3 Mathematische Grundlagen. Konvexe Analysis

3.1 Grundlagen

Die konvexe Analysis (auch Konvexitätstheorie genannt) untersucht geometrische Eigenschaften von konvexen Mengen, Funktionen und Funktionalen in linearen Räumen.
Eine typische Idee in der Mathematik ist, anstelle komplizierter, allgemeiner Probleme lineare Probleme zu betrachten und zwar derart, daß man möglichst wenig an Allgemeinheit verliert. Ein Beispiel für so einen Zugang wurde im ersten Semester betrachtet: Anstelle des allgemeinen topologischen Raumes $\mathcal{Z}$ wurde die Menge der Wahrscheinlichkeitsmaße im bidualen Raum betrachtet. Ein allgemeines Problem wurde linear (der Übergang von γ zu $\mathbf{M}_\gamma$). Die Einschränkung bestand darin, daß nur solche topologischen Räume betrachtet werden konnten, die isomorph zu ihren Kakutani-Stone Raum waren.
Aus Sicht der Mengen besteht die Idee der konvexen Analysis darin, anstelle einer konvexen Menge die Menge ihrer Tangenten (die liegen außerhalb der Menge) zu betrachten. Aus Sicht der Funktionen besteht die Idee der konvexen Analysis darin, anstelle von allgemeinen Funktionen Scharen affiner Funktionen zu betrachten, deren Supremum die allgemeine Funktion "von unten" approximiert. Es stellt sich heraus, daß so eine Approximation von konvexen und unterhalbstetigen Funktionen möglich ist.
Hauptsatz hierfür ist: *Der Durchschnitt einer beliebigen Schar abgeschlossener und konvexer Mengen ist wieder eine abgeschlossene und konvexe Menge.*

3.1.1 Konvexe Mengen

Wir sprechen im Weiteren stets von normierten linearen Räumen. Lineare Räume sind lokal konvex. Nicht normierte Räume können unangenehme Eigenschaften haben.

- **Gewichte** sind n-Tupel von nichtnegativen reellen Zahlen, deren Summe 1 ist. $\alpha_i \geq 0$, $\sum_{i=1}^n \alpha_i = 1$.
- Eine **konvexe Kombination** von Vektoren ist eine Linearkombination deren Koeffizienten Gewichte sind. Auch genannt Mittelwert.
- Die **Verbindungstrecke** zweier Vektoren x_1, x_2 ist die Menge aller konvexen Kombinationen dieser beiden Vektoren. $\overline{x_1 x_2}$
- Eine Teilmenge X heißt **konvex**, wenn sie mit zwei Vektoren auch ihre Verbindungstrecke enthält.
- **Satz:** Eine konvexe Menge enthält alle konvexen Kombinationen aller ihrer Elemente.
- Die leere Menge zählt per Definition als konvex. Damit gilt uneingeschränkt der folgende
- **Satz:** Der Durchschnitt einer beliebigen Menge konvexer Mengen ist konvex.
- Die **konvexe Hülle** $\text{conv}(A)$ einer Menge $A \in \mathfrak{X}$ ist die kleinste konvexe Menge, die A enthält.
- **Satz:** Die konvexe Hülle $\text{conv}(A)$ enthält alle konvexen Kombinationen aus A.
- **Triviale Gewichte** sind Gewichte bei denen ein Element 1 ist.
- Die **extremalen Elemente** einer konvexe Menge A, $\text{ext}(A)$ sind die Elemente der Menge A, die sich nur durch konvexe Kombinationen von Elementen aus A mit trivialen Gewichten darstellen lassen.

- Den Begriff “konkave Menge” gibt es nicht.
- **Beispiele**
  - Die extemalen Elemente eines konvexen Polyeders im $\mathbb{R}^n$ sind seine Eckpunkte.
  - Die extemalen Elemente einer abgeschlossenen Kugel im $\mathbb{R}^n$ sind ihre Oberfläche. Die Menge der extemalen Elemente einer offenen Kugel ist leer.
  - Die konvexe Hülle einer Geraden und eines Punktes außerhalb der Geraden ist der Streifen der durch die Gerade und die Gerade, die durch den Punkt geht und zur ersten Geraden parallel ist, gebildet wird. Die zweite Gerade gehört – bis auf den gegebenen Punkt – nicht zur konvexen Hülle.
  - Die Einheitskugel $\{x : \|x\| = 1\}$ ist konvex (falls $\mathcal{X}$ normiert ist).
  - Jedes Intervall $\mathcal{C}_{[a,b]}$ in $\mathcal{C}$ ist konvex.
  - Siehe weitere Beispiele in `http://de.wikipedia.org/wiki/Extremalpunkt`

Lemma: Folgende Aussagen sind äquivalent:

i) $x \in C$ ist ein extremales Element.

ii) Aus der Annahme, daß es $x_1, x_2 \in C$ mit $x = \frac{1}{2}x_1 + \frac{1}{2}x_2$ gibt, folgt $x_1 = x_2 = x$.

iii) Aus $x + x' \in C$ und $x - x' \in C$ folgt $x' = 0$. (Achtung: $x' \in C$ wird hier nicht gefordert!)

Beweis:
i) $\Longrightarrow$ *ii)* Das ist offensichtlich. Wir wählen $\alpha = 1/2$.
ii) $\Longrightarrow$ *iii)* x und x' seien gegeben. Wir setzen $x_1 = x - x'$ und $x_2 = x + x'$. Nach Voraussetzung ist $x_1, x_2 \in C$ und es gilt $x = x_1/2 + x_2/2$. Nach Voraussetzung folgt $x_1 = x_2 = x$ also $x' = 0$.
iii) $\Longrightarrow$ *i)* Es seien $x_1, x_2 \in C$, $\alpha \in (0, 1)$ und $x = \alpha x_1 + (1 - \alpha)x_2$. Wir zeigen $x_1 = x_2$.
Fall 1) Es sei $\alpha \leq 1/2$. Wir setzen $x' = \alpha(x_1 - x_2)$. Dann ist

$$\begin{aligned} x - x' &= x - \alpha(x_1 - x_2) = x_2 \in C \\ x + x' &= x + \alpha(x_1 - x_2) = 2\alpha x_1 + (1 - 2\alpha)x_2 \in C \end{aligned}$$

weil $x + x'$ wegen $1 - 2\alpha \geq 0$ eine konvexe Kombination von x_1 und x_2 ist.
Nach Voraussetzung ist dann $x' = 0$ und wegen $\alpha > 0$ folglich $x_1 = x_2$.
Fall 2) Es sei $\alpha \geq 1/2$. Wir setzen $x' = (1 - \alpha)(x_1 - x_2)$. Dann ist

$$\begin{aligned} x + x' &= x + (1 - \alpha)(x_1 - x_2) = x_1 \in C \\ x - x' &= x - (1 - \alpha)(x_1 - x_2) = (2\alpha - 1)x_1 + (2 - 2\alpha)x_2 \in C \end{aligned}$$

weil $x - x'$ wegen $2\alpha - 1 \geq 0$, $2 - 2\alpha \geq 0$ und $(2\alpha - 1) + (2 - 2\alpha) = 1$ eine konvexe Kombination von x_1 und x_2 ist.
Nach Voraussetzung ist dann $x' = 0$ und wegen $\alpha < 1$ folglich $x_1 = x_2$.
Damit ist gezeigt, daß *i)* aus *iii)* folgt und damit die Äquivalenz aller drei Aussagen. $\square$

Die wichtigsten Mengen in der konvexen Analysis sind **konvexe** und **abgeschlossene** Mengen, weil der Durchschnitt einer beliebigen Schar konvexer und abgeschlossener Mengen wieder konvex und abgeschlossen ist.

3.1.2 Konvexe Funktionen

Mit dem Begriff der konvexen Menge lassen sich konvexe Funktionen definieren. Von zentraler Bedeutung in der Konvexitätstheorie sind Suprema (und Infima). Deshalb ist es sinnvoll, Funktionen zu betrachten, die eine konvexe Menge in die erweiterten reellen Zahlen $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\} \cup \{-\infty\}$ abbilden (nicht verwechseln mit der 1- oder 2-Punkt-Kompaktifizierung!). Der Träger $\mathrm{supp}(F)$ einer Funktion $F : \mathcal{X} \to \overline{\mathbb{R}}$ ist die Menge $A \subset \mathcal{X}$, auf der F weder gleich $+\infty$ noch gleich $-\infty$ ist.

- Eine Funktion $F : \mathcal{X} \to \overline{\mathbb{R}}$ heißt **konvexe Funktion**, wenn die Menge $O_F = \{(x, y) \in \mathcal{X} \times \overline{\mathbb{R}} \mid y \geq F(x)\}$ konvex ist.

  Äquivalent: $\alpha_1 F(x_1) + \alpha_2 F(x_2) \geq F(\alpha_1 x_1 + \alpha_2 x_2)$.
- **Satz (Jensensche Ungleichung):** Für konvexe Funktionen gilt:

  Die konvexe Kombination von Funktionswerten ist größer als der Funktionswert von der konvexen Kombination.

  Der Mittelwert der Funktionswerte ist größer als der Funktionwert des Mittelwertes.

  $$\sum_{i=1}^{n} \alpha_i F(x_i) \geq F\left(\sum_{i=1}^{n} \alpha_i x_i\right)$$
- Eine Funktion F heißt **konkav**, wenn $-F$ konvex ist.
- **Satz:** Eine konvexe Funktion F ist stetig auf $\mathrm{supp}F$.
- Die Funktion $\chi_A(x) = 0$ für $x \in A$ und $+\infty$ sonst heißt **charakteristische Funktion** oder **Indikatorfunktion** der Konvexitätstheorie.
- **Beispiele** für konvexe Funktionen:
  - Die konstante Funktion $F(x) \equiv c$
  - Die Norm $F(x) = \|x\|$ ist konvex, wenn $\mathcal{X}$ ein normierter Raum ist. Die Dreiecksungleichung ist äquivalent zur Def. der Konvexität.
  - Ist $p \in \mathcal{X}^*$, dann ist $F(x) = \langle x, p \rangle$ sowohl konvex als auch konkav.
  - Die charakteristische Funktion oder Indikatorfunktion Für gegebenes $A \subset \mathcal{X}$ ist

    $$\chi_A(x) = \begin{cases} 0 & x \in A \\ +\infty & x \notin A \end{cases}$$
- **Satz:** Die Summe zweier konvexer Funktionen ist konvex.
- **Satz:** Das Produkt einer konvexen Funktion mit einer positiven reellen Zahl ist konvex.
- **Satz:** Das Supremum (im Riesz-Raum) zweier konvexer Funktionen ist konvex.

  Das läßt sich am einfachsten mit der Gleichung $O_F \cap O_G = O_{F \vee G}$ beweisen ($\vee$ ist das Symbol für das Supremum).
- Funktionen mit abgeschlossem Epigraph heißen **unterhalbstetig**.

  Unterhalbstetigkeit (äquivalente Definitionen). F ist uhs $\Longleftrightarrow$
  - Niveaumenge $\{x | F(x) \leq a\}$ für alle a abgeschlossen
  - Für alle $\tilde{x}$: $\liminf_{x \to \tilde{x}} F(x) \geq F(\tilde{x})$

  - O_F ist abgeschlossen

  Beispielsweise ist χ_A genau dann unterhalbstetig, wenn A abgeschlossen ist.
- $\overline{F}$ heißt die **unterhalbstetige Regularisierung** von F, falls $\overline{O_F} = O_{\overline{F}}$.

Durch die Definitionen der wichtigsten Begriffe für Funktionen mit Hilfe der Menge Epigraph, ist die folgende Bemerkung eine völlige Analogie zu der für Mengen:
Die wichtigsten Funktionen in der konvexen Analysis sind **konvexe** und **unterhalbstetige** Funktionen, weil das Supremum einer beliebigen Schar konvexer und unterhalbstetiger Funktionen wieder konvex und unterhalbstetig ist.

Es gelten die offensichtlichen Sätze

- **Satz:** χ_A ist konvex $\Longleftrightarrow$ A ist konvex.
- **Satz:** F ist konvex $\Longleftrightarrow$ O_F ist konvex.

Das legt es nahe, konvexe Mengen und Funktionen als entsprechende Durchschnitte/Suprema besonders einfacher solcher Mengen und Funktionen zu bilden.

3.1.3 Hyperebenen und Halbräume

Wir definieren folgende Mengen (es sei $x^* \in \mathcal{X}^*$ und $\alpha \in \mathbb{R}$)

- affine Hyperebene: $G_{x^*,\alpha} = \{x \in \mathcal{X} \mid \langle x, x^* \rangle = \alpha\}$
- oberer affiner Halbraum: $H_{x^*,\alpha} = \{x \in \mathcal{X} \mid \langle x, x^* \rangle \geq \alpha\}$
- unterer affiner Halbraum: $U_{x^*,\alpha} = \{x \in \mathcal{X} \mid \langle x, x^* \rangle \leq \alpha\}$

Affine Hyperebenen und Halbräume sind konvexe und abgeschlossene Mengen.
$G_{x^*,\alpha}$ können wir als affine Funktion (affine Hyperebene) betrachten:

$$E_{x^*,\alpha} = \langle x, x^* \rangle - \alpha$$

x^* ist der Anstieg, $-\alpha$ ist der Wert beim Durchgang durch die $\mathcal{X}$-Ebene.
Eine affine Hyperebene ist konvex und unterhalbstetig, da ihr Epigraph ein konvexer und abgeschlossener Halbraum ist. Folglich ist das Supremum einer beliebigen Schar affiner Funktionen konvex und unterhalbstetig.
Bemerkung: Ist die Schar leer, dann ist das Supremum per Definition identisch $-\infty$. Ist die Schar nicht leer, kann das Supremum nicht den Wert $-\infty$ annehmen, da eine affine Funktion nur endliche Werte annimmt. Hier darf nicht verwechselt werden, daß in der konvexen Analysis eine Funktion den Wert $\pm\infty$ annehmen darf. $F(x) = \infty$ hat Sinn, $F(\infty)$ nicht.

Von großem Interesse ist die Frage, ob die Umkehrung gilt, also ob man jede konvexe und unterhalbstetige Funktion (die $-\infty$ nicht annimmt) als Supremum einer gewissen Schar von affinen Funktionen darstellen kann. Diese – wesentlich schwerer zu beweisende – Aussage ist wahr und folgt aus dem **Satz von Hahn-Banach**.

3.1.4 Trennung von Mengen

Mann sagt, eine Hyperebene $G_{x^*,\alpha}$ **trennt** zwei Mengen A und B, falls $B \subset H_{x^*,\alpha}$, $A \subset U_{x^*,\alpha}$ oder äquivalent

$$\sup_{x\in A}\langle x, x^*\rangle \leq \inf_{x\in B}\langle x, x^*\rangle$$

Mann sagt, eine Hyperebene $G_{x^*,\alpha}$ **trennt strikt** zwei Mengen A und B, falls $B \subset H_{x^*,\alpha}$, $A \subset U_{x^*,\alpha}$, $A \cap G_{x^*,\alpha} = \emptyset$, $B \cap G_{x^*,\alpha} = \emptyset$ oder äquivalent

$$\sup_{x\in A}\langle x, x^*\rangle < \inf_{x\in B}\langle x, x^*\rangle$$

3.1.5 Der Satz von Hahn-Banach

Der Satz von Hahn-Banach stellt einen Zusammenhang zwischen affinen und konvexen Mengen her. Da man zwischen der analytischen und geometrischen Betrachtungsweise wechseln kann, gibt es den Satz in einer analytischen und einer geometrischen Formulierung.
Für die analytische Formulierung spielen Halbnormen die Rolle der konvexen Funktionen. Jede Halbnorm ist eine konvexe Funktion und jede konvexe, homogene Funktion ist eine Halbnorm: Eine Funktion $p : \mathfrak{X} \to \mathbb{R}$ heißt Halbnorm, wenn

$$\begin{aligned} p(\lambda x) &= \lambda p(x),\ \lambda \in \mathbb{R}_+ \quad \text{(Homogenität)} \\ p(x+y) &\leq p(x) + p(y) \quad \text{(Dreiecksungleichung)} \end{aligned}$$

Der Zusammenhang mit konvexen Funktionen folgt aus

$$p(\alpha x + \beta y) \leq p(\alpha x) + p(\beta y) = \alpha p(x) + \beta p(y), \quad \alpha, \beta \geq 0,\ \alpha + \beta = 1$$

Satz (Hahn-Banach) (geometrische Formulierung)
Es sei $A \subset \mathfrak{X}$ offen und konvex. Weiter sei $M \subset \mathfrak{X}$ ein affiner Unterraum und $A \cap M = \emptyset$. Dann existiert ein Funktional $x^* \in \mathfrak{X}^*$ und ein $\alpha \in \mathbb{R}$ derart, daß $M \subset G_{x^*,\alpha}$ und $A \cap G_{x^*,\alpha} = \emptyset$.
Satz (Hahn-Banach) (analytische Formulierung)
Es sei p eine Halbnorm auf $\mathfrak{X}$, $M \subset \mathfrak{X}$ ein linearer Unterraum und $\tilde{x}^*$ ein auf M definiertes lineares Funktional mit $\langle x, \tilde{x}^*\rangle \leq p(x)$ für alle $x \in M$. Dann existiert ein $x^* \in \mathfrak{X}^*$ mit $x^* = \tilde{x}^*$ auf M und $\langle x, x^*\rangle \leq p(x)$ für alle $x \in \mathfrak{X}$.
Der Satz von Hahn-Banach hat viele wichtige Folgerungen:
Folgerung 1: Es seien $A, B \subset \mathfrak{X}$ offen, konvex und disjunkt. Dann existiert eine $G_{x^*,\alpha}$, die A und B trennt.
Folgerung 2: Es seien $A, B \subset \mathfrak{X}$ konvex und disjunkt, A kompakt und B abgeschlossen. Dann existiert eine $G_{x^*,\alpha}$, die A und B trennt.
Folgerung 3: A sei konvex und habe kein leeres Inneres. Dann besteht ∂A aus Stützpunkten.
Folgerung 4: A sei abgeschlossen und konvex. Dann ist A der Durchschnitt aller abgeschlossener Halbräume, die A enthalten.

Aus dem Satz von Hahn-Banach folgt:
Satz: Jede konvexe, unterhalbstetige Funktion läßt sich als Supremum einer Schar von unter ihr liegenden affinen Hyperebenen beschreiben.
Zum **Beweis** wird zu einem gegebenen Punkt x eine Hyperebene konstruiert, die zwischen $F(x)$ und einem echt unterhalb von $F(x)$ liegenden Punkt liegt. Die Existenz einer solchen Hyperebene folgt aus Folgerung 2.

Der Satz wird hier im Raum $\mathcal{Y} = \mathcal{X} \oplus \mathbb{R}$ angewendet. Dann ist (o.B.d.A. setzen wir $-r^*$)

$$\langle y^*, y\rangle = \left\langle \begin{pmatrix} x^* \\ -r^* \end{pmatrix}, \begin{pmatrix} x \\ r \end{pmatrix} \right\rangle = \langle x^*, x\rangle - r^* \cdot r$$

Aus der Hyperebene $G_{y^*,\alpha} = \{y \in \mathcal{Y} \mid \langle y, y^*\rangle = \alpha\}$ erhalten wir nach Einsetzen und Auflösen nach r die affine Funktion

$$E_{x^*,\alpha}(x) = r = \left\langle \frac{1}{r^*} x^*, x \right\rangle - \frac{\alpha}{r^*}$$

3.2 Konvexe Dualitätstheorie

3.2.1 Die Legendretransformation

Zu einer gegebenen (nicht unbedingt konvexen) Funktion $F(x)$ betrachten wir einen Anstieg $x^* \in \mathcal{X}^*$ und nehmen an, daß für geeignet große reelle α, die affine Funktion $E_{x^*,\alpha}(x) = \langle x, x^* \rangle - \alpha$ unter F liegt. D.h., es soll

$$F(x) \quad \geq \quad E_{x^*,\alpha}(x) = \langle x, x^* \rangle - \alpha$$

gelten, was äquivalent zu

$$alpha \quad \geq \quad \langle x, x^* \rangle - F(x)$$

ist. Für gegebenes x^* sei $F^*(x^*)$ das kleinste solche α, also

$$F^*(x^*) = \sup_{x \in \mathcal{X}} \big(\langle x, x^* \rangle - F(x) \big)$$

Falls es zu einem Anstieg x^* kein solches α gibt (z.B. gibt es für die reelle Funktion $f(x) = e^x$ und den Anstieg $x^* = -1$ kein α mit $e^x \geq -x - \alpha$), setzen wir $F^*(x^*) = +\infty$.
$F^* : \mathcal{X}^* \longrightarrow \overline{\mathbb{R}}$ wird die zu F **konjugierte** oder **konjugiert konvexe** oder **polare** Funktion genannt. Die geometrische Bedeutung von $F^*(x^*)$ ist: $F^*(x^*)$ ist zu einem gegebenen Anstieg x^* die affine Hyperebene, die gerade noch unter $F(x)$ liegt.
Die Transformation $F \to F^*$ wird **Legendre-Fenchel-Transformation** (oder kurz Legendretransformation, obwohl Legendre im Gegensatz zu Fenchel wenig mit der Entwicklung der konvexen Analysis zu tun hatte) genannt.

$F^*(x^*)$ ist das Supremum affiner Funktionen. Die sind konvex und uhs. Folglich ist $F^*(x^*)$ stets konvex und uhs.
Die Konvexität läßt sich auch direkt beweisen: Es ist für alle x und y

$$F^*(x^*) \geq \langle x, x^* \rangle - F(x)$$

Hieraus folgt für $\alpha_i \geq 0$ und $\alpha_1 + \alpha_2 = 1$

$$\alpha_1 F^*(x_1^*) + \alpha_2 F^*(x_2^*) \geq \alpha_1 \big(\langle x, x_1^* \rangle - F(x) \big) + \alpha_2 \big(\langle x, x_2^* \rangle - F(x) \big) = \langle x, \alpha_1 x_1^* + \alpha_2 x_2^* \rangle - F(x)$$

(hier wurde $\alpha_1 + \alpha_2 = 1$ verwendet). Diese Ungleichung bleibt erhalten, wenn man das sup über x nimmt (die linke Seite hängt von x nicht ab). Das ergibt

$$\alpha_1 F^*(y_1) + \alpha_2 F^*(y_2) \geq \sup_x \big(\langle x, \alpha_1 x_1^* + \alpha_2 x_2^* \rangle - F(x) \big) = F^*(\alpha_1 x_1^* + \alpha_2 x_2^*)$$

3.2.2 Die Bikonjugierte

Analog heißt

$$F^{**}(x) = \sup_{x^* \in \mathcal{X}^*} \big(\langle x, x^* \rangle - F^*(x^*) \big)$$

die **doppelt konjugierte** oder **doppelt konjugiert konvexe** oder **bipolare** Funktion zu F.
Zu beachten ist, daß die Konstruktion der Legendretransformation punktweise geschiet. Da uns der Zusammenhang von F und F^{**} auf $\mathcal{X}$ interessiert, haben wir F^{**} auch nur auf $\mathcal{X}$ definiert. Analog könnte man auch F^{**} auf $\mathcal{X}^{**}$ definieren. Dann wäre F^{**} die konjugierte von F^*.

Wenn $F(x)$ eine Funktion ist, die das Supremum einer Schar von unter ihr liegenden affinen Hyperebenen – also eine konvexe und unterhalbstetige Funktion – ist, dann ist $F^{**}(x) = F(x)$, denn das Supremum einer solchen Schar ist identisch mit dem Supremum über die Schar $(\langle x, x^* \rangle - F^*(x^*))_{x^* \in \mathfrak{X}^*}$, denn $\langle x, x^* \rangle - F^*(x^*)$ ist gerade die größte aller Scharen $\langle x, x^* \rangle - \alpha$ die unter $F(x)$ liegen.

Die Eigenschaft $F^{**}(x) = F(x)$ ist eine besonders schöne Eigenschaft, die die Legendretransformation zu einer Involution macht. Diese Eigenschaft aben also alle konvexen und unterhalbstetigen Funktionen.

3.2.3 Die Youngsche Ungleichung

Aus der Definition von F^* folgt die offensichtliche Ungleichung

$$F(x) + F^*(x^*) \geq \langle x, x^* \rangle$$

Sie wird Youngsche Ungleichung genannt. Es ist klar, daß hier nichts zu beweisen ist. In manchen Aufgaben ist vom "Beweis der Youngsche Ungleichung" die Rede. Dann sind zwei Funktionen F und F^* gegeben und der Beweis der Youngsche Ungleichung – zusammen mit einer Bedingung, wann die Gleichheit angenommen wird – ist gerade der Beweis, daß F^* die konjugiert konvexe von F ist.

3.2.4 Schwache Dualität

Startet man mit einer konvexen unterhalbstetigen Funktion $F(x)$, erhält man $F^{**} = F$. Für eine allgemeine Funktion F gilt $F^{**}(x) \leq F(x)$. Diese Eigenschaft heißt schwache Dualität, weil es exakt die Eigenschaft ist, die man im Zusammenhang mit Lagrangefunktionen kennt.
Beweis: ÜA

$$\begin{aligned}
F^{**}(x) &= \sup\nolimits_{x^*} \Big(\langle x, x^* \rangle - F^*(x^*) \Big) = \sup\nolimits_{x^*} \Big(\langle x, x^* \rangle - \sup\nolimits_y \left(\langle y, x^* \rangle - F(y) \right) \Big) = \\
&= \sup\nolimits_{x^*} \Big(\langle x, x^* \rangle + \inf\nolimits_y \left(F(y) - \langle y, x^* \rangle \right) \Big) = \sup\nolimits_{x^*} \inf\nolimits_y \Big(F(y) + \langle x - y, x^* \rangle \Big) \leq \\
&\leq \inf\nolimits_y \sup\nolimits_{x^*} \Big(F(y) + \langle x - y, x^* \rangle \Big) = \inf\nolimits_y \left(F(y) + \sup\nolimits_{x^*} \langle x - y, x^* \rangle \right) = \\
&= \inf\nolimits_y \Big(F(y) + \chi_{\{x\}}(y) \Big) = F(x)
\end{aligned}$$

3.2.5 Beispiel: Charakteristische Funktionen

Es sei $F(x) = \chi_A(x)$, wir berechnen $F^*(x)$ und $F^{**}(x)$

$$\begin{aligned}
\chi_A^*(x^*) &= \sup_{x \in \mathfrak{X}} (\langle x^*, x \rangle - \chi_A(x)) = \sup_{x \in A} \langle x^*, x \rangle \\
\chi_A^{**} &= \chi_{\overline{co} A}
\end{aligned}$$

χ_A^{**} ist die charakteristische Funktion des Abschlusses der konvexen Hülle von A.
Bemerkung: Es gilt auch

$$\chi_A(x) = \inf\nolimits_{y \in A} \chi_{\{y\}}(x) = \chi_{\{0\}}(x - y)$$

3.2.6 Beispiel: Lineare Funktionen

Es sei $x_0^* \in \mathcal{X}^*$ gegeben und $F(x) = \langle x_0^*, x \rangle$, wir berechnen $F^*(x)$.

$$F^*(x^*) = \sup_{x \in \mathcal{X}} (\langle x^*, x \rangle - \langle x_0^*, x \rangle) = \sup_{x \in \mathcal{X}} \langle x^* - x_0^*, x \rangle = \chi_{\{x_0^*\}}(x^*)$$

Da F konvex und unterhalbstetig ist, ist $F^{**} = F$.

3.2.7 Beispiel: Quadratische Funktionale

Es sei $\mathbf{A} : \mathcal{X} \to \mathcal{X}^*$ symmetrisch und bijektiv (zwischen $D(\mathbf{A}) \subset \mathcal{X}$ und $R(\mathbf{A}) \subset \mathcal{X}^*$).
Es gilt die quadratische Ergänzung

$$\frac{1}{2}\langle x, \mathbf{A}x \rangle - \langle x, x^* \rangle + \frac{1}{2}\langle \mathbf{A}^{-1}x^*, x^* \rangle = \frac{1}{2}\Big\langle (x - \mathbf{A}^{-1}x^*), \mathbf{A}(x - \mathbf{A}^{-1}x^*) \Big\rangle$$

Hieraus folgt die Youngsche Ungleichung

$$\frac{1}{2}\langle x, \mathbf{A}x \rangle - \langle x, x^* \rangle + \frac{1}{2}\langle \mathbf{A}^{-1}x^*, x^* \rangle \geq 0$$

Gleichheit gilt für $\mathbf{A}x = x^*$ oder $\mathbf{A}^{-1}x^* = x$.
Es sei

$$F(x) = \frac{1}{2}\langle x, \mathbf{A}x \rangle$$

dann folgt aus der Youngsche Ungleichung, da das Maximum angenommen wird

$$\begin{aligned} F^*(x^*) &= \sup_x \left(\langle x, x^* \rangle - \frac{1}{2}\langle x, \mathbf{A}x \rangle \right) = \frac{1}{2}\langle \mathbf{A}^{-1}x^*, x^* \rangle \\ F(x) &= \sup_{x^*} \left(\langle x, x^* \rangle - \frac{1}{2}\langle \mathbf{A}^{-1}x^*, x^* \rangle \right) = \frac{1}{2}\langle x, \mathbf{A}x \rangle \end{aligned}$$

3.2.8 Die Minkowski-Summe

Im linearen Raum läßt sich die "Summe" zweier Mengen als

$$A + B = \{a + b \,|\, a \in A, b \in B\}$$

definieren. Diese Summe heißt **Minkowski-Summe**. Sie hat folgende Eigenschaften

- Die Minkowski-Summe aus konvexen Mengen ist wieder eine konvexe Menge.
- Es gilt $A + (B \cup C) = (A + B) \cup (A + C)$
- $conv(\sum_n S_n) = \sum_n conv(S_n)$
- Im $\mathbb{R}^n$ kann man sich diese Summe als die Menge vorstellen, die durch Überstreichung entsteht, wenn man mit den Randpunkten der einen Menge die Randpunkte der anderen Menge abfährt.

3.2.9 Die infimale Faltung

The infimal convolution (or epi-sum) of two functions f and g is defined as

$$(f * g)(x) = \inf_y \big(g(y) + f(x-y)\big)$$

Let $f_1, ..., f_m$ be proper, convex and lsc functions on $\mathbb{R}^n$. Then the infimal convolution is proper, convex and lsc.
The infimal convolution of two functions has a geometric interpretation: The (strict) epigraph of the infimal convolution of two functions is the Minkowski sum of the (strict) epigraphs of those functions.
Satz: Es ist $(f * g)^* = f^* + g^*$
Beweis:

$$\begin{aligned}
(f*g)^*(x^*) &= \sup_x \big(\langle x, x^*\rangle - (f*g)(x)\big) = \sup_x \big(\langle x, x^*\rangle - \inf_y \big(g(y) + f(x-y)\big)\big) = \\
&= \sup_x \big(\langle x, x^*\rangle + \sup_y \big(-g(y) - f(x-y)\big)\big) = \\
&= \sup_x \sup_y \big(\langle x, x^*\rangle - g(y) - f(x-y)\big) = \\
&= \sup_x \sup_y \big(\langle x+y, x^*\rangle - g(y) - f(x)\big) = \\
&= \sup_x \sup_y \Big[\big(\langle x, x^*\rangle - f(x)\big) + \big(\langle y, x^*\rangle - g(y)\big)\Big] = \\
&= \sup_x \big(\langle x, x^*\rangle - f(x)\big) + \sup_y \big(\langle y, x^*\rangle - g(y)\big) = \\
&= f^*(x^*) + g^*(x^*)
\end{aligned}$$

Das neutrale Element der Faltung ist $\chi_{\{0\}}$:

$$(f * \chi_{\{0\}}) = \inf_y \big(f(y) + \chi_{\{0\}}(x-y)\big)\big) = \inf_y \big(f(y) + \chi_{\{x\}}(y)\big)\big) = f(x)$$

Das kann man als Analogon der Gleichung aus der Theorie der Fouriertransformationen

$$f(x) = \int \delta(x-y) f(y) dy$$

betrachten (siehe auch das Kapitel “Harmonische Analysis”).

3.2.10 Weitere Eigenschaften und Formeln

Leicht lassen sich folgende Zusammenhänge beweisen:

$$\begin{aligned}
F^*(0) &= -\inf_x F(x) \\
F \le G &\Longrightarrow F^* \ge G^* \\
\Big(\inf_i F_i\Big)^* &= \sup_i F_i^* \\
\Big(\sup_i F_i\Big)^* &\le \inf_i F_i^* \ , i \in I \\
(\lambda F)^*(x^*) &= \lambda F^* \left(\frac{x^*}{\lambda}\right), \ \lambda > 0 \\
(F + \alpha)^* &= F^* - \alpha, \ \alpha \in \mathbb{R} \\
F_a(x) = F(x-a) &\Longrightarrow F_a^*(x^*) = F^*(x^*) + \langle x^*, a\rangle
\end{aligned}$$

3.2.11 Lagrangefunktionen und Legendretransformation I

Für gegebes $F : \mathcal{X} \to \mathbb{R}$, $x_0 \in \mathcal{X}$ und $x^* \in \mathcal{X}^*$ betrachten wir die Funktion

$$L(x^*, x) = F(x) + \langle x_0 - x, x^* \rangle$$

und fragen, wann sie eine Lagrangefunktion ist. Dazu testen wir auf starke Dualität. Es ist einerseits

$$\begin{aligned}
\inf_x L(x^*, x) &= \inf_x (F(x) + \langle x_0 - x, x^* \rangle) = -\sup_x (\langle x, x^* \rangle - F(x)) + \langle x_0, x^* \rangle = \\
&= \langle x_0, x^* \rangle - F^*(x^*) \\
\sup_{x^*} \inf_x L(x^*, x) &= \sup_{x^*} (\langle x_0, x^* \rangle - F^*(x^*)) = F^{**}(x)
\end{aligned}$$

und andereseits

$$\begin{aligned}
\sup_{x^*} L(x^*, x) &= \sup_{x^*} (F(x) + \langle x_0 - x, x^* \rangle) = F(x) + \sup_{x^*} \langle x_0 - x, x^* \rangle = F(x) + \chi_{\{x_0\}}(x) \\
\inf_x \sup_{x^*} L(x^*, x) &= \inf_x (F(x) + \chi_{\{x_0\}}(x)) = \inf_x (F(x) + \chi_{\{0\}}(x_0 - x)) = (F * \chi_{\{0\}})(x) = F(x)
\end{aligned}$$

Damit erhalten wir folgenden
Satz: $L(x^*, x)$ ist genau dann Lagrangefunktion, wenn F konvex und unterhalbstetig ist.
Bemerkung: Die Funktion $L(x^*, x)$ könnte man sich Lagrangian für folgendes bedingtes Minimumproblem vorstellen: Gesucht ist das Minimum von $F(x)$ unter der Bedingung, daß $x = x_0$. Das ist natürlich $F(x_0)$.

3.2.12 Die Faltung quadratischer Funktionale

Wir betrachten drei symmetrische, positive, invertierbare Operatoren $\mathbf{A}, \mathbf{B}, \mathbf{C} : \mathcal{X} \longrightarrow \mathcal{X}^*$ und die entsprechenden quadratischen Funktionale $F_{\mathbf{A}}(x) = \frac{1}{2}\langle x, \mathbf{A}x \rangle$, $F_{\mathbf{B}}(x) = \frac{1}{2}\langle x, \mathbf{B}x \rangle$, $F_{\mathbf{C}}(x) = \frac{1}{2}\langle x, \mathbf{C}x \rangle$. Es stellt sich heraus, daß $F_{\mathbf{A}} * F_{\mathbf{B}}$ ebenfalls ein quadratisches Funktional ist. Es gilt also $F_{\mathbf{A}} * F_{\mathbf{B}} = F_{\mathbf{C}}$ mit einem $\mathbf{C}$, das sich aus $\mathbf{A}$ und $\mathbf{B}$ ergibt. Wir berechnen $\mathbf{C}$ auf zwei Weisen:
ÜA

3.3 Konvexe Funktionen in $\mathbb{R}$

3.3.1 Die Legendretransformation für glatte, konvexe Funktionen

Es sei $F(x)$ konvex und glatt und $F'(x) = \varphi(x)$. $\varphi(x)$ ist monoton steigend. Wir nehmen fürs erste an, daß $\varphi(x)$ invertierbar ist und $\varphi^{-1}(y)$ die inverse Funktion zu φ ist. φ^{-1} ist ebenfalls monoton steigend.
Wir berechnen die Legendretransformation

$$F^*(y) = \sup_{x\in\mathbb{R}} \big(xy - F(x)\big)$$

unter Annahme ausreichender Glattheit. Es ist das Maximum von $xy - F(x)$ bezüglich x gesucht. Man erhält es durch Nullsetzen der Ableitung:

$$\frac{d}{dx}\big(xy - F(x)\big) = y - \varphi(x) = 0 \quad\Longrightarrow\quad x_0 = \varphi^{-1}(y)$$

Der Wert an dieser Stelle ist dann

$$F^*(y) = \sup_{x\in\mathbb{R}} \big(xy - F(x)\big) = x_0 y - F(x_0) = y\varphi^{-1}(y) - F\big(\varphi^{-1}(y)\big)$$

Für die Ableitung von $F^*(y)$ erhält man

$$\begin{aligned}
\frac{d}{dy}F^*(y) &= \frac{d}{dy}\Big(y\varphi^{-1}(y) - F\big(\varphi^{-1}(y)\big)\Big) = \\
&= \varphi^{-1}(y) + y\big(\varphi^{-1}\big)'(y) - F'\big(\varphi^{-1}(y)\big)\big(\varphi^{-1}\big)'(y) = \\
&= \varphi^{-1}(y)
\end{aligned}$$

wegen $F'\big(\varphi^{-1}(y)\big) = \varphi\big(\varphi^{-1}(y)\big) = y$. Die Funktionen $F(x)$ und $F^*(y)$ haben also zueinander inverse Ableitungen. Das liefert eine einfache Methode zur Konstruktion der Legendretransformation.

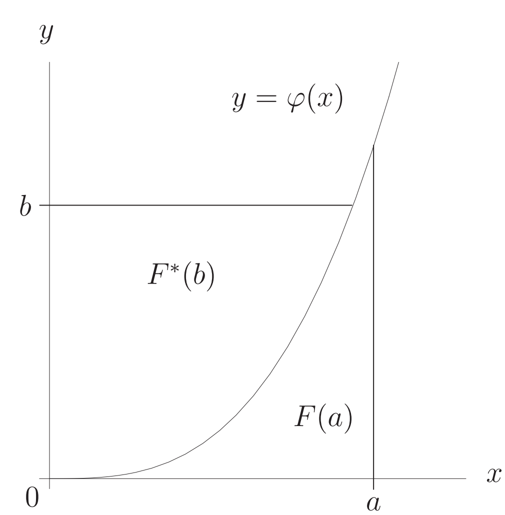
Bemerkung: Die Legendretransformation spielt eine herausragende Rolle in der Physik. Leider wird sie in den meisten Physikbüchern als $F^*(y) = y\varphi^{-1}(y) - F\big(\varphi^{-1}(y)\big)$ mit $F'(x) = \varphi(x)$ definiert. Hier ist der geometrische Sinn als Supremum über affine Funktionen nur noch schwer zu erkennen.

3.3.2 Geometrische Bedeutung der Legendretransformation

Die Youngsche Ungleichung wird an dieser Stelle zur Gleichheit:

$$F^*(y) + F(x) = xy,\ y = \varphi(x),\ x = \varphi^{-1}(y)$$

Diese Identität und die Youngsche Ungleichung haben eine einfache geometrische Interpretation. Es sei o.B.d.A. $\varphi(0) = 0$.



3.3.3 Die Bestimmung der Integrationskonstanten

Berechnet man $F^*(y)$ als Stammfunktion von $\psi(y)$, erhält man $F^*(y)$ nicht eindeutig (Integrationskonstante!). Dagegen ist $F^*(y)$ als Supremum natürlich eindeutig festgelegt. Die Integrationskonstante kann man durch Berechnung eines festen Werte ermitteln. So folgt aus

$$F^*(y) = \sup_x \big(xy - F(x)\big)$$

$$F^*(0) = \sup_x \big(-F(x)\big) = -\inf_x F(x)$$

$-F^*(0)$ ist der Funktionswert von $F(x)$, wo der Anstieg $= 0$ ist. Gibt es so einen Punkt nicht, weil $F(x)$ nach unten unbeschränkt ist, muß man sich einen geeigneten anderen Wert suchen, dessen Anstieg existiert.

Beispiel: Wir berechnen $F^*(y)$ für eine nach unten unbeschränkte Funktion $F(x)$:

$$\begin{aligned} F(x) &= x + e^x \\ \varphi(x) &= 1 + e^x = y \\ \psi(y) &= \log(y-1) \\ F^*(y) &= (y-1)\big(\log(y-1) - 1\big) + C \end{aligned}$$

An der Stelle $x = 0$ ist $F(0) = 1$ und $F'(0) = 2$. Folglich ist $F^*(2) = -F(0) = -1$ und damit $C = 0$.

3.3.4 Das Supremum über beschränkten Mengen

Betrachtet man

$$F^*(y) = \sup_{x\in[a,b]} \big(xy - F(x)\big)$$

dann ist das so, als ob die Funktion $F(x)$ über $[a, b]$ hinaus $= \infty$ gesetzt wird.
Endliche Werte nimmt $F^*(y)$ nur für $y \in [\varphi(a), \varphi(b)]$ an. Darüber hinaus wird $F^*(y)$ linear fortgesetzt. Es gilt

$$F^*(\varphi(b)) = b\varphi(b) - F(b)$$

Für die konjugierte gilt also

$$\begin{aligned} F^*(y) &= y\varphi(b) - F(b),\ y \ge \varphi(b) \\ &= y\varphi(a) - F(a),\ y \le \varphi(a) \end{aligned}$$

3.3.5 Darstellung einer Funktion durch seine Tangenten

Es sei $T_{x_0}(x)$ die Tangente an eine konvexe Funktion im Punkt x_0. Sie hat die Gleichung

$$T_{x_0}(x) = F(x_0) + (x - x_0)F'(x_0) = x\varphi(x_0) - \big(x_0\varphi(x_0) - F(x_0)\big) = x\varphi(x_0) - F^*\big(\varphi(x_0)\big)$$

und liegt unterhalb von F. Es gilt also die Ungleichung (sie folgt z.B. aus der Taylorreihe)

$$F(x) \ge T_{x_0}(x) = x\varphi(x_0) - F^*\big(\varphi(x_0)\big)$$

Das ist wieder die Youngsche Ungleichung.

$$F(x) = \sup_{x_0} T_{x_0}(x) = \sup_{x_0} \Big(x\varphi(x_0) - F^*\big(\varphi(x_0)\big)\Big)$$

3.3.6 Beispiele

$f(x)$	$\mathrm{dom}(f)$	$f^*(x^*)$	$\mathrm{dom}(f^*)$
$f(ax)(a \neq 0)$	X	$f^*\left(\frac{x^*}{a}\right)$	X^*
$f(x+b)$	X	$f^*(x^*) - \langle b, x^* \rangle$	X^*
$af(x)(a > 0)$	X	$af^*\left(\frac{x^*}{a}\right)$	X^*
$\alpha + \beta \cdot f(\gamma x + \delta)$	X	$-\alpha - \frac{\delta}{\gamma} x^\star + \beta \cdot f^\star \left(\frac{x^\star}{\beta\gamma}\right) \quad (\beta > 0)$	X^*
$\frac{\lvert x\rvert^p}{p}(p > 1)$	$\mathbb{R}$	$\frac{\lvert x^*\rvert^q}{q}(\frac{1}{p} + \frac{1}{q} = 1)$	$\mathbb{R}$
$\frac{-x^p}{p}(0 < p < 1)$	$\mathbb{R}_+$	$\frac{-(-x^*)^q}{q}(\frac{1}{p} + \frac{1}{q} = 1)$	$\mathbb{R}_{--}$
$\sqrt{1+x^2}$	$\mathbb{R}$	$-\sqrt{1-(x^*)^2}$	$[-1, 1]$
$-\log(x)$	$\mathbb{R}_{++}$	$-(1 + \log(-x^*))$	$\mathbb{R}_{--}$
e^x	$\mathbb{R}$	$\begin{cases} x^* \log(x^*) - x^* & \text{if } x^* > 0 \\ 0 & \text{if } x^* = 0 \end{cases}$	$\mathbb{R}_+$
$\log(1+e^x)$	$\mathbb{R}$	$\begin{cases} x^* \log(x^*) + (1-x^*)\log(1-x^*) & \text{if } 0 < x^* < 1 \\ 0 & \text{if } x^* = 0, 1 \end{cases}$	$[0, 1]$
$-\log(1-e^x)$	$\mathbb{R}$	$\begin{cases} x^* \log(x^*) - (1+x^*)\log(1+x^*) & \text{if } x^* > 0 \\ 0 & \text{if } x^* = 0 \end{cases}$	$\mathbb{R}_+$

3.3.7 Integraldarstellungen für konvexe Funktionen

Aus der Integraldarstellung der konvexen Kombination folgt

$$\frac{1}{\alpha_2}F(x_1) + \frac{1}{\alpha_1}F(x_2) - \frac{1}{\alpha_1\alpha_2}F(x_0) = \int_{x_1}^{x_2} \int_{\alpha_1 x_1 + \alpha_2 x}^{\alpha_1 x + \alpha_2 x_2} F''(x')dx'dx, \quad x_0 = \alpha_1 x_1 + \alpha_2 x_2$$

$$\begin{aligned}
0 &\leq \frac{1}{\alpha_2}F(x_1) + \frac{1}{\alpha_1}F(x_2) - \frac{1}{\alpha_1\alpha_2}F(\alpha_1 x_1 + \alpha_2 x_2) = \int_{x_1}^{x_2}\int_{x_1}^{x} F''\left(\alpha_1 x' + \alpha_2 x\right)dx'dx = && (16)\\
&= \frac{1}{\alpha_1}\int_{x_1}^{x_2}\Big(F'(x) - F'(\alpha_1 x_1 + \alpha_2 x)\Big)dx = && (17)\\
&= \int_{x_1}^{x_2}\int_{x_1}^{x} \frac{F'(x) - F'(\alpha_1 x_1 + \alpha_2 x)}{x - (\alpha_1 x_1 + \alpha_2 x)} = && (18)\\
&= \frac{1}{\alpha_2}\int_{\alpha_1 x_1 + \alpha_2 x_2}^{x_1} F'(x)dx + \frac{1}{\alpha_1}\int_{\alpha_1 x_1 + \alpha_2 x_2}^{x_2} F'(x)dx && (19)
\end{aligned}$$

$$\begin{aligned}
0 &\leq F(x_2) - F(x_1) - F'(x_1)(x_2 - x_1) = I(x_2, x_1) = \\
&= \int_{x_1}^{x_2}\int_{x_1}^{x} F''(x')\,dx'dx = \int_{x_1}^{x_2}\int_{x}^{x_2} F''(x)\,dx'dx = \int_{x_1}^{x_2}(x_2 - x)F''(x)\,dx = && (20)\\
&= (x_2 - x_1)^2\int_0^1(1-t)F''\big((1-t)x_1 + tx_2\big)dt = (x_2 - x_1)^2\int_0^1 tF''\big((1-t)x_2 + tx_1\big)dt = && (21)\\
&= \int_{x_1}^{x_2}\big(F'(x) - F'(x_1)\big)dx = \int_{x_1}^{x_2}\int_{x_1}^{x}\frac{F'(x) - F'(x_1)}{x - x_1}dx && (22)
\end{aligned}$$

3.3.8 Zusammenfassung. Formeln

Es sei $F : \mathbb{R} \to \mathbb{R}$, glatt, konvex, $F'(x) = \varphi(x)$. Wenn F konvex ist, dann ist φ monoton, und damit invertierbar. Es sei $\varphi(x) = y \iff \psi(y) = x$.

$$F(x) \to F'(x) = \varphi(x) = y \quad \iff \quad x = \psi(y) = F^{*\prime}(y) \to F^*(y)$$

Monotonieungleichung:

$$0 \le (x_2 - x_1)\big(\varphi(x_2) - \varphi(x_1)\big), \quad \forall x_1, x_2$$

$$(x_2 - x_1)\varphi(x_1) \le F(x_2) - F(x_1), \quad \forall x_1, x_2$$

Zwischenwertsatz:

$$\varphi(x_1) \le \frac{F(x_2) - F(x_1)}{x_2 - x_1} =: \varphi(\tilde{x}) \le \varphi(x_2), \quad x_1 \le \tilde{x} \le x_2$$

LegendreTransformation:

$$\begin{aligned} F^*(y) &= \sup_x (xy - F(x)) \\ F(x) &= \sup_y (xy - F^*(y)) \end{aligned}$$

Youngsche Ungleichung:

$$F^*(y) + F(x) \ge xy, \quad \forall x, y$$

Youngsche Ungleichung. Gleichheit:

$$\begin{aligned} x_0 y_0 &= F^*(y_0) + F(x_0), \quad \varphi(x_0) = y_0, \ \psi(y_0) = x_0 \\ F^*(y_0) &= y_0 \psi(y_0) - F(\psi(y_0)) \\ F(x_0) &= x_0 \varphi(x_0) - F^*(\varphi(x_0)) \end{aligned}$$

(Un)Gleichung der Tangente: $T_{x_0}(x)$ ist Tangente an $F(x)$ im Punkt x_0.

$$F(x) \quad \ge \quad xy_0 - F^*(y_0) = (x - x_0)\varphi(x_0) + F(x_0) = T_{x_0}(x)$$

3.3.9 Invertierbare konvexe Funktionen

$f(x)$ sei konvex, invertierbar und monoton wachsend.

$$\begin{aligned} f(\alpha_1 x_1 + \alpha_2 x_2) &\geq \alpha_1 f(x_1) + \alpha_2 f(x_2) \\ f(\alpha_1 f^{-1}(y_1) + \alpha_2 f^{-1}(y_2)) &\geq \alpha_1 y_1 + \alpha_2 y_2 \\ \alpha_1 f^{-1}(y_1) + \alpha_2 f^{-1}(y_2) &\geq f^{-1}(\alpha_1 y_1 + \alpha_2 y_2) \end{aligned}$$

Das heißt, die inverse einer konvexen und monoton wachsenden Funktion ist konkav. Analog ist die inverse einer konvexen und monoton fallenden Funktion wieder konvex.
Angenommen, f und g sind zueinander invers. Dann folgt nach Differenzieren

$$\begin{aligned} f(g(x)) &= x \\ f'(g(x))g'(x) &= 1 \\ \big(g'(x)\big)^2 f''(g(x)) &= -f'(g(x))g''(x) \end{aligned}$$

Aus der letzten Gleichung folgt, daß für monoton fallendes f die zweiten Ableitungen von f und g das gleiche Vorzeichen haben.

3.3.10 Darstellung konvexer Funktionen als Supremum von Tangenten

Es sei $F(x)$ eine konvexe Funktion. Es gibt folgende Darstellungen:

$$\begin{aligned} F(x) &= F(x_0) + F'(x_0)(x - x_0) + \frac{1}{2}(x - x_0)^2 F''(\xi) \\ F(x) &\geq F(x_0) + F'(x_0)(x - x_0) \\ F(x) &= \sup_{x_0} \big(F(x_0) + F'(x_0)(x - x_0)\big) \end{aligned}$$

Die Tangente $g_{x_0}(x)$ im Punkt x_0 hat die Gleichung

$$g_{x_0}(x) = F'(x_0)(x - x_0) + F(x_0)$$

Die Funktion $F(x)$ läßt sich darstellen als

$$F(x) = \sup_{x_0} \Big(F'(x_0)(x - x_0) + F(x_0)\Big) = \sup_{x_0} \Big(xF'(x_0) + F(x_0) - x_0 F'(x_0)\Big)$$

Gleicheit wird für $x_0 = x$ angenommen.
Es ist

$$F(x) \geq F'(x_0)(x - x_0) + F(x_0)$$

Das ist die Konvexitätsungleichung.
Es sei

$$y_0 = F'(x_0) = \varphi(x_0) \quad \Longleftrightarrow \quad x_0 = \varphi^{-1}(y_0)$$

dann gilt

$$F(x) = \sup_{y_0 \in R(\varphi)} \Big(y_0 x - y_0 \varphi^{-1}(y_0) + F\big(\varphi^{-1}(y_0)\big)\Big)$$

Tatsächlich kann man $R(\varphi)$ ausdehnen auf ganz $\mathbb{R}$ indem man die Funktion geeignet $\pm\infty$ fortsetzt. Es gilt also

$$F(x) = \sup_y \Big(yx - G(y)\Big)$$

mit

$$G(y) = y\varphi^{-1}(y) - F\big(\varphi^{-1}(y)\big)$$

Das ist gerade konjugierte: $G(y) = F^*(x)$. Wenn F konvex ist, ist F als Legendretransf. seiner Legendretransf. So könnte man auch die Legendretransf. einführen.

$$G(y) = \sup_x \big(xy - F(x)\big)$$

3.4 Stückweise affine Funktionen

sind besonders interessant, weil die Klasse dieser Funktionen invariant bezüglich konvexer Konjugation ist.
Wir betrachten konvexe Funktionen $t = F(x)$, ihre Subdiffenrentiale $y = \varphi(x)$, deren inverse $x = \varphi^{-1}(y)$ und die konjugiert konvexen $s = F^*(y)$.
Solche Funktionen können gegeben sein durch:

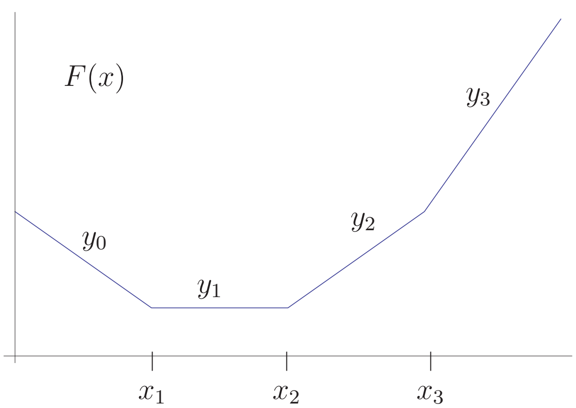
- Zwei monotone Folgen $(x_i)_{i=1}^n$ (Stützpunkte) und $(y_i)_{i=0}^n$ (Anstiege)
- Das Maximum über eine Menge von Geraden $F(x) = \max(a_i + y_i x)$.
- Als Polygon $(x_i, t_i)_{i=0}^{n+1}$

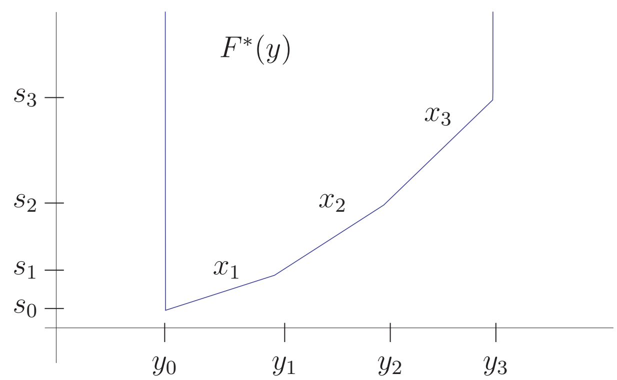
3.4.1 Gegebene Stützpunkte und Anstiege

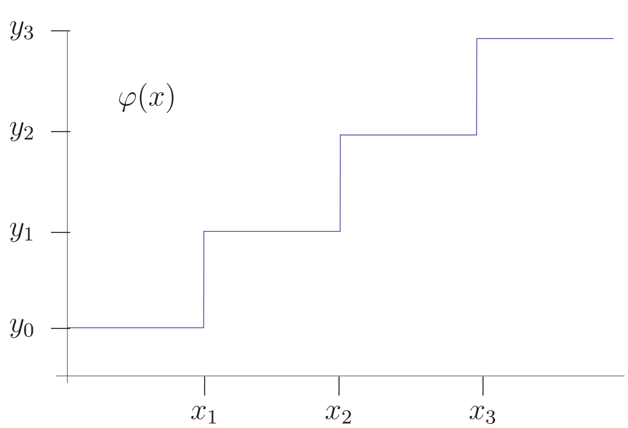
Eine stückweise affine konvexe Funktion $F(x)$ sei gegeben durch zwei monoton wachsende Folgen

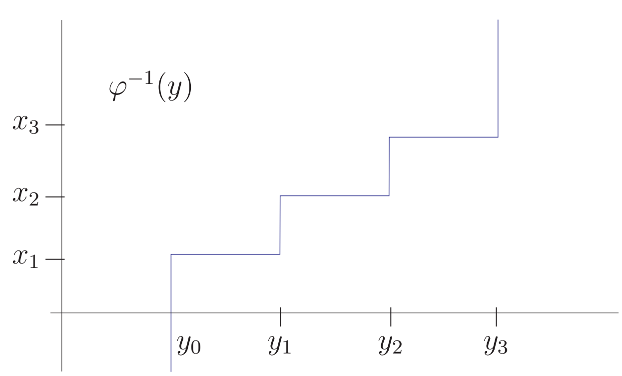
$$\begin{array}{ccccccccc} y_0 & \le & y_1 & \le & y_2 & \le & \cdots & \le & y_n \\ & & x_1 & \le & x_2 & \le & \cdots & \le & x_n \end{array} \tag{23}$$

oder allgemein $(x_{j-1} - x_j)(y_{i-1} - y_i) \ge 0$. Hier sind die x_i die Abszissenwerte und die y_i die Anstiege.
Die Ableitung $F'(x) = \varphi(x) = y$ kann man in einem (x, y)-Koordinatensystem darstellen.









Die Asymmetrie mit den ∞-Werten kommt dadurch, daß Beschränktheit beim Übergang von $\varphi(x)$ zu $\varphi^{-1}(y)$ in Unbeschränktheit übergeht. Man könnte es symmetrischer machen, indem man $y_0 = -\infty$ setzt. Dann geht der linkste Zweig von $F(x)$ parallel zur y-Achse nach oben und dafür geht das linke Geradenstück mit Anstieg x_1 von $F^*(x)$ bis nach $-\infty$.
Exakte Symmetrie ist erfüllt, falls für die Anstiege gilt $x_1 = -y_n$, $x_2 = -y_{n-1}$, ..., $x_n = -y_1$. Das hat Folgen für die Intervalle. Wegen $x_1 = -y_n$ (das ist der Punkt mit ∞ steilem Anstieg) ist 0 Symmetrieachse und es gilt $F(x) = F^*(-y)$. Das bedeutet

$$-F^*(-y) = \inf \big(xy + F(x)\big)$$

und sieht nach optimal transport und nach meinem $A_{gp}(t)$ aus.
Bild malen ?

3.4.2 Gegeben als Maximum

Es sei $F(x)$ gegeben als

$$F(x) = \max_{0\le i\le n}(a_i + y_i x)$$

Das Maximum über eine beliebige Menge von Geraden ergibt stets eine stückweise affine konvexe Funktion. Dabei kann es passieren, daß einige dieser Geraden keine Rolle spielen (weggelassen werden könnten). Damit das nicht der Fall ist, müssen die a_i und y_i bestimmt Bedingungen erfüllen, die letztlich auf die Bedingung (35) hinauslaufen. Wir setzen voraus, daß man keine der Geraden $a_i + y_i x$ weglassen kann. Dann ist

$$F(x) = a_i + y_i x,\ \ x_i \le x \le x_{i+1}$$

mit gewissen $x_0, ..., x_{n+1}$, die bestimmt werden müssen. Es ist $x_0 = -\infty$, $x_{n+1} = +\infty$ und für die anderen x_i gilt

$$a_{i-1} + y_{i-1}x_i = a_i + y_i x_i$$

also

$$x_i = \frac{a_{i-1} - a_i}{y_i - y_{i-1}}$$

Entsprechend (35) müssen die x_i und y_i monoton wachsend sein, das bedeutet $y_0 \le y_1 \le y_2 \le \cdots \le y_n$ und

$$\begin{aligned}
x_i \le x_{i+1} \quad &\Longleftrightarrow \quad \frac{a_{i-1} - a_i}{y_i - y_{i-1}} \le \frac{a_i - a_{i+1}}{y_{i+1} - y_i} \\
&\Longleftrightarrow \quad 0 \le a_{i-1}(y_i - y_{i+1}) + a_i(y_{i+1} - y_{i-1}) + a_{i+1}(y_{i-1} - y_i) \\
&\Longleftrightarrow \quad 0 \le y_{i-1}(a_{i+1} - a_i) + y_i(a_{i-1} - a_{i+1}) + y_{i+1}(a_i - a_{i-1}) \\
&\Longleftrightarrow \quad a_{i-1}(y_{i+1} - y_i) + a_{i+1}(y_i - y_{i-1}) \le a_i(y_{i+1} - y_{i-1}) \\
&\Longleftrightarrow \quad a_i \ge a_{i-1}\frac{y_{i+1} - y_i}{y_{i+1} - y_{i-1}} + a_{i+1}\frac{y_i - y_{i-1}}{y_{i+1} - y_{i-1}}
\end{aligned}$$

Das letzte ist eine Konkavitätsbedingung der a_i bezüglich der y_i.

Für die Funktionswerte erhält man

$$t_i = F(x_i) = a_{i-1} + y_{i-1}x_i = a_i + y_i x_i = \frac{a_{i-1}y_i - a_i y_{i-1}}{y_{i+1} - y_{i-1}}$$

Das Subdifferential ist das Polygon

$$\begin{aligned}\Phi = (-\infty, y_0) &\to (x_1, y_0) \to (x_1, y_1) \to (x_2, y_1) \to (x_2, y_2) \to \cdots \\ &\to (x_{i-1}, y_{i-1}) \to (x_i, y_{i-1}) \to (x_i, y_i) \to \cdots \to (x_n, y_n) \to (+\infty, y_n)\end{aligned}$$

$$\begin{aligned}F^*(y) &= \sup_x \big(xy - F(x)\big) = \max_{0\le i\le n}\left(\sup_{x_i \le x \le x_{i+1}} \big(xy - a_i - y_i x\big)\right) = \\ &= \max_{0\le i\le n}\left(-a_i + \sup_{x_i\le x\le x_{i+1}} x(y - y_i)\right) = \\ &= \max_{0\le i\le n}\left(-a_i + \begin{cases} x_i(y-y_i), & y \le y_i \\ x_{i+1}(y-y_i), & y_i \le y\end{cases}\right) = \\ &= \max_{0\le i\le n}\left(\begin{cases} -a_i + x_i(y-y_i), & y \le y_i \\ -a_i + x_{i+1}(y-y_i), & y_i \le y\end{cases}\right) = \\ &= \max_{1\le i\le n}\big(-a_i - x_i y_i + x_i y\big) = \\ &= \max_{1\le i\le n}\big(b_i + x_i y\big)\end{aligned}$$

mit $b_i = -a_i - x_i y_i$.
Man erhält

$$F^*(y) = b_i + x_i y = -a_i - x_i y_i,\ \ y_{i-1} \le y \le y_i$$

und $F^*(y_i) = s_i = -a_i$. Analog erhält man $t_i = -b_i$.

3.4.3 Gegebenes Polygon

Gegeben seien die Punkte

$$(t_0, s_0), (t_1, s_1), (t_2, s_2), ..., (t_n, s_n)$$

Gesucht ist eine Darstellung er konkaven Hülle diese Punkte. Das sei die Funktion $T(t)$. Es ist

$$T(t) = \min_{0\le k\le n-1}(y_k + h_{k+1}t)$$

mit

$$\begin{aligned}h_{k+1} &= \frac{s_{k+1} - s_k}{t_{k+1} - t_k} \\ y_k &= s_k - h_{k+1}t_k = \frac{s_k t_{k+1} - s_{k+1}t_k}{t_{k+1} - t_k}\end{aligned}$$

Das ergibt

$$\begin{aligned}T(t) &= \min_{0\le k\le n-1}\left(\frac{s_k t_{k+1} - s_{k+1}t_k + ts_{k+1} - ts_k}{t_{k+1} - t_k}\right) = \\ &= \min_{0\le k\le n-1}\left(\frac{s_k(t_{k+1} - t) + s_{k+1}(t - t_k)}{t_{k+1} - t_k}\right)\end{aligned}$$

Das ist gerade die konvexe Kombination der Funktionswerte.
Die $y_k + h_{k+1}t$ sind Geradenstücke. Aus der letzten Darstellung sieht man, daß $T(t)$ in den Punkten t_k und t_{k+1} gerade die Werte s_k und s_{k+1} annimt, wie es sein soll.
Es ergibt sich eine stückweise affine Funktion, die durch die beiden Folgen (Stützpunkte und Anstiege) $t_0, t_1, ..., t_n$ und $h_1, h_2, ..., h_n$ definiert ist.
Im allgemeinen kann passieren, daß manche Punkte keine Rolle spielen. Das ist nicht der Fall, wenn Stützpunkte und Anstiege monoton sind. Angenommen, die Stützpunkte sind monoton, dann muß $h_{k+1} \geq h_k$ gelten. Das ist gerade der Fall, wenn

$$s_k t_{k-1} - s_{k+1} t_{k-1} - s_{k-1} t_k + s_{k+1} t_k + s_{k-1} t_{k+1} - s_k t_{k+1} \geq 0$$

gilt. Das sieht wie die Konvexitätsbedingung aus.

3.4.4 Zusammenfassung

$F(x)$	$F^*(y)$
$\max\limits_{0 \leq i \leq n} (-s_i + y_i x)$	$\max\limits_{1 \leq i \leq n} (-t_i + x_i y)$
geht durch (x_i, t_i)	geht durch (y_i, s_i)
$x_i = \dfrac{s_i - s_{i-1}}{y_i - y_{i-1}}$	$y_i = \dfrac{t_i - t_{i-1}}{x_i - x_{i-1}}$
$t_i = x_i y_i - s_i$	$s_i = x_i y_i - t_i$
$= \dfrac{s_i y_{i-1} - s_{i-1} y_i}{y_i - y_{i-1}}$	
$\max\limits_{0 \leq i \leq n} \left(t_{i-1} \dfrac{x_i - x}{x_i - x_{i-1}} + t_i \dfrac{x - x_{i-1}}{x_i - x_{i-1}} \right)$	$\max\limits_{1 \leq i \leq n} \left(s_{i-1} \dfrac{y_i - y}{y_i - y_{i-1}} + s_i \dfrac{y - y_{i-1}}{y_i - y_{i-1}} \right)$

3.4.5 Gedanken

- Warum ist

$$\max_{0 \leq i \leq n} \left(\begin{cases} -a_i + x_i(y - y_i), & y \leq y_i \\ -a_i + x_{i+1}(y - y_i), & y_i \leq y \end{cases} \right) = \max_{1 \leq i \leq n} \big(- a_i - x_i y_i + x_i y \big)$$

- Kann man auch direkt von $F(x) = \max ...$ auf $F^*(y) = \max ...$ schließen? Es ist

$$\begin{aligned} F^*(y) &= \sup_x \left(xy - \max_{0 \leq i \leq n} (a_i + y_i x) \right) = \sup_x \left(\min_{0 \leq i \leq n} (xy - a_i - y_i x) \right) = \\ &= \operatorname{conv} \left(\min_{0 \leq i \leq n} \left(- a_i + \sup_x x(y - y_i) \right) \right) = \operatorname{conv} \left(\min_{0 \leq i \leq n} \left(- a_i + \chi_{\{y_i\}}(y) \right) \right) \end{aligned}$$

  Das sieht aus wie die konvexe Hülle um die Punkte $(y_i, -a_i)$. Und das ist richtig! Nur das min ist überflüssig.

- Jede konvexe Funktion kann man von oben (an den Rändern unendlich steile) oder von unten (an den Rändern affine Funktionen) durch stückweise affine Funktionen approximieren.

3.5 Differenzierbarkeit und Subdifferenzierbarkeit

3.5.1 Das Subdifferential

Wir haben gesehen, daß es für manche Funktionen $F : \mathcal{X} \to \mathbb{R}$ – etwa konvexe unterhalbstetige – in einem Punkt x eine affine Funktion gibt, die F im Punkt x berühert und unter F liegt. In diesem Fall existiert ein $x^* = x^*(x)$ mit

$$F(y) \geq F(x) + \langle y - x, x^* \rangle,\ y \in \mathcal{X}$$

In diesem Fall wird gesagt, daß F im Punkt x **subdifferenzierbar** ist und der Anstieg x^* der entsprechenden affinen Funktion in diesem Punkt heißt **Subgradient**.
Es ist klar, daß es mehrere solche affine Funktionen in einem Punkt geben kann, wenn F in diesem Punkt nicht glatt ist. Die Menge aller Subgradienten heißt **Subdifferential** $\partial F(x)$ von F in x.

$$\partial F(x) = \{x^* \in \mathcal{X}^* | F(y) \geq F(x) + \langle y - x, x^* \rangle,\ y \in \mathcal{X}\}$$

Das Subdifferential ist eine Teilmenge des dualen Raumes: $\partial F(x) \subset \mathcal{X}^*$.
Falls F in x nicht subdifferenzierbar ist, wird $\partial F(x) = \emptyset$ festgelegt.

Beispiel 1: Wir betrachten die Funktion $F : \mathbb{R} \to \mathbb{R}$ definiert durch $F(x) = 1$ für $x > 0$, sonst ist $F(x) = 0$. Es ist $\partial F(0) = \{0\}$.

Beispiel 1: Wir betrachten die Funktion $F : \mathbb{R} \to \mathbb{R}$ definiert durch $F(x) = 1$ für $x \geq 0$, sonst ist $F(x) = 0$. Es ist $\partial F(0) = \emptyset$.

Satz: Es sei x_0 derart, daß $F(x_0) < \infty$. Dann gilt

$$F(x_0) = \min_{x \in \mathcal{X}} F(x) \quad \Longleftrightarrow \quad 0 \in \partial F(x_0)$$

Beweis: Folgt aus der Definition. □
Analog zu $\partial F(x)$ läßt sich der Subgradient von F^* im Punkt x^* definieren. Es gilt $\partial F^*(x^*) \subset \mathcal{X}^{**}$. Tatsächlich kann es sein, daß für ein $x^{**} \in \partial F^*(x^*)$ auch $x^{**} \in \mathcal{X}$ gilt. Wir identifizieren dann x^{**} mit dem entsprechenden $x \in \mathcal{X}$.

Satz: Es gilt für alle $x \in \mathcal{X}$:

$$\begin{aligned} x^* \in \partial F(x) &\iff F(x) + F^*(x^*) = \langle x, x^* \rangle \\ x \in \partial F^*(x^*) &\iff F^{**}(x) + F^*(x^*) = \langle x, x^* \rangle \end{aligned}$$

Beweis: Die erste Zeile folgt aus folgender Äquivalenzkette:

$$\begin{aligned} x^* \in \partial F(x) &\iff F(y) \geq F(x) + \langle y - x, x^* \rangle,\ y \in \mathcal{X} \\ &\iff \langle x, x^* \rangle - F(x) \geq \langle y, x^* \rangle - F(y),\ y \in \mathcal{X} \\ &\iff \langle x, x^* \rangle - F(x) \geq F^*(x^*) \\ &\iff \langle x, x^* \rangle \geq F(x) + F^*(x^*) \end{aligned}$$

Die umgekehrte Ungl. $\langle x, x^* \rangle \leq F(x) + F^*(x^*)$ gilt stets, woraus die Behauptung folgt. □
Der Zusammenhang zwischen Subdifferential und der Eigenschaft von Funktionen konvex und unterhalbstetig zu sein, liefern folgende Sätze.

Satz: Es sei $\partial F(x) \neq \emptyset$. Dann gilt $F^{**}(x) = F(x)$ und

$$x^* \in \partial F(x) \implies x \in \partial F^*(x^*)$$

Beweis: Es sei $x^* \in \partial F(x)$. Aus $F^{**}(x) \geq \langle x, x^* \rangle - F(x)$ und $F^{**}(x) \leq F(x)$ folgt

$$\langle x, x^* \rangle \leq F^{**}(x) + F^*(x^*) \leq F(x) + F^*(x^*) = \langle x, x^* \rangle$$

Also gilt auch $F^{**}(x) + F^*(x^*) = \langle x, x^* \rangle$ und damit nach dem letzten Satz $x \in \partial F^*(x^*)$. □

Satz: Ist F konvex und unterhalbstetig, dann gilt

$$x^* \in \partial F(x) \iff x \in \partial F^*(x^*)$$

Der **Beweis** folgt aus den letzten Sätzen.

Der folgende Satz zeigt einen Zusammenhang zwischen Subdifferential und Minkowskisumme:
Satz: Es gilt

$$\partial F(x) + \partial G(x) \subset \partial(F + G)(x)$$

Sind darüber hinaus F und G konvex und beschränkt in der Nähe von x (dann sind sie stetig), dann gilt

$$\partial F(x) + \partial G(x) = \partial(F + G)(x)$$

3.5.2 Zusammenfassung. Rechenregeln. Sätze

- $\lambda > 0 \implies \partial(\lambda F)(x) = \lambda \partial F(x)$
- $\partial F(x) \neq \emptyset \implies F(x) = F^{**}(x)$
- $F(x) = F^{**}(x) \implies \partial F(x) = \partial F^{**}(x)$
- $F(x_0) = \min_{x \in \mathfrak{X}} F(x) \iff 0 \in \partial F(x)$
- $x^* \in \partial F(x) \iff F(x) + F^*(x^*) = \langle x, x^* \rangle$
- $\partial F(x)$ ist konvex und vage abgeschlossen
- $x^* \in \partial F(x) \implies x \in \partial F^*(x^*)$
- $x^* \in \partial F(x) \iff x \in \partial F^*(x^*)$, falls F regulär ist.
- F sei konvex und beschränkt in einer Umgebung von x, dann ist $\partial F(x) \neq \emptyset$.
-
-
-

Wir haben gesehen, daß konvexe Funktionen subdifferenzierbar sind. Jetzt betrachten wir den Zusammenhang zur üblichen Differenzierbarkeit im Banachraum. Die entsprechenden Definitionen dort werden oft auf die bekannten Definitionen von reellen Funktionen zurückgeführt.

3.5.3 Differentialquotienten reeller Funktionen

Es sei $f : \mathbb{R} \to \mathbb{R}$ konvex. Wir definieren eine Funktion $d : \mathbb{R} \to \mathbb{R}$ durch

$$d(t) = \frac{1}{t}\big(f(x+t) - f(x)\big)$$

Es gilt folgender
Satz: Auf $t \in (0, \infty)$ ist $d(t)$ monoton wachsend. Außerdem gilt dort $d(-t) \leq d(t)$.
Beweis: Es sei $0 < s < t$ und

$$x + s = \frac{t-s}{t}\ x + \frac{s}{t}\ (x+t)$$

eine konvexe Kombination der Punkt x und $x + t$. Da f konvex ist, folgt

$$f(x+s) \leq \frac{t-s}{t} f(x) + \frac{s}{t} f(x+t)$$

und damit

$$d(s) = \frac{1}{s}\big(f(x+s) - f(x)\big) \leq \frac{1}{t}\big(f(x+t) - f(x)\big) = d(s)$$

Außerdem folgt für $t > 0$ aus der Konvexität von f

$$f(x) \leq \frac{1}{2} f(x-t) + \frac{1}{2} f(x+t)$$

und hieraus

$$d(-t) = \frac{1}{-t}\big(f(x-t) - f(x)\big) \leq \frac{1}{t}\big(f(x+t) - f(x)\big) = d(t)$$

3.5.4 Differentialgeometrie für Funktionen $F : \mathcal{X} \to \mathbb{R}$

Für subdifferenzierbare Funktionen F gilt

$$F(x') - F(x) \geq \langle x' - x, x^* \rangle,\ x' \in \mathcal{X} \tag{24}$$

Wir können aus der linken Seite einen reellen Differenzenquotienten machen, indem wir $y - x$ aus einem festen Unterraum wählen. Wir wählen deshalb ein $y \in \mathcal{X}$ und setzen $x' = x + ty$ mit reellem $t > 0$. Das ergibt

$$\frac{1}{t}\big(F(x+ty) - F(x)\big) \geq \langle y, x^* \rangle \tag{25}$$

Die rechte Seite hängt nicht von t ab. Man kann also den Zusammenhang des Grenzwertes $t \to +0$ auf der linken Seite mit der rechten Seite untersuchen.
Die **Richtungsableitung** einer Funktion $F : \mathcal{X} \to \mathbb{R}$ am Punkt x entlang y ist definiert durch den Limes

$$D_y F(x) = \lim_{t \to +0} \frac{F(x+ty) - F(x)}{y}$$

falls dieser existiert.
Es gilt folglich

$$D_y F(x) \geq \langle y, x^* \rangle,\ x^* \in \partial F(x) \tag{26}$$

Für $x' \to x$ geht (68) in eine Gleichheit über. Das würde man auch für (26) erwarten, wenn der Subgradient $\partial F(x)$ aus einem Element bestünde. Gilt in (26) Gleichheit, dann sieht man, daß $D_y F(x)$ linear von der Richtung y abhängt.
Falls ein lineares beschränktes Funktional $x^* \in \mathfrak{X}^*$ existiert (das von x abhängt), sodaß

$$D_y F(x) = \langle x^*, y \rangle,\ y \in \mathfrak{X}$$

dann heißt $x^*(x) = F'(x) \in \mathfrak{X}^*$ **Gateauxableitung** von F in x. f heißt in x **Gateaux-diff.bar**.
Satz: Falls $F'(x)$ existiert, dann gilt $\partial F(x) = \{F'(x)\}$.
Falls F in x eine Gateauxableitung $F'(x)$ besitzt und außerdem

$$\lim_{0 \neq y \to 0} \frac{1}{\|y\|} \big| F(x+y) - F(x) - \langle y, F'(x) \rangle \big| = 0$$

gilt, dann heißt $F'(x)$ die **Frechet-Ableitung** von F in x