $\lambda \in \{0,1\}$ werden ausgeschlossen, denn für $\lambda = 0$ ergibt sich die Trivialität $f(y_2) \leq f(y_2)$ und Entsprechendes gilt für $\lambda = 1$. Daher muss in diesen Fällen nichts gezeigt werden.

Der Satz 3.2.1 wird oft als die Definition für die Charakterisierung der Konvexität von Funktionen gewählt, obwohl dieser Ansatz zu technischen Problemen führen kann, da der Fall $\infty - \infty$ auftreten könnte [10].

Daher wird im Folgenden eine Verallgemeinerung betrachtet und $X = \mathbb{R}^n$ angenommen.

Satz 3.2.2.
Sei $f \in \mathcal{F}$. Die Funktion f ist genau dann konvex, wenn

$$f(\lambda x_1 + (1-\lambda)x_2) < \lambda\beta_1 + (1-\lambda)\beta_2 \tag{3.8}$$

gilt, mit $x_1, x_2 \in X$, $\lambda \in (0,1)$ und immer, wenn $f(x_i) < \beta_i$ mit $i \in \{1,2\}$ ist.

Beweis. Sei $f \in \mathcal{F}$. Bewiesen wird, dass f genau dann konvex ist, wenn die Ungleichung (3.8) erfüllt ist, wobei $x_1, x_2 \in X$, $\lambda \in (0,1)$ und $f(x_i) < \beta_i$ für $i \in \{1,2\}$ ist.

Angenommen, die Funktion f ist konvex. Das bedeutet, es soll gezeigt werden, dass für alle $x_1, x_2 \in X$ und alle $\lambda \in (0,1)$ die Ungleichung (3.8) gilt.
Seien $x_1, x_2 \in X$, $\lambda \in (0,1)$.
Durch die Konvexität von f folgt

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda)f(x_2).$$

Gesetzt wird $\beta_i > f(x_i)$ für $i \in \{1,2\}$ und damit gilt

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda)f(x_2) < \lambda\beta_1 + (1-\lambda)\beta_2.$$

Damit ist gezeigt, dass die Ungleichung

$$f(\lambda x_1 + (1-\lambda)x_2) < \lambda\beta_1 + (1-\lambda)\beta_2$$

gilt, wenn f konvex ist.

Angenommen, die Ungleichung

$$f(\lambda x_1 + (1-\lambda)x_2) < \lambda\beta_1 + (1-\lambda)\beta_2$$

gilt für $x_1, x_2 \in X$, $\lambda \in (0,1)$ und $f(x_i) < \beta_i$.
Es wird gezeigt, dass f konvex ist. Das bedeutet, dass für alle $x_1, x_2 \in X$, alle $\lambda \in (0,1)$

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda)f(x_2)$$

gezeigt werden muss.

Seien $x_1, x_2 \in X$ und $\lambda \in (0,1)$. Es gilt

$$f(x_i) < \beta_i \text{ für } i \in \{1,2\}.$$

Somit folgt

$$\lambda f(x_1) + (1-\lambda)f(x_2) < \lambda\beta_1 + (1-\lambda)\beta_2.$$

Dies impliziert, dass

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda)f(x_2) < \lambda\beta_1 + (1-\lambda)\beta_2$$

gilt. Da dies für jede Wahl von $x_1, x_2 \in X$ und $\lambda \in (0,1)$ gilt, folgt, dass f konvex ist.

□

Weitere Eigenschaften von konvexen Funktionen werden im folgenden Satz zusammengefasst.

Satz 3.2.3.
Seien $f, f_1, \ldots, f_m$ konvexe Funktionen und $\alpha, \alpha_1, \ldots, \alpha_m \in \mathbb{R}_{\geq 0}$, dann gilt

i.) αf *ist konvex*

ii.) $f_1 + f_2$ *ist konvex*

iii.) $\sum_{i=1}^{m} \alpha_i f_i$ *ist konvex*

Beweis. Für einen Beweis zu (i), (ii) und (iii) siehe [11]. □

Satz 3.2.4.
Sei $X \subset \mathbb{R}^n$ eine offene konvexe Menge mit $X \neq \emptyset$ und sei $f\colon X \to \mathbb{R}$ eine differenzierbare Funktion.
Dann ist f konvex genau dann, wenn

$$f(x) \geq f(x_0) + \nabla f(x_0)^\top (x - x_0) \text{ für alle } x, x_0 \in X. \tag{3.9}$$

Beweis. Sei f eine konvexe Funktion. Für $\lambda \in (0,1)$ gilt

$$\begin{aligned} & f(\lambda x + (1-\lambda)x_0) \leq \lambda f(x) + (1-\lambda) f(x_0) \\ \Leftrightarrow\ & f(x_0 + \lambda(x - x_0)) \leq \lambda f(x) + f(x_0) - \lambda f(x_0) \\ & \qquad = \lambda(f(x) - f(x_0)) + f(x_0). \end{aligned}$$

Damit ergibt sich

$$\frac{f(x_0 + \lambda(x - x_0)) - f(x_0)}{\lambda} \leq f(x) - f(x_0).$$

Für λ gegen 0 erhält man

$$\nabla f(x_0)^\top (x - x_0) \leq f(x) - f(x_0). \tag{3.10}$$

Es wird angenommen, dass die Ungleichung (3.10) bzw. (3.9) gilt. Seien $x_1, x_2 \in X$ beliebig gewählt und setze $x = \lambda x_1 + (1-\lambda)x_2$ mit $\lambda \in (0,1)$.
Dann gilt

$$\begin{aligned} \nabla f(x_0)^\top (x_1 - x_0) &\leq f(x_1) - f(x_0), \\ \nabla f(x_0)^\top (x_2 - x_0) &\leq f(x_2) - f(x_0). \end{aligned}$$

Äquivalent dazu ist

$$f(x_1) \geq f(x_0) + \nabla f(x_0)^\top (x_1 - x_0), \tag{3.11}$$
$$f(x_2) \geq f(x_0) + \nabla f(x_0)^\top (x_2 - x_0). \tag{3.12}$$

Multipliziert man die Ungleichung (3.11) mit λ und (3.12) mit $(1-\lambda)$, erhält man

$$\lambda f(x_1) \geq \lambda(f(x_0) + \nabla f(x_0)^\top (x_1 - x_0))$$

und

$$(1-\lambda)f(x_2) \geq (1-\lambda)(f(x_0) + \nabla f(x_0)^\top (x_2 - x_0)).$$

Durch Addition der Ungleichungen und Nutzung von $x = \lambda x_1 + (1-\lambda)x_2$ folgt

$$\begin{aligned} \lambda f(x_1) + (1-\lambda)f(x_2) &\geq f(x_0) + \nabla f(x_0)^\top (\lambda x_1 + (1-\lambda)x_2 - x_0) \\ &= f(x_0) + \nabla f(x_0)^\top (x - x_0). \end{aligned} \tag{3.13}$$

Die Ungleichung (3.13) zeigt die Konvexität der Funktion f. Dies folgt, da die Funktion f im Punkt x kleiner gleich einer linearen Kombination ihrer Werte an den beliebig gewählten Punkten x_1 und x_2 ist. Damit sind beide Richtungen gezeigt. □

Im Falle von zweimal stetig differenzierbaren reellen Funktionen kann folgender Satz verwendet werden.

Satz 3.2.5.

i.) Sei $f\colon (\alpha, \beta) \to \mathbb{R}$ zweimal stetig differenzierbar. Die Funktion f ist genau dann konvex, wenn für alle $x \in (\alpha, \beta)$ Folgendes gilt

$$\frac{\partial^2 f}{\partial x^2} \geq 0$$

ii.) Sei $g\colon C \to \mathbb{R}$ zweimal stetig differenzierbar mit $C \subseteq \mathbb{R}^n$ offen und konvex. Die Funktion g ist genau dann konvex, wenn für alle $x \in C$ die Hesse-Matrix

$$Hg(x) = \left(\frac{\partial^2 g}{\partial x_i \partial x_j}(x) \mid i, j = 1, \dots n \right)$$

positiv semi-definit ist

Beweis. Für einen detaillierten Beweis zu (i) und (ii) siehe [12]. □

Der Satz 3.2.5 behandelt differenzierbare Funktionen und ihre Konvexität. Das Ideal solcher Funktionen ist nicht immer gegeben, deswegen werden fortan Funktionen begutachtet, die sich nicht in Gänze überall differenzieren lassen, wie zum Beispiel die bereits erwähnte Betragsfunktion.

4 Der Subgradient

In diesem Abschnitt wird die Differenzierbarkeit konvexer Funktionen und die Rolle des Subgradienten untersucht. Begonnen wird dabei mit einem Beispiel, das im Weiteren zu den Grundlagen der Differenzierbarkeit führt und anschließend werden die Definition und die Eigenschaften von Subgradienten erläutert, um ein tieferes Verständnis für konvexe Funktionen zu erlangen.

4.1 Einstieg

Beispiel 4.1.1.
Betrachtet wird die Abbildung 4.1, die den Graphen der Funktion $f(x) = |x|$ darstellt.

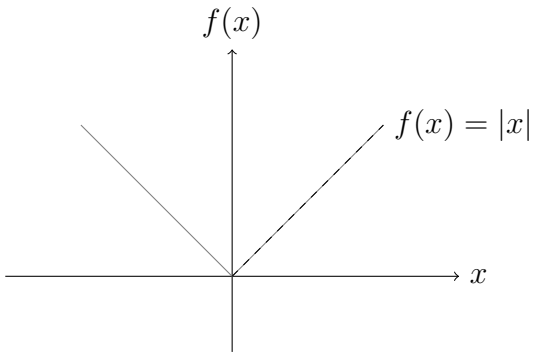

Abbildung 4.1 Graph der Betragsfunktion

Wie bereits gezeigt wurde, ist die Funktion an der Stelle $x = 0$ nicht differenzierbar. Es bleibt jedoch nachzuweisen, dass die Funktion konvex ist. Die Abbildung 4.1 deutet bereits auf die Konvexität der Funktion hin. Dabei ist zu beachten, dass f nach $\mathbb{R}_{\geq 0}$ abbildet. Zur Begründung der Konvexität bietet sich Satz 3.2.1 an.

Beweis. Zu zeigen ist, dass für die Funktion $f(x) = |x|$ mit $x \in \mathbb{R}$ für alle $x_1, x_2 \in \mathbb{R}$ und $\lambda \in (0, 1)$

$$f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2)$$

gilt.
Daher seien $x_1, x_2 \in \mathbb{R}$ beliebig und $\lambda \in (0, 1)$.
Dann gilt

$$\begin{aligned} f(\lambda x_1 + (1 - \lambda)x_2) &= |\lambda x_1 + (1 - \lambda)x_2| \\ &\leq |\lambda x_1| + |(1 - \lambda)x_2| \\ &= \lambda f(x_1) + (1 - \lambda)f(x_2). \end{aligned}$$

x_1, x_2 wurden beliebig gewählt und damit ist die Betragsfunktion konvex. □

Es zeigt sich, dass die Betragsfunktion eine nicht überall differenzierbare konvexe Funktion ist.

Trotz der Nicht-Differnzierbarkeit an der Stelle $x_0 = 0$ zeigt die Betragsfunktion, dass sie an allen Punkten außer $x_0 = 0$ differenzierbar ist. Jedoch weist sie andere Differenzierbarkeitseigenschaften auf, die im Folgenden eingeführt werden.

Definition 4.1.2 (einseitige Richtungsableitung)**.**
Eine Funktion $f\colon \mathbb{R}^n \to \overline{\mathbb{R}}$ heißt einseitig richtungsdifferenzierbar in $x \in \mathbb{R}^n$, wenn für alle $y \in \mathbb{R}^n$ der Grenzwert

$$D_y^+ f(x) := \lim_{h \to 0^+} \frac{f(x+hy) - f(x)}{h}$$

existiert. Man beachte, dass

$$-D_{-y}^- f(x) = \lim_{h \to 0^-} \frac{f(x+hy) - f(x)}{h},$$

sodass die einseitige Richtungsableitung genau dann zweiseitig ist, wenn $D_{-y}f(x)$ existiert und

$$D_{-y}f(x) = -D_y f(x)$$

gilt.

Bemerkung 4.1.3.
Die Definition 4.1.2 ist bekannt, da sie lediglich einen Spezialfall der partiellen Ableitung darstellt. Dies lässt sich verdeutlichen durch die Betrachtung des Ausdruckes

$$D_{e_i} f(x) := \lim_{h \to 0} \frac{f(x+he_i) - f(x)}{h},$$

wobei e_i der i-te Einheitsvektor ist.

Satz 4.1.4.
Sei $f \in \mathcal{F}$ konvex und an der Stelle $x \in X$ differenzierbar. Für alle $y \in X$ ist der einseitige Differenzenquotient eine wachsende Funktion mit $h > 0$, sodass $D_y^+ f(x)$ existert und

$$D_y^+ f(x) = \inf_{h>0} \frac{f(x+hy) - f(x)}{h}.$$

Außerdem ist $D_y^+ f(x)$ eine positive homogene konvexe Funktion mit $D_0^+ f(x) = 0$ und es gilt

$$-D_{(-y)}^+ f(x) \leq D_y^+ f(x) \text{ für alle } y \in \mathbb{R}^n.$$

Beweis. Für einen detaillierten Beweis zu Satz 4.1.4 siehe [13]. □

4.2 Das Subdifferential

Definition 4.2.1 (Subgradient)**.**
Sei $f \in \mathcal{F}$ konvex, dann heißt der Vektor $x' \in X$ Subgradient der Funktion f an der Stelle $x_0 \in X$, wenn

$$f(x) \geq f(x_0) + \langle x', x - x_0 \rangle \text{ für alle } x \in X \tag{4.1}$$

oder

$$f(x) - f(x_0) \geq \langle x', x - x_0 \rangle \text{ für alle } x \in X \text{ } \textit{(Subgradientungleichung)} \tag{4.2}$$

gilt.

Definition 4.2.2 (Subdifferential)**.**
Das Subdifferential $\partial f(x_0)$ ist die Menge aller Subgradienten der Funktion f an der Stelle x_0.

Bemerkung 4.2.3.
Intuitiv bedeutet der Subgradient in $\mathbb{R}$, dass der entsprechende Graph der gewählten Funktion f oberhalb der Geraden g liegt, wobei

$$g(x) := x'(x - x_0) + f(x_0)$$

durch den Punkt $(x_0, f(x_0))$ verläuft und die Steigung x' besitzt. Das Subdifferential ist damit die Menge

$$G := \{(x, g(x)) \in \mathbb{R}^2 \mid g(x) = x'(x - x_0) + f(x_0)\},$$

die wiederum alle Geraden beschreibt, über denen sich der Graph befindet.

Zurück zum anfänglichen Beispiel.

Beispiel 4.2.4.
Für die Betragsfunktion besteht die Behauptung, dass das Subdifferential

$$\partial f(x_0) = \begin{cases} \{1\} & , \text{ für } x_0 > 0 \\ [-1, 1] & , \text{ für } x_0 = 0 \\ \{-1\} & , \text{ für } x_0 < 0 \end{cases}$$

ist.

Beweis. Sei $f(x) = |x|$.
Es wird gezeigt, dass für $\partial f(0) = [-1, 1]$ gilt.
Man halte ein beliebigens $x' \in \partial f(0)$ fest, dann gilt durch Definition 4.2.1 und für alle $x \in \mathbb{R}$

$$x'(x - 0) = x'x \leq f(x) - f(0) = |x|.$$

Dadurch erhält man für $x = 1$

$$x' \cdot 1 = x' \leq 1.$$

Für $x = -1$ ergibt sich

$$x' \cdot (-1) = -x' \leq |-1| = 1$$

und äquivalent dazu

$$-1 \leq x'.$$

Daher ist $x' \in [-1, 1]$ und damit ist $\partial f(0) \subset [-1, 1]$.

Nun wird ein beliebiges $x' \in [-1, 1]$ festgehalten.
Es gilt

$$f(x') = |x'| \leq 1.$$

Für alle $x \in \mathbb{R}$ folgt

$$x'(x - 0) = x'x \leq |x'x| = \underbrace{|x'|}_{\leq 1} |x| \leq |x| = |x| - |0| = f(x) - f(0).$$

Somit ist $x' \in \partial f(0)$ und damit ist $[-1, 1] \subset \partial f(0)$.

Wenn $x_0 > 0$, dann ist $f(x_0) = |x_0| = x_0$. Die Funktion ist an der Stelle x_0 differenzierbar mit $\frac{d}{d\,x} f(x_0) = 1 =: x'$. Es soll (4.2) gezeigt werden.
Da für $f(x) = |x|$ und für $f(x_0) = x_0$ gilt, wird diese Ungleichung zu

$$\begin{aligned} |x| - x_0 \geq x'(x - x_0) &\Leftrightarrow |x| - x_0 \geq 1 \cdot (x - x_0) \\ &\Leftrightarrow |x| - x_0 \geq x - x_0 \\ &\Leftrightarrow |x| \geq x, \end{aligned}$$

was immer wahr ist. Daher ist $\partial f(x_0) = \{1\}$ für $x_0 > 0$.
Ähnliches gilt für $x_0 < 0$. □

Es sollen nicht nur Beispiele gezeigt werden, für die sich überall das Subdifferential bilden lässt, sondern auch welche für Funktionen, die zwar konvex sind, aber sich nicht überall subdifferenzieren lassen.

Beispiel 4.2.5.
Die Funktion

$$f(x) = \begin{cases} -\sqrt{1 - x^2}, & \text{für } x \in [-1, 1] \\ +\infty, & \text{sonst} \end{cases}$$

stellt eine nicht überall subdifferenzierbare Funktion dar. Um diese Behauptung zu bestätigen, wird die Subgradientenungleichung benötigt.

Beweis. Zu zeigen ist, dass die Funktion sich nicht überall subdifferenzieren lässt. Tatsächlich ist die Funktion für $x \in (-1, 1)$ differenzierbar, dies kann anhand der Kettenregel nachgewiesen werden.
Dazu sei $x \in (-1, 1)$ beliebig. Zuerst wird der Differentialquotient der inneren Funktion

$g(x) := 1 - x^2$ bestimmt und anschließend der Differentialquotient der äußeren Funktion $h(x) := \sqrt{x}$. Schließlich werden beide Differentialquotienten miteinander multipliziert. Der Differentialquotient der inneren Funktion $g(x)$ ist

$$g'(x) = -2x.$$

Der Differentialquotient der äußeren Funktion $h(x)$ ist

$$h'(x) = \frac{1}{2}(1 - x^2)^{-\frac{1}{2}}.$$

Daher ist das Differential der Funktion f

$$f'(x) = \frac{1}{2}(1 - x^2)^{-\frac{1}{2}} \cdot (-2x) = -\frac{x}{\sqrt{1 - x^2}}.$$

Damit ist die Funktion differenzierbar für $x \in (-1, 1)$.
Beim Betrachten von $x \in \mathbb{R} \setminus (-1, 1)$ fällt Folgendes auf

$$\partial f(1) = \emptyset.$$

Denn nach Einsetzen in die Subgradientenungleichung existiert kein $x' \in \mathbb{R}$ mit

$$-\sqrt{1 - x^2} \geq x'(x - 1).$$

□

4.3 Die Subdifferenzierbarkeit

Satz 4.3.1.
Sei $f \in \mathcal{F}$ eine konvexe Funktion, die in $x \in \mathbb{R}^n$ differenzierbar ist. Dann ist x' der Subgradient von f genau dann, wenn

$$D_y^+ f(x) \geq \langle x', y \rangle$$

für alle $y \in \mathbb{R}^n$.

Beweis. Setze $z = x + hy$. Verwendet wird die Definition der Subgradientenungleichung, welche in die folgende Form übersetzt werden kann

$$\frac{f(x + hy) - f(x)}{h} \geq \langle x', y \rangle,$$

für alle $y \in \mathbb{R}^n$ und $h > 0$. Auf die Verwendung des Satzes über homogene konvexe Funktionen folgt nun die Aussage des Satzes. □

Satz 4.3.2.
Sei f eine konvexe Funktion und x eine Stelle, an der die Funktion differenzierbar ist. Wenn f subdifferenzierbar an der Stelle x ist, dann ist f eine echte konvexe Funktion.
Falls f nicht subdifferenzierbar an der Stelle x ist, dann muss es ein $y \in \mathbb{R}^n$ geben, sodass

$$D_{-y} f(x) = -D_y f(x) = -\infty.$$

Beweis. Gezeigt werden soll, dass f eine echte konvexe Funktion ist, im Falle, dass f subdifferenzierbar ist.
Dadurch, dass f an der Stelle x differenzierbar ist, existiert der Gradient $\nabla f(x)$. Somit ist f auch subdifferenzierbar an der Stelle x. Daraus folgt, dass der Gradient $\nabla f(x)$ das einzige Element des Subdifferentials $\partial f(x)$ ist. Das bedeutet $\partial f(x) = \{\nabla f(x)\}$.
Für eine differenzierbare Funktion ist der Gradient ein Subgradient. Daher gilt für alle $y \in \mathbb{R}^n$

$$f(y) \geq f(x) + \nabla f(x)^\top (y - x).$$

Dies ist die Bedingung für Konvexität bei differenzierbaren Funktionen. Da f bereits als konvex gegeben ist, ist f eine echte konvexe Funktion.
Falls f nicht subdifferenzierbar an der Stelle x ist, ist zu zeigen, dass ein $y \in \mathbb{R}^n$ existiert mit

$$D_{-y} f(x) = -D_y f(x) = -\infty.$$

Durch die Voraussetzung, dass f nicht subdifferenzierbar an der Stelle x ist, ist das Subdifferential $\partial f(x)$ leer. Das bedeutet, es existiert kein $g \in \mathbb{R}^n$, das für alle $y \in \mathbb{R}^n$ die Ungleichung

$$f(y) \geq f(x) + g^\top (y - x)$$

erfüllt.
Da f nicht subdifferenzierbar ist, bedeutet dies, dass es Richtungen gibt, in denen die einseitigen Ableitungen der Funktion f unendlich sind. Daher sind die Richtungsableitungen $D_{-y} f(x)$ und $D_y f(x)$ an der Stelle x in bestimmten Richtungen unbeschränkt.
Für die Richtungsableitung in Richtung y und $-y$ kann

$$D_y f(x) = \lim_{h \to 0^+} \frac{f(x + hy) - f(x)}{h},$$
$$D_{-y} f(x) = \lim_{h \to 0^+} \frac{f(x - hy) - f(x)}{h}$$

geschrieben werden. Da keine Subgradienten existieren, müssen diese Grenzen in einigen Richtungen unbeschränkt sein, d.h.,

$$D_y f(x) = +\infty \quad \text{und} \quad D_{-y} f(x) = -\infty$$

für ein gewisses y. □

5 Das Subgradientenverfahren

Optimierungsprobleme sind im Alltagsleben weit verbreitet. Vom Gärtner, der eine bestimmte Fläche mit einer Vielzahl von Tulpen und Rosen bepflanzen möchte, um den Ertrag zu maximieren, bis hin zum Apotheker, der die maximale Wirkung $W(x,t)$ eines Medikaments analysiert, wobei x die Dosis des Medikaments und t die Zeit in Stunden darstellt - Optimierungsprobleme sind allgegenwärtig.

5.1 Konvexe Optimierung

Die Problemstellung einer konvexen Optimierungsaufgabe wird durch eine konvexe Menge $C \in \mathbb{R}^n$, die Definitionsmenge und eine konvexe Funktion $f\colon C \to \mathbb{R}$, die Zielfunktion, definiert. Die Aufgabe besteht darin, ein Minimum der Funktion f auf C zu suchen. Dies wird in der folgenden Form beschieben

$$\min_{x \in C} f(x).$$

Falls $C = \mathbb{R}^n$ ist, dann spricht man von einer Optimierungsaufgabe ohne Nebenbedingung. Weiter wird die folgende Definition des Minimums eingeführt.

Definition 5.1.1 ((Globales) Minimum)**.**
Sei $x^* \in C$ wobei $C \subseteq \mathbb{R}^n$, dann heißt x^* ein (globaler) Minimalpunkt von f auf C, falls

$$f(x) \geq f(x^*) \text{ für alle } x \in C.$$

Definition 5.1.2 (Lokales Minimum)**.**
Sei $x^* \in C$ wobei $C \subseteq \mathbb{R}^n$, dann heißt x^* ein lokaler Minimalpunkt von f auf C, falls ein $\epsilon > 0$ existiert mit

$$f(x) \geq f(x^*) \text{ für alle } x \in C \cap \mathcal{B}(x^*, \epsilon), \tag{5.1}$$

wobei $\mathcal{B}(x^*, \epsilon)$ die offene Kugel um x^* mit Radius ϵ bezeichnet.

Ein lokales Minimum bedeutet, dass in der Umgebung von x^* (innerhalb eines bestimmten Abstands $\epsilon > 0$) kein anderer Punkt in C einen kleineren Funktionswert hat, als der Punkt x^* aufweist.

Bemerkung 5.1.3.
Beim Betrachten der Funktion $f(x) = \cos(x)$, ist festzustellen, dass die Funktion nicht konvex ist und mehrere lokale Minimalpunkte besitzt.

Wiederum sind bei konvexer Optimierung alle Minimalpunkte global. Diese Tatsache führt zu folgendem Satz.

Satz 5.1.4.
Sei $C \in \mathbb{R}^n$ konvex mit $C \neq \emptyset$ und $f\colon C \to \mathbb{R}$ eine konvexe Funktion. Dann ist jeder lokale Minimalpunkt ein globaler Minimalpunkt der Funktion f auf der konvexen Menge C und die Menge der Minimalpunkte

$$\{x \in C \mid f(x) \leq f(y) \text{ für alle } y \in C\}$$

ist ebenso konvex.

Beweis. Sei $x^* \in C$ ein lokaler Minimalpunkt einer Funktion f, d. h., dass nach Definition ein $\epsilon > 0$ existiert mit (5.1). Es folgt für ein weiteres beliebiges $y \in C$, wobei $y \neq x^*$, mit

$$\lambda := \frac{1}{2} \min \left\{ 1, \frac{1}{\epsilon ||y - x^*||} \right\}$$

aufgrund der Konvexität der Menge C, dass

$$x^* + \lambda(y - x^*) \in C \cap \mathcal{B}(x^*, \epsilon)$$

ist. Durch (5.1) erhält man

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

Weiter folgt

$$\begin{aligned} & f(x^*) \leq \lambda f(y) + (1 - \lambda) f(x^*) \\ \Leftrightarrow \quad & f(x^*) \leq \lambda f(y) + f(x^*) - \lambda f(x^*) \\ \Leftrightarrow \quad & 0 \leq \lambda f(y) - \lambda f(x^*) \\ \Leftrightarrow \quad & 0 \leq \lambda (f(y) - f(x^*)). \end{aligned}$$

Aufgrund dessen, dass $\lambda > 0$ ist, gilt

$$0 \leq f(y) - f(x^*).$$

Damit ist $f(x^*) \leq f(y)$.

Seien $x_1, x_2 \in \{x \in C \mid f(x) \leq f(y) \text{ für alle } y \in C\}$, dann sind x_1 und x_2 globale Minimalpunkte, und es gilt für alle $y_1, y_2 \in C$

$$f(x_1) \leq f(y_1)$$

und

$$f(x_2) \leq f(y_2).$$

Insbesondere gilt dies auch für $y_1 = x_2$ bzw. $y_2 = x_1$. Dabei ist

$$f(x_1) \leq f(x_2) \text{ und } f(x_2) \leq f(x_1).$$

Die Konsequenz ist

$$f(x_1) = f(x_2).$$

Weiter gilt, dass für $\lambda \in [0,1]$

$$\begin{aligned} f(\lambda x_2 + (1-\lambda)x_1) &\leq (1-\lambda)f(x_1) + \lambda f(x_2) \\ &= f(x_1). \end{aligned}$$

Daher ist

$$\lambda x_2 + (1-\lambda)x_1 \in \{x \in C \mid f(x) \leq f(y) \text{ für alle } y \in C\}.$$

Somit ist die Menge der Minimalpunkte konvex. □

Die Eindeutigkeit der Lösung konvexer Optimierungsprobleme ist nicht immer gegeben. Um die Eindeutigkeit zu gewährleisten, erweist sich folgender Satz als hilfreich.

Satz 5.1.5.
Sei $C \in \mathbb{R}^n$ konvex mit $C \neq \emptyset$ und $f\colon C \to \mathbb{R}$ sei strikt konvex. Dann gilt bei Existenz eines Minimalpunktes, dass dieser eindeutig ist.

Beweis. Sei $x^* \in C$ der Minimalpunkt der Funktion f auf der konvexen Menge C.
Durch Satz 5.1.4 ist x^* ein globaler Minimalpunkt. Ist $y \in C$ ein beliebiger Punkt mit $f(y) = f(x^*)$, dann ist dieser auch ein globaler Minimalpunkt der Funktion f auf der konvexen Menge C. Außerdem gilt aufgrund der Konvexität der Menge C, dass

$$z := \frac{1}{2}x^* + (1-\frac{1}{2})y = \frac{1}{2}(x^* + y) \in C.$$

Wäre nun $x^* \neq y$, so würde gegen die Striktheit der Konvexität der Funktion f verstoßen, denn durch

$$f(x^*) = f(y)$$

folgt

$$f(z) < \frac{1}{2}(f(x^*) + f(y)) = f(x^*)$$

und x^* wäre nicht eindeutig. Daher muss $y = x^*$ gelten. Dann gilt für $y \in C$ mit $y \neq x^*$, dass $f(y) > f(x^*)$. □

Der folgende Satz soll die Existenz (mindestens) eines Minimalpunktes untermauern.

Satz 5.1.6.
Sei $C \in \mathbb{R}^n$ abgeschlossen. Weiter sei $f\colon C \to \mathbb{R}$ eine stetige Funktion. Nun sei C beschränkt oder $\lim\limits_{||x|| \to \infty} f(x) = +\infty$.
Dann hat

$$\min_{x \in C} f(x)$$

(mindestens) eine Lösung.

Beweis. Sei $(x_n)_{n\in\mathbb{N}}$ eine Minimalfolge. Damit gilt

$$\lim_{x\in\mathbb{R}^n}(x_n) = \inf_{x\in\mathbb{R}^n} f(x).$$

Somit ist die gegebene Folge nach Voraussetzung beschränkt. Weiter gibt es eine konvergente Teilfolge $(x_k)_{k\in\mathbb{N}}$ mit dem Grenzwert $x^* := \lim_{k\to\infty} f(x_k) = f(x^*)$.
Aufgrund der Abgeschlossenheit von C und der Stetigkeit der Funktion f ergibt sich

$$\inf_{x\in\mathbb{R}^n} f(x) = \lim_{k\to\infty} f(x_k) = f(x^*).$$

Hiermit ist der Satz 5.1.6 und damit die Existenz einer Lösung gezeigt. □

Beispiel 5.1.7.
Bei Betrachtung der Funktion $f\colon (0,+\infty) \to \mathbb{R}$, mit $f(x) = \frac{1}{x}$, lässt sich feststellen, dass die Voraussetzungen von Satz 5.1.6 nicht erfüllt sind und damit im Voraus bereits klar ist, dass die Zielfunktion für

$$\min_{x\in C} f(x)$$

keine Lösung hat.

Satz 5.1.8.
Sei erneut $C \in \mathbb{R}^n$ konvex. Weiter sei gegeben, dass $f\colon \mathbb{R}^n \to \mathbb{R}$ konvex auf C und f richtungsdifferenzierbar in $x^* \in C$ ist. Dann sind die folgenden Aussagen äquivalent

i.) x^* ist eine Lösung von $\min_{x\in C} f(x)$

ii.) $D^+_{(x-x^*)} f(x^*) \geq 0$ für alle $x \in C$

Beweis. Für einen detaillierten Beweis zu Satz 5.1.8 siehe [14]. □

5.2 Das Lagrange-Multiplikatorenverfahren

Im Folgenden wird das Lagrange-Multiplikatorenverfahren betrachtet, das $\min_{x\in C} f(x)$ löst. Dieses Verfahren löst jedoch nur zweimal stetig differenzierbare Funktionen.

Beispiel 5.2.1.
Betrachtet wird folgendes Optimierungsproblem mit einer Restriktion

Minimiere	$f(x,y) = x^2 + y^2$
unter der Nebenbedingung	$x + y \geq 1.$

Es handelt sich dabei um eine bekannte Methode zur Lösungsfindung für zweimal stetig differenzierbare Funktionen. Zunächst werden wieder die Zielfunktion und die Nebenbedingungen definiert

Minimiere	$f(x,y) = x^2 + y^2$
unter den Nebenbedingungen	$x + y \geq 1,$
	$x \geq 0,$
	$y \geq 0.$

Für die Ungleichung $x + y \geq 1$ wird ein Lagrange-Multiplikator λ eingeführt. Die Lagrange-Funktion lautet

$$L(x, y, \lambda) = x^2 + y^2 + \lambda(1 - x - y).$$

Die notwendigen Bedingungen für ein Minimum sind

$$\frac{\partial L}{\partial x} = 2x - \lambda = 0 \tag{5.2}$$
$$\frac{\partial L}{\partial y} = 2y - \lambda = 0 \tag{5.3}$$
$$\frac{\partial L}{\partial \lambda} = 1 - x - y = 0 \tag{5.4}$$

Aus (5.2) und (5.3) erhält man

$$2x = \lambda \text{ und } 2y = \lambda.$$

Daraus folgt

$$2x = 2y \Leftrightarrow x = y.$$

$x = y$ wird in (5.4) eingesetzt

$$\begin{aligned} 1 - x - y &= 1 - x - x \\ &= 1 - 2x = 0 \\ \Leftrightarrow 2x &= 1 \\ \Leftrightarrow x &= \frac{1}{2}. \end{aligned}$$

Damit gilt auch

$$y = \frac{1}{2}.$$

Durch Prüfung der Bedingungen resultiert

$$\begin{aligned} x + y = \frac{1}{2} + \frac{1}{2} = 1 \qquad & \text{(Bedingung erfüllt)} \\ x \geq 0, \quad y \geq 0 \qquad & \text{(Bedingungen erfüllt).} \end{aligned}$$

Einsetzen von $x = \frac{1}{2}$ und $y = \frac{1}{2}$ in die Zielfunktion liefert

$$f\left(\frac{1}{2}, \frac{1}{2}\right) = \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2 = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}.$$

Das Minimum der Zielfunktion $f(x, y) = x^2 + y^2$ unter den gegebenen Nebenbedingungen wird bei $x = \frac{1}{2}$ und $y = \frac{1}{2}$ erreicht. Der minimale Wert der Zielfunktion ist $\frac{1}{2}$ und somit erhält man den Minimalpunkt $(0,5, 0,5, 0,5)$. Die Veranschaulichungen in Abbildung 5.1 und 5.2 stellen genau diesen Minimalpunkt dar.

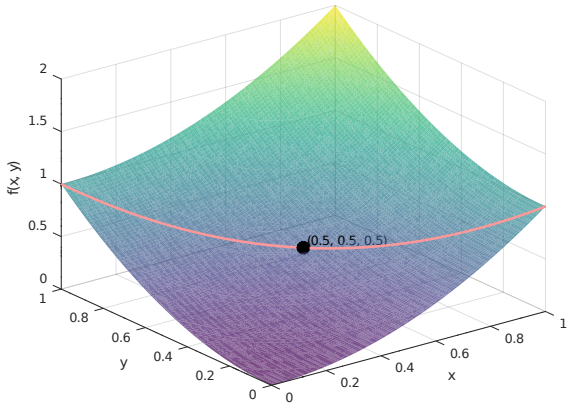

Abbildung 5.1 Minimalpunkt der Funktion $f(x, y) = x^2 + y^2$

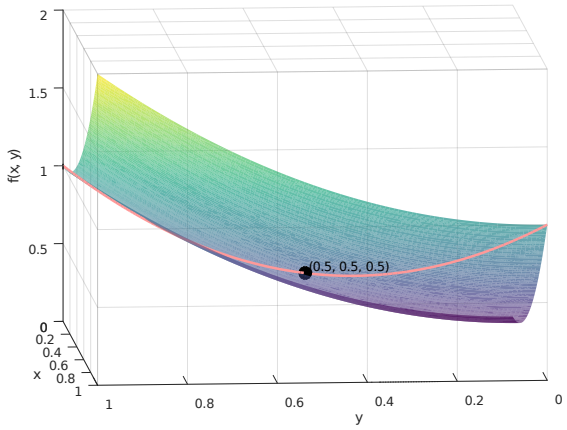

Abbildung 5.2 Andere Pespektive der Abbildung 5.1

Beim Betrachten der Funktion und der Abbildungen 5.1 und 5.2 fällt auf, dass die Funktion f konvex ist, was durch Satz 3.2.5 (ii) gezeigt werden kann.
Die Funktion sei gegeben durch

$$f(x, y) = x^2 + y^2.$$

Die ersten Ableitungen sind

$$f_x = \frac{\partial f}{\partial x} = 2x,$$

$$f_y = \frac{\partial f}{\partial y} = 2y.$$

Die zweiten Ableitungen sind

$$f_{xx} = \frac{\partial^2 f}{\partial x^2} = 2,$$

$$f_{yy} = \frac{\partial^2 f}{\partial y^2} = 2,$$

$$f_{xy} = \frac{\partial^2 f}{\partial x \partial y} = 0,$$

$$f_{yx} = \frac{\partial^2 f}{\partial y \partial x} = 0.$$

Die Hesse-Matrix H ist gegeben durch

$$H = \begin{pmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}.$$

Um die Definitheit der Hesse-Matrix zu bestimmen, betrachtet man die Eigenwerte. Da die Matrix diagonal ist, sind die Eigenwerte einfach die Diagonalelemente

$$\text{Eigenwerte: } \lambda_1 = 2, \quad \lambda_2 = 2.$$

Somit sind beide Eigenwerte positiv und die Hesse-Matrix daher positiv definit bzw. positiv semi-definit. Dementsprechend ist die Funktion nach Satz 3.2.5 (ii) konvex.

Erneut ist zu sehen, dass die Funktion überall differenzierbar ist, was nach Voraussetzung gegeben sein muss, um dieses Verfahren anzuwenden.

Bemerkung 5.2.2.
Ein weiterer Ansatz, um zu zeigen, dass $f(x, y) = x^2 + y^2$ konvex ist, ist, Satz 3.2.5 (i) mit Satz 3.2.3 (ii) zu kombinieren. Dies erspart die Berechnung der Hesse-Matrix und bei komplexeren Funktionen auch das Berechnen der Determinante.

Das Lagrange-Multiplikatorenverfahren und andere Verfahren, wie der Simplex-Algorithmus zum Lösen von Optimierungsproblemen, setzt voraus, dass die Zielfunktionen differenzierbar sind. Diese Methoden sind daher nicht direkt auf nicht-differenzierbare konvexe Optimierungsprobleme anwendbar.
Man betrachtet zum Beispiel die Funktion

$$f(x, y) = |x| + |y|.$$

Wie in Abbildung 5.3 gezeigt, ist $f(x, y)$ an der Stelle $(x, y) = (0, 0)$ nicht differenzierbar.
In solchen Fällen, in denen die Differenzierbarkeit nicht gegeben ist, ist das Subgradientenverfahren ein nützliches Werkzeug. Dieses Verfahren ist speziell für konvexe Funktionen konzipiert, die an bestimmten Punkten nicht differenzierbar sind und ermöglicht es, auch in diesen Situationen nach Lösungen zu suchen. Weiter werden Optimierungsprobleme ohne Restriktionen untersucht.

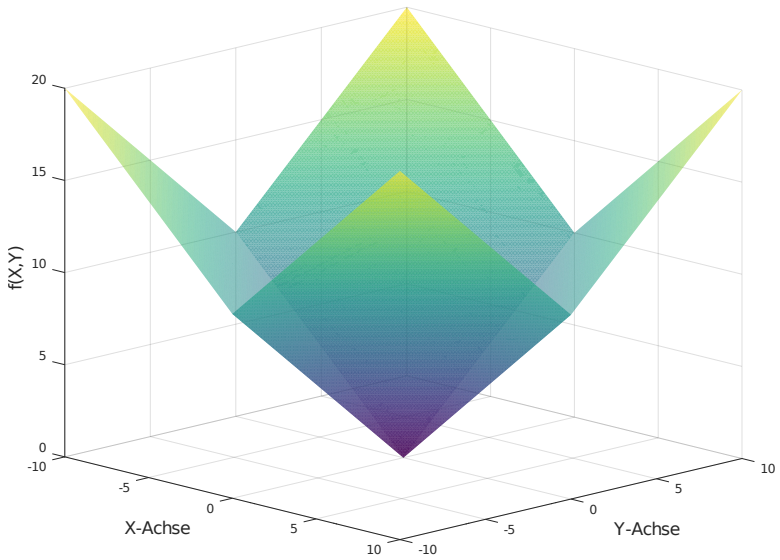

Abbildung 5.3 Graph der Funktion $f(x, y) = |x| + |y|$

5.3 Subgradientenverfahren

Zunächst wird das Subgradientenverfahren ohne spezifische Bedingungen, wie beispielsweise die Wahl der Schrittweite, eingeführt. Anschließend wird untersucht, inwieweit die Schrittweite die Konvergenz beeinflusst und welche Probleme bzw. Erkenntnisse daraus resultieren.

Algorithmus 5.3.1 (Subgradientenverfahren)**.**
Es werden notwendige Definitionen eingeführt. Es sei $k \in \mathbb{N} \cup \{0\}$.

- Startpunkt x_0 im Definitionsbereich D der Funktion f
- Anfangsschrittweite $\alpha_0 > 0$
- Ein Subgradient $\overline{x} \in \partial f(x_k)$

Weiter werden folgende Schritte ausgeführt.

i.) **Initialisierung:**
- Setzen des Startpunkts x_0
- Wählen einer Anfangsschrittweite α_0
- Festsetzen eines Abbruchkriteriums $\overline{x}_k = 0_v$ mit $0_v \in \mathbb{R}^n$

ii.) **Iteration:**

Wiederholen der folgenden Schritte.

a) **Subgradientenberechnung**
- Berechnen eines Subgradienten $\overline{x}_k$ der Funktion f an der aktuellen Stelle x_k. Ein Subgradient erfüllt die Ungleichung

$$f(x) \geq f(x_k) + \langle \overline{x}_k, (x - x_k) \rangle \text{ für alle } x \in D$$

b) **Aktualisierung**
- Setzen von $\omega_k = \frac{\overline{x}_k}{||\overline{x}_k||}$
- Bestimmen des nächsten Punktes x_{k+1} durch

$$x_{k+1} = x_k - \alpha_k \omega_k$$

c) **Wiederholung**
- Setzen von k auf $k + 1$ und Wiederholen des Vorgangs beginnend mit a)

Bemerkung 5.3.2.
Das Abbruchkriterium ist optimistisch, wird jedoch praktisch ohne weitere Einschränkungen nie erfüllt (vgl. Beispiel 5.4.3).

Beispiel 5.3.3.
Die Betragsfunktion hat ein eindeutig bestimmtes Minimum in $x^* = 0$. Wenn der Algorithmus 5.3.1 in x^* startet und $\overline{x}_0 \neq 0$, dann ist das Abbruchkriterium nicht erreichbar.
Durch die Wahl dieses Startpunktes würde man sich sogar vom Optimum entfernen.

5.4 Konvergenzanalyse

Damit sich der Algorithmus 5.3.1 eignet, müssen gewisse Bedingungen vom Startpunkt und der Schrittweite erfüllt sein.

Lemma 5.4.1.
Sei f eine konvexe Funktion und x^* ein beliebiger Minimalpunkt der Funktion f.
Weiter ist x_k kein Minimalpunkt der Funktion f, dann existiert ein $\mathrm{A}_k > 0$, sodass für den Algorithmus 5.3.1 folgende Ungleichung gilt

$$||x_{k+1} - x^*|| < ||x_k - x^*|| \text{ für alle } \alpha_k \in (0, \mathrm{A}_k). \tag{5.5}$$

Beweis. Sei $\alpha_k > 0$ beliebig gewählt, wobei $x_{k+1} = x_k - \alpha_k \omega_k$ durch den Algorithmus 5.3.1 definiert ist.
Dann gilt

$$\begin{aligned} ||x_{k+1} - x^*||^2 &= ||x_k - \alpha_k \omega_k - x^*||^2 \\ &= ||x_k - \alpha_k \frac{\overline{x}_k}{||\overline{x}_k||} - x^*||^2 \\ &= ||x_k - x^*||^2 - 2\alpha_k \langle \frac{\overline{x}_k}{||\overline{x}_k||}, x_k - x^* \rangle + \alpha_k^2 \langle \frac{\overline{x}_k}{||\overline{x}_k||}, \frac{\overline{x}_k}{||\overline{x}_k||} \rangle \\ &= ||x_k - x^*||^2 - 2\alpha_k \langle \frac{\overline{x}_k}{||\overline{x}_k||}, x_k - x^* \rangle + \alpha_k^2. \end{aligned} \tag{5.6}$$

Nach Voraussetzung gilt, dass x_k kein Minimalpunkt der Funktion f ist. Demnach gilt

$$f(x^*) \leq f(x_k).$$

Da $\overline{x}_k \in \partial f(x_k)$ ist, folgt

$$0 > f(x^*) - f(x_k) \geq \langle \overline{x}_k, x^* - x_k \rangle = -\langle \frac{\overline{x}_k}{||\overline{x}_k||}, x_k - x^* \rangle ||\overline{x}_k||.$$

Dies resultiert aus der *Subgradientenungleichung* (4.1). Damit gilt

$$-2\alpha_k \langle \frac{\overline{x}_k}{||\overline{x}_k||}, x_k - x^* \rangle + \alpha_k^2 < 0 \text{ für } \alpha_k \in (0, \mathrm{A}_k), \tag{5.7}$$

wobei $\mathrm{A}_k := 2\langle \frac{\overline{x}_k}{||\overline{x}_k||}, x_k - x^* \rangle$ gewählt wird. Mit (5.6) und (5.7) wird (5.5) gefolgert. Damit ist alles gezeigt. □

Bemerkung 5.4.2.
Es zeigt sich, dass die Wahl der Schrittweite ein wesentliches Kriterium für das Subgradientenverfahren ist. Durch die geometrische Interpretation in [15] und die daraus resultierenden Erkenntnisse über die Schrittweite

$$\lim_{k \to \infty} \alpha_k = 0 \quad \text{und} \quad \sum_{k=0}^{\infty} \alpha_k = \infty. \tag{5.8}$$

Beispiel 5.4.3.
Es wird erneut die Betragsfunktion $f(x) = |x|$ betrachtet. Die Schrittweite wird wie folgt gesetzt $\alpha_k = \frac{2}{k+1}$ mit $k \in \mathbb{N} \cup \{0\}$. Damit ist die Bedingung (5.8) erfüllt, denn wird $n = k+1$ mit $n \in \mathbb{N}$ gesetzt, dann folgt

$$\begin{aligned}\lim_{k\to\infty} \frac{2}{k+1} &= \lim_{n\to\infty} 2\frac{1}{n} \\ &= 2 \lim_{n\to\infty} \frac{1}{n} = 2 \cdot 0 = 0.\end{aligned}$$

Weiter gilt

$$\begin{aligned}\sum_{k=0}^{\infty} \frac{2}{k+1} &= 2 \sum_{k=0}^{\infty} \frac{1}{k+1} \\ &= 2 \sum_{n=1}^{\infty} \frac{1}{n} = 2 \cdot \infty.\end{aligned}$$

Daraus resultiert eine zweifache Skalierung der harmonischen Reihe, welche außerdem divergent ist.
Die Tabelle 5.1 zeigt die Iterationen des Subgradientenverfahrens mit dem Startpunkt $x_0 = 2$ und der gewählten Schrittweite.

k	x_k	$f(x_k)$	7	$-0,064$	$0,064$	15	$-0,007$	$0,007$
0	$2,000$	$2,000$	8	$0,158$	$0,158$	16	$0,110$	$0,110$
1	$1,000$	$1,000$	9	$-0,042$	$0,042$	17	$-0,001$	$0,001$
2	$0,333$	$0,333$	10	$0,140$	$0,140$	18	$0,104$	$0,104$
3	$-0,167$	$0,167$	11	$-0,027$	$0,027$	19	$0,004$	$0,004$
4	$0,233$	$0,233$	12	$0,127$	$0,127$	20	$-0,091$	$0,091$
5	$-0,100$	$0,100$	13	$-0,016$	$0,016$	21	$-0,0001$	$0,0001$
6	$0,186$	$0,186$	14	$0,117$	$0,117$	22	$0,087$	$0,087$

Tabelle 5.1 Werte des Subgradientenverfahrens von $f(x) = |x|$

Die Abbruchbedingung $\overline{x}_k = 0_v$ ist nicht erfüllt, denn für den Gradienten gilt

$$\nabla f(x_k) \in \{-1, 1\}.$$

Es ist festzuhalten, dass die Folge der Funktionswerte nicht monoton fallend ist, denn betrachtet man die Werte bei $k = 11$ und $k = 12$ gilt

$$f(x_{11}) \geq f(x_{12})$$

und je nach Wahl des Startwertes in Verbindung mit der Schrittweite, ist die Konvergenz langsam. Hierzu kann die Tabelle 6.1 betrachtet werden, wobei der Startwert bei $x_0 = 10$ liegt.

Damit die vom Subgradientenverfahren berechnete Konvergenz der Folge $(f(x_k))_{k\in\mathbb{N}}$ bewiesen werden kann, werden zuerst Hilfslemmata gezeigt.

Lemma 5.4.4.
Sei $(\alpha_k)_{k\in\mathbb{N}}$ eine Folge mit $\alpha_k \in \mathbb{R}_{>0}$ und die Bedingung (5.8) sei ebenso erfüllt. Weiter seien

$$\theta_n := \sum_{k=0}^{n} \alpha_k \text{ und } \phi_n := \sum_{k=0}^{n} \alpha_k^2 \text{ für alle } n \in \mathbb{N},$$

dann gilt

$$\lim_{n\to\infty} \frac{\phi_n}{\theta_n} = 0.$$

Beweis. Sei $(\alpha_k)_{k\in\mathbb{N}}$ die besagte Folge. Dann sind θ_n und ϕ_n wie folgt definiert

$$\theta_n := \sum_{k=0}^{n} \alpha_k \quad \text{und} \quad \phi_n := \sum_{k=0}^{n} \alpha_k^2 \quad \text{für alle } n \in \mathbb{N}.$$

Nach Voraussetzung gilt $\lim_{k\to\infty} \alpha_k = 0$, somit existiert für jedes $N \in \mathbb{N}$ mit $N > k$ ein N_0 mit $N > N_0 > k$, sodass

$$\alpha_{N_0} > \alpha_{N_0}^2.$$

Das bedeutet, dass für hinreichend große n

$$\sum_{k=N_0}^{n} \alpha_k^2 < \sum_{k=N_0}^{n} \alpha_k$$

gilt. Daraus folgt, dass θ_n die Summe ϕ_n dominiert, und somit gilt

$$\lim_{n\to\infty} \frac{\phi_n}{\theta_n} = 0.$$

□

Lemma 5.4.5.
Sei $f\colon \mathbb{R}^n \to \mathbb{R}$ eine konvexe Funktion. Weiter seien $\tilde{x}, z \in \mathbb{R}^n$ mit $f(z) > f(\tilde{x})$, $0_v \neq x^* \in \partial f(z)$ und $H_z := \{x \in \mathbb{R}^n \mid \langle x^*, \tilde{x} - z\rangle = 0\}$, dann gilt für das Minimierungsproblem

$$\min_{x\in H_z} \frac{1}{2}||x - \tilde{x}||^2$$

eine eindeutig bestimmte Lösung

$$\overline{\overline{x}} = \tilde{x} + \lambda x^* \text{ mit } \lambda = \frac{\langle x^*, z - \tilde{x}\rangle}{||x^*||^2}.$$

Außerdem ist $\lambda > 0$, für $||\overline{\overline{x}} - \tilde{x}|| = \lambda||x^*||$ und $f(\overline{\overline{x}}) \geq f(z)$.

Beweis. Setze $\rho(x) = \frac{1}{2}\|x - \tilde{x}\|^2$ mit der Nebenbedingung $\langle x^*, \tilde{x} - z\rangle = 0$. Die Funktion ρ ist eine quadratische Funktion und daher differenzierbar und insbesondere strikt konvex. Die Nebenbedingung

$$\langle x^*, \tilde{x} - z\rangle = {x^*}^\top(\tilde{x} - z)$$

ist eine lineare Abbildung. Somit kann das Lagrange-Multiplikatorenverfahren, wie in Beispiel 5.2.1 gezeigt, angewendet werden. Damit erhält man

$$\nabla p(\overline{\overline{x}}) - \lambda x^* = \overline{\overline{x}} - \tilde{x} - \lambda x^* = 0_v \tag{5.9}$$

und

$$\langle x^*, \overline{\overline{x}} - z \rangle = 0. \tag{5.10}$$

Durch (5.9) erhält man

$$\overline{\overline{x}} = \tilde{x} + \lambda x^*.$$

Durch Einsetzen in (5.10) folgt

$$0 = \langle x^*, \tilde{x} - \lambda x^* - z \rangle = \langle x^*, \tilde{x} - z \rangle + \lambda \left\| x^* \right\|^2$$

und damit gilt

$$\lambda = \frac{\langle x^*, z - \tilde{x} \rangle}{||x^*||^2} = -\frac{\langle x^*, \tilde{x} - z \rangle}{\left\| x^* \right\|^2}.$$

Dies ist die Darstellung von λ.

Weiter gilt nach Voraussetzung, dass

$$f(z) > f(\tilde{x})$$

und damit auch äquivalent

$$0 > f(\tilde{x}) - f(z).$$

Durch die Konvexität der Funktion und der damit geltenden *Subgradientenungleichung* folgt

$$0 > f(\tilde{x}) - f(z) \geq x^{*\top}(\tilde{x} - z).$$

Damit ist $x^{*\top}(\tilde{x} - z) > 0$, demnach ist auch $\lambda > 0$. Mit (5.9) erhält man

$$\left\| \overline{\overline{x}} - \tilde{x} \right\| = \lambda \left\| x^* \right\|.$$

Ebenso wird aus der *Subgradientenungleichung* für x^*

$$f(\overline{\overline{x}}) - f(z) \geq x^{*\top}(\overline{\overline{x}} - z) = 0$$

erhalten. Aufgrund (5.10) gilt diese Aussage. □

Bemerkung 5.4.6.
Wie im Beispiel 5.4.3 gezeigt, ist die Folge $(f(x_k))_{k\in\mathbb{N}}$ im Allgemeinen nicht monoton fallend. Daraus folgt, dass auch die Funktionswertefolge im Subgradientenverfahren im Allgemeinen keine fallende Monotonie aufweist. Daher wird nun eine Folge definiert, die die idealen Funktionswerte $(\tilde{f}_k)_{k\in\mathbb{N}}$ beschreibt, die durch das Subgradientenverfahren erzeugt werden. Diese Folge ist wie folgt definiert

$$\tilde{f}_k := \min\{f(x_i) \mid i = 0, 1, \ldots, k\}.$$

Satz 5.4.7 (Konvergenz des Subgradientenverfahrens)**.**
Sei f eine konvexe Funktion. Für die Schrittweitefolge soll die Bedingung (5.8) *erfüllt sein. Bricht das Subgradientenverfahren nicht nach endlich vielen Schritten ab, dann gilt*

$$\lim_{k\to\infty} \tilde{f}_k = f^* := \inf_{x\in\mathbb{R}^n} f(x),$$

wobei $f^* = -\infty$ sein kann.

Beweis. Angenommen, die Aussage ist falsch. Dann existiert ein $\tilde{x} \in \mathbb{R}^n$ und ein $\eta > 0$, sodass

$$f(x_k) \geq \tilde{f}_k \geq f(\tilde{x}) + \eta \text{ für alle } k \in \mathbb{N} \cup \{0\}. \tag{5.11}$$

Nach (5.6) gilt für $k \in \mathbb{N}$

$$||x_{k+1} - \tilde{x}||^2 = ||x_k - \tilde{x}||^2 - 2\alpha_k \langle \frac{\overline{x}_k}{||\overline{x}_k||}, x_k - \tilde{x}\rangle + \alpha_k^2.$$

Es gilt $\overline{x}_k \neq 0_v$, aufgrund dessen, dass das Subgradientenverfahren nicht nach endlich vielen Iterationen abbricht. Wird Lemma 5.4.5 angewendet, wobei Folgendes gesetzt wird

$$z = x_k,\ x^* = \overline{x}_k \text{ und } \lambda = \frac{\langle \frac{\overline{x}_k}{||\overline{x}_k||}, x_k - \tilde{x}\rangle}{\overline{x}_k} > 0$$

und $\langle \frac{\overline{x}_k}{||\overline{x}_k||}, x_k - \tilde{x}\rangle > 0$ für alle $k \in \mathbb{N}$. Weiter sei

$$\delta_m := \min_{0\leq k\leq m} \langle \frac{\overline{x}_k}{||\overline{x}_k||}, x_k - \tilde{x}\rangle \text{ für } m \geq 1.$$

Damit gilt für k

$$||x_{k+1} - \tilde{x}||^2 = ||x_k - \tilde{x}||^2 - 2\alpha_k\delta_m + \alpha_k^2.$$

Durch Betrachtung der folgenden Aufsummierung

$$||x_{k+1} - \tilde{x}||^2 + 2\delta_m \sum_{k=0}^{m} \alpha_k \leq ||x_k - \tilde{x}||^2 + \sum_{k=0}^{m} \alpha_k^2,$$

gilt

$$\lim_{m\to\infty} \delta_m = 0. \tag{5.12}$$

Für $m \geq 1$ sei ${}^m k \in \{0, \dots m\}$ ein Index mit $\delta_m = \langle \frac{\overline{x}_k}{||\overline{x}_k||}, x_k - \tilde{x}\rangle_{{}^m k}$. Sei weiter $K := \{{}^m k \mid m \leq 1\}$. Somit gilt nach (5.12)

$$\lim_{s\in K} \langle \frac{\overline{x}_s}{||\overline{x}_s||}, x_s - \tilde{x}\rangle = 0. \tag{5.13}$$

Daher existiert ein $M > 0$ mit $0 \leq \langle \frac{\overline{x}_s}{||\overline{x}_s||}, x_s - \tilde{x}\rangle \leq M$ für alle $s \in K$. Nach Lemma 5.4.5, wobei Folgendes gesetzt wird

$$z = x_s,\ x^* = \overline{x}_s,$$

hat das Minimierungsproblem $\min_{x \in H} \frac{1}{2}||x - \tilde{x}||^2$ mit

$$H := H_{x_s}\{x \in \mathbb{R}^n \mid \langle \overline{x}_s, x - x_s \rangle = 0\}$$

die eindeutig bestimmte Lösung

$$\overline{\overline{x}}_s = \tilde{x} + \lambda_s \overline{x}_s \text{ mit } \lambda_s = \frac{\langle \overline{x}_s, x_s - \tilde{x} \rangle}{||\overline{x}_s||^2} = \frac{\langle \frac{\overline{x}_s}{||\overline{x}_s||}, x_s - \tilde{x} \rangle}{||\overline{x}_s||}$$

und demnach gilt

$$||\overline{\overline{x}}_s - \tilde{x}_s|| = \langle \frac{\overline{x}_s}{||\overline{x}_s||}, x_s - \tilde{x} \rangle \leq M \text{ für alle } s \in K.$$

Nach der Lipschitz-Stetigkeit von f, existiert ein $L > 0$ mit

$$|f(\tilde{x}) - f(\tilde{x}_s)| \leq L||\tilde{x} - \tilde{x}_s|| = L\langle \frac{\overline{x}_s}{||\overline{x}_s||}, x_s - \tilde{x} \rangle \text{ für alle } s \in K.$$

Nach Lemma (5.4.5) gilt weiter

$$f(\tilde{x}_s) \geq f(x_s) \geq f(\tilde{x})$$

und damit

$$0 \leq f(x_s) - f(\tilde{x}) \leq f(\overline{\overline{x}}_s) - f(\tilde{x}) \leq L\langle \frac{\overline{x}_s}{||\overline{x}_s||}, x_s - \tilde{x} \rangle \text{ für alle } s \in K.$$

Mit dem Resultat aus (5.13) gilt

$$\lim_{l \in K} f(x_s) - f(\tilde{x}) = 0$$

und ist damit einen Widerspruch zu (5.11).

Zusammenfassend besagt der Satz, dass unter der Voraussetzung, dass die Schrittweite die notwendige Bedingung erfüllt und das Subgradientenverfahren nicht nach endlich vielen Schritten abbricht, dass die Folge der idealen, bisher gefundenen Funktionswerte gegen das Infimum der Funktion konvergiert. In anderen Worten nähern sich die iterativen Funktionswerte des Verfahrens dem bestmöglichen Wert der Funktion an, vorausgesetzt, dass das Verfahren lange genug fortgesetzt wird. □

Somit wurde das Subgradientenverfahren eingeführt und dessen Konvergenz nachgewiesen.

6 Diskussion

Im Rahmen dieser Arbeit wurde das Subgradientenverfahren eingeführt. Der Weg führte dabei über die Grundlagen und Eigenschaften konvexer Mengen und darauf folgend konvexer Funktionen, deren Gegebenheiten die Basis für das Verständnis dieses Gebietes bilden. Die Untersuchung der Differenzierbarkeit konvexer Funktionen und die Erarbeitung des Subgradienten unter Benutzung eines geeigneten Beispiels, führte schließlich zur detaillierten Darstellung des Subgradientenverfahrens und der dafür nötigen theoretischen Elemente.

Würde für das gegebene Beispiel 5.4.3 der Startwert $x_0 = 10$ und die Schrittweite $\alpha_k = \frac{2}{k+1}$ gewählt werden, würde die Tabelle 6.1 nach der 105. Iteration immer noch keine akzeptable Lösung für die vermeintlich einfache Funktion zeigen. Dies zeigt, dass die Wahl der Parameter für die Benutzer:innen eine herausfordernde Aufgabe darstellt.
Eine weitere Schwierigkeit bei der Verwendung des Subgradientenverfahrens liegt in der Wahl einer geeigneten Abbruchbedingung. Das Verfahren kann eine Näherung an das Minimum nicht gewährleisten, was die Findung verlässlicher Lösungen erheblich erschwert.
Der Grund für die langsame Konvergenz des Algorithmus liegt darin, dass der Subgradient $\overline{x}_k \in \partial f(x_k)$ selbst oft unzureichende Informationen über das lokale Verhalten der Funktion liefert.

Eine vielversprechende Möglichkeit, um den Schwierigkeiten des Subgradientenverfahrens entgegenzuwirken, besteht darin, ein approximatives Abstiegverfahren zu entwickeln, das eine genauere Annäherung an das Subdifferential ermöglicht.
Hierbei kann das ϵ-Subdifferential eingeführt werden, um die Berechnungen zu verfeinern und die Leistung des Algorithmus zu verbessern. Das ϵ-Subdifferential ermöglicht es, die Approximation des Subdifferentials auf ein gewünschtes Maß an Genauigkeit zu erhöhen, indem es die Unsicherheit im Subgradienten berücksichtigt und damit die Konvergenzgeschwindigkeit erhöht. Jedoch sind auch hier gewisse Herausforderungen und Hürden zu erwarten (vgl. [16]).

Neben dem Subgradientenverfahren gibt es weitere Verfahren, die bei der Lösung von Optimierungsproblemen von nicht-differenzierbaren Funktionen hilfreich sein können. An dieser Stelle kann beispielhaft das Proximal-Gradient-Verfahren genannt werden.

Weiterführend interessant wäre, diese Verfahren miteinander zu vergleichen und die dementsprechende Theorie für das Proximal-Gradient-Verfahren einzuführen.

k	x_k	k	x_k	k	x_k
1	8.000000	26	2.291161	51	0.962374
2	7.000000	27	2.217086	52	0.923912
3	6.333333	28	2.145658	53	0.886176
4	5.833333	29	2.076692	54	0.849139
5	5.433333	30	2.010026	55	0.812776
6	5.100000	31	1.945510	56	0.777061
7	4.814286	32	1.883010	57	0.741974
8	4.564286	33	1.822404	58	0.707491
9	4.342063	34	1.763580	59	0.673593
10	4.142063	35	1.706437	60	0.640259
11	3.960245	36	1.650882	61	0.607472
12	3.793579	37	1.596828	62	0.575214
13	3.639732	38	1.544196	63	0.543468
14	3.496875	39	1.492914	64	0.512236
15	3.363542	40	1.442914	65	0.481501
16	3.238542	41	1.394133	66	0.451245
17	3.120895	42	1.346514	67	0.421453
18	3.009784	43	1.300003	68	0.392108
19	2.904521	44	1.254548	69	0.363195
20	2.804521	45	1.210104	70	0.334698
21	2.709283	46	1.166626	71	0.306602
22	2.618373	47	1.124072	72	0.278891
23	2.531417	48	1.082406	73	0.251551
24	2.448084	49	1.041589	74	0.224566
25	2.368084	50	1.001589	75	0.197922
76	0.170973	86	0.019399	96	-0.020774
77	0.144999	87	-0.003589	97	0.003612
78	0.119358	88	0.019138	98	-0.020341
79	0.094041	89	-0.003326	99	0.003874
80	0.069041	90	0.018877	100	-0.020101
81	0.044350	91	-0.003063	101	0.004137
82	0.019960	92	0.018616	102	-0.019860
83	-0.004137	93	-0.002801	103	0.004399
84	0.019673	94	0.018355	104	-0.019620
85	-0.003856	95	-0.002538	105	0.004660

Tabelle 6.1 Subgradientenverfahren: Werte von $f(x) = |x|$ bei $x_0 = 10$

Literaturverzeichnis

[1] Rudin, W.: *Principles of Mathematical Analysis*. 3rd. New York : McGraw-Hill, 1976. – 103–119 S.

[2] Forster, O.: *Analysis 1: Differential- und Integralrechnung einer Veränderlichen*. 4. Wiesbaden, Germany : Vieweg & Teubner, 1998 (vieweg studium; Grundkurs Mathematik). – 169–173 S. – ISBN 3–528–37224–9

[3] Rockafellar, R. T.: *Convex analysis*. Princeton, N. J. : Princeton University Press, 1970 (Princeton Mathematical Series)

[4] Forster, O.: *Analysis 2: Differentialrechnung im $\mathbb{R}^n$, gewöhnliche Differentialgleichungen*. 5. Wiesbaden, Germany : Vieweg & Teubner, 1984 (vieweg studium; Grundkurs Mathematik). – 10 S. – ISBN 3–528–37231–1

[5] Rudin, W.: *Real and Complex Analysis*. 3. New York, NY : McGraw-Hill Professional, 1986 (McGraw-Hill series in higher mathematics). – 63–64 S. – ISBN 0–07–054234–1

[6] Burkard, R. E. ; Zimmermann, U.: *Lineare Optimierungsmodelle*. Berlin, Heidelberg : Springer Berlin Heidelberg, 2012. – 125 S. http://dx.doi.org/10.1007/978-3-642-28673-5_1. http://dx.doi.org/10.1007/978-3-642-28673-5_1. – ISBN 978–3–642–28673–5

[7] Boyd, S. ; Vandenberghe, L.: *Convex Optimization*. Cambridge : Cambridge University Press, 2004. – 68 ff. S.

[8] Rockafellar, R. T.: *Convex analysis*. Princeton, N. J. : Princeton University Press, 1970 (Princeton Mathematical Series). – 30 S.

[9] Rockafellar, R. T.: *Convex analysis*. Princeton, N. J. : Princeton University Press, 1970 (Princeton Mathematical Series). – 24 S.

[10] Rockafellar, R. T.: *Convex analysis*. Princeton, N. J. : Princeton University Press, 1970 (Princeton Mathematical Series). – 25 S.

[11] Sun, W. ; Yuan, Y. X.: *Optimization Theory and Methods: Nonlinear Programming*. Springer US, 2006 (Springer Optimization and Its Applications). – 40 S. https://books.google.at/books?id=o0BYHLhhPJMC. – ISBN 9780387249766

[12] Rockafellar, R. T.: *Convex analysis*. Princeton, N. J. : Princeton University Press, 1970 (Princeton Mathematical Series). – 26–27 S.

[13] Rockafellar, R. T.: *Convex analysis*. Princeton, N. J. : Princeton University Press, 1970 (Princeton Mathematical Series). – 213–214 S.

[14] Alt, W.: *Numerische Verfahren der konvexen, nichtglatten Optimierung*. Springer-Verlag, 2013. – 17–18 S. https://link.springer.com/book/10.1007/978-3-322-80083-1?error=cookies_not_supported&code=7ea1ff8c-2dcd-438b-873d-a6ac7db472cf. – ISBN 9783322800831

[15] Alt, W.: *Numerische Verfahren der konvexen, nichtglatten Optimierung*. Springer-Verlag, 2013. – 71–72 S. https://link.springer.com/book/10.1007/978-3-322-80083-1?error=cookies_not_supported&code=7ea1ff8c-2dcd-438b-873d-a6ac7db472cf. – ISBN 9783322800831

[16] Alt, W.: *Numerische Verfahren der konvexen, nichtglatten Optimierung*. Springer-Verlag, 2013. – 79 ff. S. https://link.springer.com/book/10.1007/978-3-322-80083-1?error=cookies_not_supported&code=7ea1ff8c-2dcd-438b-873d-a6ac7db472cf. – ISBN 9783322800831

Eigenständigkeitserklärung

Hiermit bestätige ich, Maximiliano Fontana, Matrikelnummer 102868, dass ich die vorliegende Arbeit selbständig und ohne unzulässige Hilfe verfasst, keine anderen als die angegebenen Quellen und Hilfsmittel benutzt, sowie die wörtlich und sinngemäß übernommenen Passagen aus anderen Werken kenntlich gemacht habe. Die Arbeit ist weder von mir noch von einer anderen Person an der Universität Passau oder an einer anderen Hochschule zur Erlangung eines akademischen Grades eingereicht worden.
Der Plagiatsprüfung mittels einer Plagiatssoftware stimme ich zu. Mit der Zustimmung zur Verwendung einer Plagiatssoftware räume ich der Universität Passau zum Zwecke der Überprüfung meiner schriftlichen Arbeit in anonymisierter Form ein einfaches Nutzungsrecht ein.

Graz, den 1. September 2024

Maximiliano Fontana