9 Mathematische Grundlagen. Konvexe Analysis

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.
Aus Sicht der Mengen besteht die Idee der konvexen Analysis darin, anstelle einer konvexen Menge die Menge ihrer Tengenten (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, daß der Durchschnitt einer beliebigen Schar abgeschlossener und konvexer Mengen wieder eine abgeschlossene und konvexe Mengen ist.

9.1 Konvexe Mengen. Konvexe Hülle und extremale Elemente

Wir beschreiben hier zwei grundlegende Begriffe der Konvexitätstheorie: die konvexe Hülle und die extremalen Elemente.
Ein wichtiger Begriff ist die konvexe Kombination von Punkten, die eine spezielle lineare Kombination ist. Damit ist klar, daß man von konvexen Mengen nur in einem linearen Raum sprechen kann.
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** (oder Extremalpunkte) einer konvexe Menge A, $\text{extr}(A)$ sind die Elemente der Menge A, die sich nur durch konvexe Kombinationen von Elementen aus A mit trivialen Gewichten darstellen lassen. Wir schreiben häufig $A_e = \text{extr}(A)$.

Genauer: Es sei $C \subset \mathcal{X}$ eine konvexe Menge in einem linearen Raum. Ein Element $x \in C$ heißt extremal, wenn aus $x = \alpha x_1 + (1-\alpha)x_2$, $x_1, x_2 \in C$, $0 < \alpha < 1$ folgt: $x_1 = x_2$.

Das ist so zu verstehen: Die Annahme, daß sich x als konvexe Kombination zweier verschiedener Elemente darstellen läßt, schlägt fehl.

- Den Begriff "konkave Menge" gibt es nicht.

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: ÜA 30)
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 $\alpha \le 1/2$. 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 $1 - 2\alpha \ge 0$ 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 $\alpha \ge 1/2$. 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 $2\alpha - 1 \ge 0$, $2 - 2\alpha \ge 0$ 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. $\square$

9.1.1 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 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`

9.1.2 Bemerkungen

- Intuitiv könnte man annehmen, daß die extremalen Elementen einer konvexen Menge soetwas wie "der Rand der Ränder" der Menge ist, so wie es beim konvexen Polyeders im $\mathbb{R}^n$ der Fall ist. Im allgemeinen ist das nicht so.
- Die konvexe Hülle einer abgeschlossenen Menge muß nicht abgeschlossen sein. Ein Gegenbeispiel ist die konvexe Menge einer Geraden und eines Punktes außerhalb der Geraden in $\mathbb{R}^2$.
- Die extremalen Elementen einer abgeschlossenen konvexen Menge muß keine abgeschlossenen Menge sein. Ein Gegenbeispiel ist der Doppelkegel aus der Aufgabe.
- Die offene Kugel im $\mathbb{R}^n$ ist konvex, besitzt aber keine extremalen Elementen.
- Die Begriffe "konvexe Hülle" und "extremalen Elemente" sind in gewissem Sinn invers zueinander. Es kann vorkommen, daß der besondere Fall

  $$C = \mathrm{conv}(A),\ A = \mathrm{extr}(C) \tag{33}$$

  das heißt

  $$A = \mathrm{extr}(\mathrm{conv}(A)),\ C = \mathrm{conv}(\mathrm{extr}(C)) \tag{34}$$

  eintritt. Beim konvexen Polyeders C im $\mathbb{R}^n$ ist z.B. $C =$conv(extr(C)). der Fall.

  Die Umkehrung muß nicht der Fall sein: Wir betrachten im $\mathbb{R}^2$ drei Punkte, die ein Dreieck bilden und einen Punkt im Inneren des Dreiecks. Dann ist die konvexe Hülle dieser vier Punkte das Dreieck. Die extremalen Elemente des Dreiecks sind aber nur die drei Eckpunkte.

  Allgemein gilt beim konvexen Polyeders C im $\mathbb{R}^n$: $C =$conv(extr(C)) und $A \supset$extr(conv(C)).

  Es ist klar, daß (33) bzw. (34) in gewissem Sinne erstrebenswert ist. A ist soewas wie die Basis von C. Man hat mit A genau die Punkte gefunden, die C generieren.

  Grob gesagt gilt: Falls $C \neq \mathrm{conv}(\mathrm{extr}(C))$, dann ist C nicht "eckig" und "abgeschlossen" genug. Falls $A \neq \mathrm{extr}(\mathrm{conv}(A))$, hat man zu viele Elemente in A gewählt um C zu bilden.

9.1.3 Übungsaufgaben

ÜA Bestimme die konvexe Hülle einer Geraden und eines Punktes außerhalb der Geraden im $\mathbb{R}^2$.

Lösung: 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.

ÜA Bestimme die extremalen Elemente folgender Menge im $\mathbb{R}^3$: Wir betrachten zwei identische schiefe, rechtwinklige Kegel und setzten sie an den Grundflächen aneinander.

Lösung: Die Menge der Extremalpunkte dieses Doppelkegels besteht aus den Kegelspitzen und allen Punkten der Grundfläche ohne dem Schnittpunkt der Grundflächemit der Verbindungsstrecke der beiden Kegelspitzen.

9.1.4 Konvexe Mengen in $\mathcal{C}$, $\mathcal{C}^*$ und $\mathcal{C}^{**}$

$$\begin{aligned}
\mathcal{R}(\mathcal{Z}) &= \left\{ g \in \mathcal{C}(\mathcal{Z}) \;\middle|\; g \geq 0,\ \|g\| \leq 1 \right\} \\
\mathcal{P}(\mathcal{Z}) &= \left\{ p \in \mathcal{C}^*(\mathcal{Z}) \;\middle|\; p \geq 0,\ \|p\| = 1 \right\} \\
\mathcal{Q}(\mathcal{Z}) &= \left\{ \xi \in \mathcal{C}^{**}(\mathcal{Z}) \;\middle|\; \xi \geq 0,\ \|\xi\| \leq 1 \right\}
\end{aligned}$$

In $\mathcal{R}(\mathcal{Z})$ gibt es im allgemeinen, z.B. wenn $\mathcal{Z}$ nur aus einer Zusammenhangskomponetne besteht, bis auf 1 und 0 keine extremalen Elemente, denn für alle $g \in \mathcal{S}(\mathcal{Z})$ gilt

$$g = \frac{1}{2}\Big(2g - g^2\Big) + \frac{1}{2}g^2,$$

wobei der erste Summand wegen $2g - g^2 = 1 - (1-g)^2$ und der zweite offensichtlich in $\mathcal{S}(\mathcal{Z})$ liegt. Extremale Elemente dürfen nur die Werte 0 und 1 annehmen. Es gilt also

$$\begin{aligned}
\mathcal{R}_e(\mathcal{Z}) &= \Big\{ 0,\ 1 \Big\} \\
\mathcal{P}_e(\mathcal{Z}) &= \left\{ \delta_z \;\middle|\; z \in \mathcal{Z} \right\} \\
\mathcal{Q}_e(\mathcal{Z}) &= \left\{ \chi_B \;\middle|\; B \in \mathcal{B}(\mathcal{Z}) \right\}
\end{aligned}$$

$\mathcal{R}(\mathcal{Z})$ ist abgeschlossen, aber nicht schwach abgeschlossen. Der schwache Abschluß von $\mathcal{R}(\mathcal{Z})$ ist $\mathcal{Q}(\mathcal{Z})$.

9.2 Konvexe Funktionen. Ungleichungen der Konvexitätstheorie

9.2.1 Konvexe Funktionen

Wir bezeichnen im Weiteren mit

$$\mathcal{P}_n = \{p \in \mathbb{R}^n | p_i \geq 0, \sum_{i=1}^n p_i = 1\}$$

den $(n-1)$-dimensionalen Simplex der Gewichte in $\mathbb{R}^n$.
Für einen beliebigen Vektor $g \in \mathcal{X}^n$, bezeichnen wir mit

$$\langle g, p \rangle = \sum_{i=1}^n g_i p_i$$

den Mittelwert von g bezüglich p.
Mit $\mathbb{1}$ bezeichnen wir den konstanten Vektor $\mathbb{1}_i = 1$. Es ist also $\langle \mathbb{1}, p \rangle = 1$.
Es sei $C \in \mathcal{X}$ eine konvexe Menge in einem normierten Raum $\mathcal{X}$. Ein Funktion $F : C \to \mathbb{R}$ heißt **konvex**, wenn ihr **Epigraph**
eine konvexe Menge in $\mathcal{X} \times \mathbb{R}$ ist.
Ein Funktion $F : C \to \mathbb{R}$ heißt **konkav**, wenn $-F$ konvex ist.
Es seien $x_1, x_2 \in C$ und $\overline{x} = \langle x, p \rangle$ mit einem gewissen $p \in \mathcal{P}_2$. Es sei $S(x)$ die Gerade (Sekante) durch die Punkte $(x_1, F(x_1))$ und $(x_2, F(x_2))$. Dann ist offensichtlich $F : C \to \mathbb{R}$ genau dann konvex, wenn $F(\overline{x}) \leq S(\overline{x})$ für beliebige $x \in \mathcal{X}^2$ und $p \in \mathcal{P}_2$.

Ausgeschrieben bedeutet das

$$F(p_1g_1 + p_2g_2) \leq p_1F(g_1) + p_2F(g_2)$$

Im weiteren sei $C = [a, b] \subset \mathbb{R}$ ein abgeschlossenes Intervall, $F : C \to \mathbb{R}$ eine konvexe Funktion, $p \in \mathcal{P}_n$, $g_i \in C$, $i = 1, ..., n$. $g_{\min} = \min_i g_i$, $g_{\max} = \max_i g_i$ und $F(g) = (F(g_1), ..., F(g_n))$.
Es gelten folgende Ungleichungen:

9.2.2 Mittelungleichung (MU)

Lemma: Mittelungleichung (MU):

$$g_{\min} \leq \langle g, p \rangle \leq g_{\max}$$

Beweis: trivial

9.2.3 Jensensche Ungleichung (JU)

Lemma: Jensensche Ungleichung (JU):

$$F\big(\langle g, p \rangle\big) \leq \langle F(g), p \rangle$$

Beweis: mit vollständiger Induktion über n.
IA: Für $n = 2$ ist die JU äquivalent zur Definition der Konvexität.
IV: Wir nehmen an, für n gilt die JU.

$$F\left(p_1g_1 + ...p_ng_n\right) \leq p_1F(g_1) + ...p_nF(g_n)$$

IB: Wir nehmen an, für $n + 1$ gilt die JU ebenfalls.

$$F\left(p_1g_1 + ...p_ng_n + p_{n+1}g_{n+1}\right) \leq p_1F(g_1) + ...p_nF(g_n) + p_{n+1}F(g_{n+1})$$

IS: Nach IA gilt JU für $n = 2$. Folglich gilt

$$F\left((1 - p_{n+1})\frac{p_1g_1 + ...p_ng_n}{1 - p_{n+1}} + p_{n+1}g_{n+1}\right) \leq (1 - p_{n+1})F\left(\frac{p_1g_1 + ...p_ng_n}{1 - p_{n+1}}\right) + p_{n+1}F(g_{n+1})$$

Wir setzen $q_i = p_i/(1 - p_{n+1})$. Dann ist $q = (q_1, ..., q_n) \in \mathcal{P}_n$ und nach IV gilt

$$F\left(\frac{p_1g_1 + ...p_ng_n}{1 - p_{n+1}}\right) = F\left(q_1g_1 + ...q_ng_n\right) \leq q_1F(g_1) + ... + q_nF(g_n)$$

Zusammen mit (**??**) ergibt das die IB. □

Bemerkung: Gleichheit gilt in der JU in einem von folgenden Fällen:

- $F(x)$ ist affin für $x \in [g_{\min}, g_{\max}]$.
- g ist Vielfaches von $\mathbb{1}$ (d.h. $g_i = g_j\ \forall i, j$)
- Für ein i gilt $p_i = 1$.

Hinzu kommen noch verschiedene Kombinationen wie $p_i = 0$ für gewisse i und g ist in den anderen Indizes konstant.
Bemerkung: Die JU ist in folgendem Sinn äquivalent zur Konvexität von F: Falls JU für beliebige g und p gilt, dann ist F konvex.
Ein interessantes Gegenbeispiel hierzu ist der häufig auftretende Fall, daß man nur solche g mit konstantem Mittelwert $g_0 = \langle g, p \rangle$ betrachtet. Dann kann JU gelten, obwohl F nicht konvex ist.

9.2.4 Äquivalenz der Ungleichungen

Lemma: Folgende Aussagen sind äquivalent:

i) $p \in \mathcal{P}_n$

ii) Es gilt MU $\forall g$

iii) Es gilt JU $\forall g$, $\forall F$

Beweis:
i) $\Longrightarrow$ *ii)* und *i)* $\Longrightarrow$ *iii)* wurden bereits bewiesen. Wir beweisen *ii)* $\Longrightarrow$ *i)* und *iii)* $\Longrightarrow$ *ii)*.
ii) $\Longrightarrow$ *i)*: Da MU für alle g gelten soll, wählen wir $g = e_i$ den Einheitsvektor. Dann ist $g_{\min} = 0$ und aus der MU folgt $p_i \geq 0$. Wir wählen $g = \mathbb{1}$, dann ist $g_{\min} = g_{\max} = 1$ und es folgt $\langle \mathbb{1}, p \rangle = 0$, also $p \in \mathcal{P}_n$.
iii) $\Longrightarrow$ *ii)*: Zu zeigen ist $g_{\min} \leq \langle g, p \rangle$ und $\langle g, p \rangle \leq g_{\max}$.
Zum indirekten Beweis von $\langle g, p \rangle \leq g_{\max}$ nehmen wir an, daß es ein g mit $\langle g, p \rangle > g_{\max}$ gibt. Wir wählen c so, daß $\langle g, p \rangle > c > g_{\max}$.
Da JU für alle konvexen F gelten soll, wählen wir $F = F_c$ und erhalten

$$0 = F_c\big(\langle g, p \rangle\big) \leq \langle F_c(g), p \rangle = (g_{\max} - c)\langle \mathbb{1}, p \rangle = g_{\max} - c$$

folglich $c \leq g_{\max}$, ein Widerspruch zur Konstruktion von c.
Zum Beweis der Ungleichung $g_{\min} \leq \langle g, p \rangle$ wählen wir eine entsprechende konvexe Funktion, geeignet gespiegelt.
Damit ist das Lemma bewiesen. □

Bemerkung: Dieses Lemma entspricht der Äquivalenz zwischen Wahrscheinlichkeitsmaßen und mittelnden Funktionalen (siehe Punkt 7.2.1 auf S. 123).

9.3 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: Eien Funktion $p : \mathcal{X} \to \mathbb{R}$ 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

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

Es sei $G_{x^*,\alpha} = \{x \in \mathcal{X} \mid \langle x, x^* \rangle = \alpha\}$ eine Hyperebene in $\mathcal{X}$.
Satz (Hahn-Banach) (geometrische Formulierung)
Es sei $A \subset \mathcal{X}$ offen und konvex. Weiter sei $M \subset \mathcal{X}$ ein affiner Unterraum und $A \cap M = \emptyset$. Dann existiert ein Funktional $x^* \in \mathcal{X}^*$ 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 $\mathcal{X}$, $M \subset \mathcal{X}$ ein linearer Unterraum und $\tilde{x}^*$ ein auf M definiertes lineares Funktional mit $\langle x, \tilde{x}^* \rangle \le p(x)$ für alle $x \in M$. Dann existiert ein $x^* \in \mathcal{X}^*$ mit $x^* = \tilde{x}^*$ auf M und $\langle x, x^* \rangle \le p(x)$ für alle $x \in \mathcal{X}$.
Der Satz von Hahn-Banach hat viele wichtige Folgerungen:
Folgerung 1: Es seien $A, B \subset \mathcal{X}$ offen, konvex und disjunkt. Dann existiert eine $G_{x^*,\alpha}$, die A und B trennt.
Folgerung 2: Es seien $A, B \subset \mathcal{X}$ 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.
Der Durchschnitt einer beliebigen Menge konvexer Mengen ist konvex. Der Durchschnitt einer beliebigen Menge abgeschlossener Mengen ist abgeschlossen. Folglich ist das Supremum einer beliebigen Menge von affine Hyperebenen einen konvexe unterhalbstetige Funktion. Äquivalent ist das Supremum einer beliebigen Menge affiner Funktionen konvex und unterhalbstetig.
Tatsächlich gilt auch die Umkehrung:
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

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

9.4 Konvexe Dualitätstheorie

Wie bereits erwähnt, ist eine Idee der Konvexitätstheorie, anstelle einer konvexen Funktion die Menge ihrer Tangenten zu betrachten. In gewissem Sinne sind eine Funktion und ihre Tangenten dual zueinander.

9.4.1 Die Legendretransformation

Wir betrachten zu einer gegebenen Funktion $F(x)$ eine "unter ihr liegende" affine Funktion (Hyperebene) $E(x) = \langle x, x^* \rangle - \alpha$. Für gegebenen "Anstieg" x^* und ausreichend großem α sollte das möglich sein. D.h., wir nehmen an, daß es ein α gibt, sodaß

$$\begin{aligned} F(x) &\ge E(x) = \langle x, x^* \rangle - \alpha \quad \Longleftrightarrow \\ \alpha &\ge \langle x, x^* \rangle - F(x) \end{aligned}$$

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

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

$F^* : \mathcal{X}^* \longrightarrow \overline{\mathbb{R}}$ 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, sie also berüher t.

$F^*(x^*)$ 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

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

9.4.2 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. Die Transformation $F \to F^*$ wird **Legendre-Fenchel-Transformation** (kurz Legendretransformation) genannt.
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 ist, dann ist $F^{**}(x) = F(x)$, denn das Supremum einer solchen Schar ist identisch mit dem Supremum über die Schar $(\langle x, x^* \rangle - F^*(x^*))_{x^* \in \mathcal{X}^*}$, denn $\langle x, x^* \rangle - F^*(x^*)$ ist gerade die größte aller Scharen $\langle x, x^* \rangle - \alpha$ die unter $F(x)$ liegen.

9.4.3 Dualität

Die Eigenschaft $F^{**}(x) = F(x)$ ist eine besonders schöne Eigenschaft, die die Legendretransformation zu einer Involution macht. Es ist deshalb sinnvoll, die Menge der Funktionen zu beschreiben, die diese Eigenschaft haben, also die, die sich als Supremum einer Schar von unter ihr liegenden affinen Hyperebenen beschreiben lassen.

9.5 Konvexe Funktionen in $\mathbb{R}$

9.5.1 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.

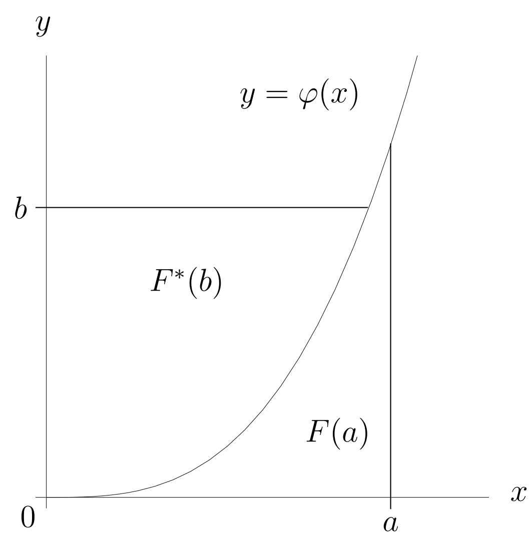
Bemerkung: Die Legendretransformation spielt eine herausragende Rolle in der Physik. Leider wird sie in den meisten Physikbüchern als $F^*(y) = y\varphi^{-1}(y) - F\big(\varphi^{-1}(y)\big)$ mit $F'(x) = \varphi(x)$ definiert. Hier ist der geometrische Sinn als Supremum über affine Funktionen nur noch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

9.5.2 Die Youngsche Ungleichung
Geometrische Bedeutung der Legendretransformation

Aus der Definition von F^* folgt die offensichtliche Ungleichung

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

Sie wird Youngsche Ungleichung genannt. An der Stelle $y = \varphi(x)$ oder $x = \varphi^{-1}(y)$ wird sie zur Gleichheit.
Diese Identität und die Youngsche Ungleichung haben eine einfache geometrische Interpretation. Es sei o.B.d.A. $\varphi(0) = 0$.



Es ist klar, daß an der Youngsche Ungleichung 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.

9.5.3 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 $C = 0$.

9.5.4 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

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

9.5.5 Beispiele

$f(x)$	dom(f)	$f^*(x^*)$	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^*
$\frac{\|x\|^p}{p}(p > 1)$	$\mathbb{R}$	$\frac{\|x^*\|^q}{q}(\frac{1}{p} + \frac{1}{q} = 1)$	$\mathbb{R}$
$\frac{-x^p}{p}(0 < p < 1)$	$\mathbb{R}_+$	$\frac{-(-x^*)^q}{q}(\frac{1}{p} + \frac{1}{q} = 1)$	$\mathbb{R}_{--}$
$\sqrt{1+x^2}$	$\mathbb{R}$	$-\sqrt{1-(x^*)^2}$	$[-1, 1]$
$-\log(x)$	$\mathbb{R}_{++}$	$-(1 + \log(-x^*))$	$\mathbb{R}_{--}$
e^x	$\mathbb{R}$	$\begin{cases} x^* \log(x^*) - x^* & \text{if } x^* > 0 \\ 0 & \text{if } x^* = 0 \end{cases}$	$\mathbb{R}_+$
$\log(1 + e^x)$	$\mathbb{R}$	$\begin{cases} x^* \log(x^*) + (1-x^*)\log(1-x^*) & \text{if } 0 < x^* < 1 \\ 0 & \text{if } x^* = 0, 1 \end{cases}$	$[0, 1]$
$-\log(1 - e^x)$	$\mathbb{R}$	$\begin{cases} x^* \log(x^*) - (1+x^*)\log(1+x^*) & \text{if } x^* > 0 \\ 0 & \text{if } x^* = 0 \end{cases}$	$\mathbb{R}_+$

9.5.6 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)

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

Das ist wieder die Youngsche Ungleichung.

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