Convex Analysis

Gert Wanka
Chemnitz University of Technology

Contents

1 Convex Functions

Convex functions (functionals) **play an important role** in **many fields** of **mathematics** such as **optimization, control theory, operations research, geometry, differential equations, functional analysis** etc., as well as in **applied sciences** and **practice**, e.g. in **economics**, **finance**.

They have a lot of **interesting** and **fruitful properties, e.g. continuity** and **differentiability** properties or the fact that a **local minimum turns out** to be a **global minimum** etc. They even allow to **establish** a **proper** and **general theory of convex functions**, moreover **together** with **convex sets**, the so-called theory of **Convex Analysis**, which is **important** not only for **itself** but also for its many **applications** in the theory of **convex** and also **non-convex optimization**.

Within this lecture on Convex Analysis we **do not want** to **develop** various **basic facts** on **convex sets**, because our **intention** is more to come **faster** to the relevant and **essential results** for **convex functions**. They will only be **summarized without proofs**, sometimes at the point where we need them within the representation of the lecture, as so-called "Standard Preliminaries". One could also see them as material of an Appendix.

The **main Chapters** of our **lecture** are devoted to **convex functions** (functionals, respectively), so-called **conjugate functions, convex optimization, subdifferential calculus, duality assertions** and **variational equalities** supplemented by some **special topics (Cones, Convex Processes, saddle-functions etc.)**.

The exercises proposed to the reader within this lecture are solved in the Appendix.

Definition 1.1. *Let X be a given linear space (we consider only real linear spaces), denote $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$ and consider a function $f : X \to \overline{\mathbb{R}}$.*

(a) **f** *is called* **convex** *if (and only if)*

$$f\big((1-\lambda)x_1 + \lambda x_2\big) \leq (1-\lambda)f(x_1) + \lambda f(x_2) \quad \textit{(Jensen's inequality)}$$

for all $x_1, x_2 \in X$ and $\lambda \in (0,1)$ for which the right-hand side is meaningful *(i.e. x_1, x_2 with $f(x_1)$ and $f(x_2)$ simultaneously infinite with opposite signs are not to be considered). If Jensen's inequality is fulfilled without equality for all $x_1 \neq x_2$, then the function f is called* **strictly convex**. *We call the function f* **concave** *if $-f$ is convex.*

(b) *The set defined by*

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

is called the **effective domain** *of f (cf. Figure 1.1).*

(c) *The* **epigraph** *of the function f is the set (cf. Figure 1.1)*

$$\operatorname{epi} f = \{(x, \alpha) \in X \times \mathbb{R} : f(x) \leq \alpha\}.$$

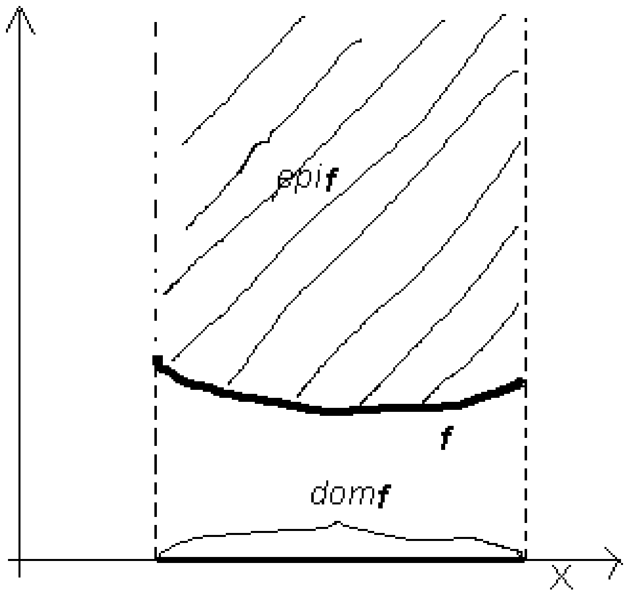

Figure 1.1

(d) *The function f is said to be* **proper** *if $\operatorname{dom} f \neq \emptyset$ and $f(x) > -\infty \ \ \forall x \in X$.*

Proposition 1.1. *Let X be a given linear space and the function $f : X \to \overline{\mathbb{R}}$. Then the following statements are valid.*

(a) *f is convex if and only if $\operatorname{epi} f$ is convex.*

(b) If f is convex then $\operatorname{dom} f$ *is convex.*

Proof.

(a) *Necessity.* Let f be convex and $(x_1, \alpha_1), (x_2, \alpha_2) \in \operatorname{epi} f$, that is $f(x_1) \le \alpha_1$ and $f(x_2) \le \alpha_2$. Let also be $\lambda \in (0, 1)$. From the convexity of f and the definition of the epigraph we have

$$\begin{aligned} f\big(\lambda x_1 + (1-\lambda)x_2\big) &\le \lambda f(x_1) + (1-\lambda) f(x_2) \\ &\le \lambda \alpha_1 + (1-\lambda)\alpha_2. \end{aligned}$$

This means that $\big(\lambda x_1 + (1-\lambda)x_2, \lambda\alpha_1 + (1-\lambda)\alpha_2\big) \in \operatorname{epi} f$ or $\lambda(x_1, \alpha_1) + (1-\lambda)(x_2, \alpha_2) \in \operatorname{epi} f$, i.e. $\operatorname{epi} f$ is convex.

Sufficiency. Let $\operatorname{epi} f$ be convex and this implies (cf. (b)) that $\operatorname{dom} f$ is convex. It is sufficient to verify Jensen's inequality over $\operatorname{dom} f$. Thus, let us take x_1, $x_2 \in \operatorname{dom} f$ and choose a, b such that $f(x_1) \le a$ and $f(x_2) \le b$, i.e. $(x_1, a), (x_2, b) \in \operatorname{epi} f$. By the assumptions follows $\big(\lambda(x_1, a) + (1-\lambda)(x_2, b)\big) \in \operatorname{epi} f$ for all $\lambda \in [0, 1]$ and this implies

$$f(\lambda x_1 + (1-\lambda)x_2) \le \lambda a + (1-\lambda) b.$$

If $f(x_1)$ and $f(x_2)$ are finite, one can take $a = f(x_1)$ and $b = f(x_2)$ to conclude the assertion of the proposition.

If either $f(x_1)$ or $f(x_2)$ is $-\infty$ one can let tend a or b to $-\infty$ and thus Jensen's inequality is also fulfilled.

(b) Clear, because $\operatorname{epi} f$ is convex and therefore $\operatorname{dom} f$ as its projection, too.$\square$

Remarks: Why do we allow the value $+\infty$ for proper functions?

(i) Let f be a functional $f : A \subset X \to \mathbb{R}$ and define $\tilde{f} : X \to \overline{\mathbb{R}}$ by

$$\tilde{f} = \begin{cases} f(x), & \text{if } x \in A, \\ +\infty, & \text{if } x \notin A. \end{cases}$$

The functional $\tilde{f}$ is convex if and only if A is convex and $f : A \to \mathbb{R}$ is convex. Therefore we need only to consider functions defined on X everywhere.

(ii) Let $A \subset X$. Define **the indicator functional of** A

$$\chi_A : X \to \overline{\mathbb{R}}, \quad \chi_A(x) = \begin{cases} 0, & \text{if } x \in A, \\ +\infty, & \text{if } x \notin A. \end{cases}$$

Then A is a convex subset if and only if χ_A is convex ($\operatorname{dom} \chi_A = A$).

Thus the study of convex sets can be reduced to the study of convex functions.

Remark: We do not prove trivialities within this lecture. Therefore we only mention shortly some simple facts about convex functions.

(i) If f is convex and $\lambda \geq 0$ then λf is convex.

(ii) If $f,\ g$ are convex then $f + g$ is convex.

(iii) If $(f_i)_{i \in I}$ is any family of convex functions from X into $\overline{\mathbb{R}}$, their pointwise supremum $f : X \to \overline{\mathbb{R}}$, $f(x) = \sup\limits_{i \in I} f_i(x)$ is convex, because $\operatorname{epi} f = \bigcap\limits_{i \in I} \operatorname{epi} f_i$ that is convex as intersection of convex sets (see Figure 1.2).

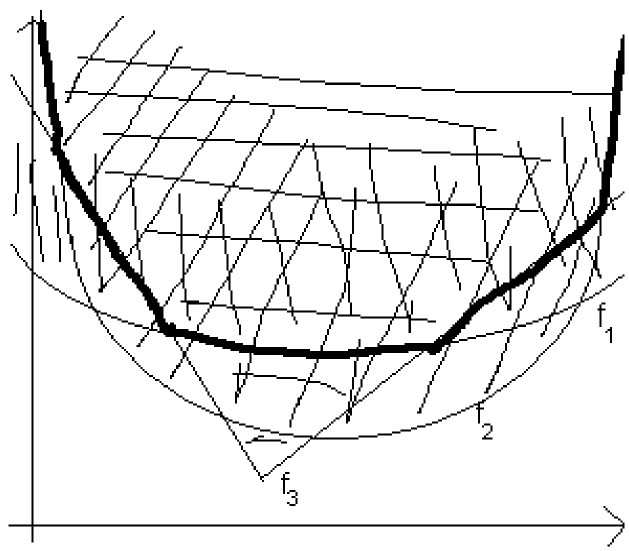

Figure 1.2

Examples 1.1.

(i) Let X be a real normed space. Then $f : X \to \mathbb{R}, f(x) = \|x\|^n$, $n \geq 1$ is a convex function. If n=1 we can apply the triangle inequality; the case $n > 1$ remains as an exercise for the reader.

(ii) Analogously for the function $f : X \to \mathbb{R}, f(x) = \|x - \bar{x}\|^n$, where $\bar{x}$ is a fixed point from X.

(iii) Let X be a reflexive real Banach space (i.e. $X^{**} = (X^*)^* \cong X$). By X^* we denote the **topological dual space** to X, i.e. the space of linear continuous functionals defined on X. The value of a functional $x^* \in X^*$ at a point $x \in X$ is usually denoted by $\langle x^*, x\rangle := x^*(x)$.

The linearity of $x^* \in X^*$ is obvious, for all $\alpha, \beta \in \mathbb{R}$ and $x_1, x_2 \in X$ it is sure that $\langle x^*, \alpha x_1 + \beta x_2\rangle = \alpha\langle x^*, x_1\rangle + \beta\langle x^*, x_2\rangle$.

When $X = \mathbb{R}^n$, by considering the value of the functional $x^* \in (\mathbb{R}^n)^*$ at some point $x \in \mathbb{R}^n$ we obtain actually the Euclidean scalar product $\langle x^*, x\rangle = \sum\limits_{i=1}^{n} x_i^* x_i$. If $x = (x_1, ..., x_n)^T \in \mathbb{R}^n$ we have $x^* = (x_1^*, ..., x_n^*)^T \in \mathbb{R}^n$ and that implies $\mathbb{R}^{n*} \cong \mathbb{R}^n$ and further $\mathbb{R}^{n**} \cong \mathbb{R}^n$.

For X **Hilbert space** it follows that $x^*(x) = \langle x^*, x\rangle$ is a **scalar product**. We have $x^* \in X^* \cong X$ and this leads to $X^{**} = X$.

Now (cf. above) let be X **a reflexive real Banach space** and let be $B : X \to X^*$ a linear bounded (continuous)(i.e. $B \in L(X, X^*)$) non-negative self-adjoint operator (mapping), i.e. $\langle Bx, x\rangle \geq 0\ \forall x \in X$ and $B^* : X^{**} = X \to X^*,\ B^* = B$.

Exercise 1.1.

(i) The function $f : X \to \mathbb{R}, f(x) = \langle Bx, x\rangle$ is convex.

(ii) If we consider $X = \mathbb{R}^n$, $B = (b_{ij})_{i,j=1,...,n}$ a $n \times n$ symmetric positive semi-definite matrix and the quadratic function $f(x) = \langle Bx, x\rangle = \langle x, Bx\rangle = \sum\limits_{i,j=1}^{n} b_{ij} x_i x_j \geq 0$, $x = (x_1, ..., x_n)^T$, then f is convex.

The following result is well-known from the basic lecture courses of analysis and optimization, respectively.

Proposition 1.2. *Let f be a twice continuously differentiable real-valued function on an open convex set C in $\mathbb{R}^n$. Then f is convex on C if and only if its Hessian matrix $Q_x = (q_{ij}(x))_{i,j=1,..,n}$, $x = (x_1, ..., x_n)^T$ where $q_{ij}(x) = \frac{\partial^2 f}{\partial x_i \partial x_j}(x_1, ..., x_n)$ is positive semi-definite for every $x = (x_1, ..., x_n)^T \in C$.*

Definition 1.2. *Let $f_1, ..., f_m$ be proper functions on a linear space X. Then we call the function*

$$f(x) := \inf\{f_1(x_1) + ... + f_m(x_m) : x_1 + ... + x_m = x, \ x_i \in X, \ i = 1, ..., m\}$$

infimal convolution, *usually denoted $f = f_1 \square f_2 \square ... \square f_m = \mathop{\square}\limits_{i=1}^{m} f_i$.*

Exercise 1.2. Show that the function f is convex. Hint: it comes from the fact that for only two functions f_1, f_2, after setting $x_2 \to y$ and $x_1 \to x - y$ there is

$$(f_1 \square f_2)(x) = \inf\{f_1(x - y) + f_2(y)\},$$

but this is analogous to the classical formula for integral convolution $(\int f_1(x - y) f_2(y) dy)$.

Examples 1.2.

(i) Considering the indicator function of the set {a}

$$\delta(x|a) = \begin{cases} 0, & \text{if } x = a, \\ +\infty, & \text{otherwise,} \end{cases}$$

then $f \square \delta(\cdot|a) = f(x - a)$ (translation of the graph of f horizontally by a). If $a = 0$ then $f \square \delta(\cdot|0) = f(x)$ (identity).

(ii) If $f(x) = \|x\|$ and $g(x) = \chi_C(x)$, $x \in X$ (the indicator function of the set C), then $(f \square g)(x) = \inf\limits_{y \in X} \{\|x - y\| + \chi_C(y)\} = \inf\limits_{y \in C} \|x - y\| = d(x, y)$, which is actually a distance function. This means that the distance function for a convex set C is a convex function.

Remark: There is another representation of $f_1 \square f_2$,

$$f_1 \square f_2 = \inf\{\mu : \exists x \in X \text{ such that } (x, \mu) \in (\operatorname{epi} f_1 + \operatorname{epi} f_2)\}.$$

In this way $f_1 \Box f_2$ **may be defined** even in the case when f_1 and f_2 are **not** necessarily **proper convex functions**, i.e. for any functions mapping from X to $[-\infty, +\infty]$.

The **operation "$\Box$"** has the following properties

- **commutativity:** $f_1 \Box f_2 = f_2 \Box f_1$;
- **associativity:** $f_1 \Box (f_2 \Box f_3) = (f_1 \Box f_2) \Box f_3$;
- **preserves the convexity**;
- $\delta(\cdot, 0)$ acts as its **identity element**.

(iii) **Support functional** of a convex set $C \subset X$

$$S_C(x) = S(x|C) = \sup_{y \in C} \langle x, y \rangle .$$

Geometrical interpretation. (cf. Figure 1.3)

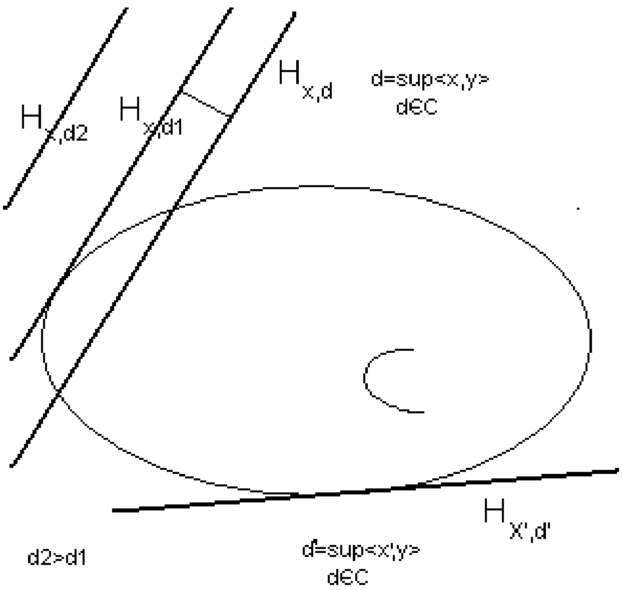

Figure 1.3

Consider a hyperplane in X, $H_{X,d} = \{y : \langle x, y \rangle = d\}$, with $d = \sup_{y \in C} \langle x, y \rangle$. The tangent hyperplanes are supporting the set C.

$S_C(x)$ is convex as the pointwise supremum of a certain collection of linear functions (i.e. convex) $\langle \cdot, y \rangle$ as y ranges over C. (Family of functions $f_y(x) = \langle x, y \rangle$, $y \in C$ which is an index set.)

(iv) A **gauge** of a set C is defined by

$$\begin{cases} \gamma_C(x) & = \gamma(x|C) = \inf\{\lambda : \lambda \geq 0, x \in \lambda C\}, \ x \in X, \\ \gamma_C(x) & = +\infty, \quad \text{if there is no } \lambda \geq 0 \text{ such that } x \in \lambda C, \end{cases}$$

and is also called **Minkowski (gauge) functional** (cf. Figure 1.4).

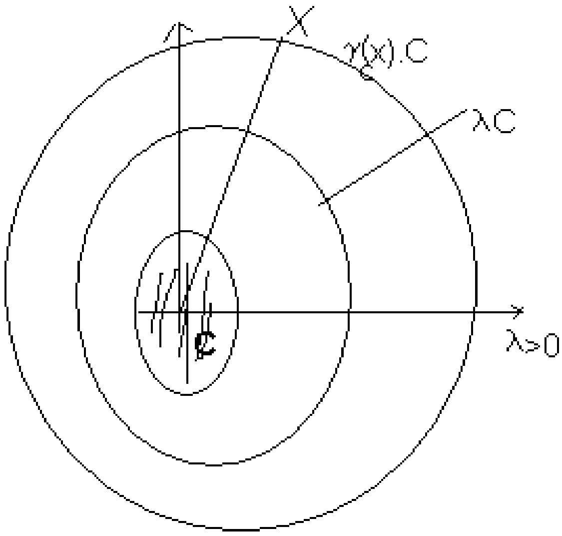

Figure 1.4

Exercise 1.3. If C is convex, then $\gamma_C(x)$ is convex, too.

Definition 1.3. *A functional* $p : X \to \bar{\mathbb{R}}$ *is said to be* **sublinear** *if*

(i) $p(tx) = tp(x)$ *whenever* $t \geq 0$ *(positive homogeneousity),*

(ii) $p(x + y) \leq p(x) + p(y)$ *(subadditivity).*

Remark: Of course, it is trivial to show that a sublinear functional is also convex.

In this sense a **sublinear functional** is a **generalization** of a **norm** in a linear space.

Remark: Usually, defining a gauge by a convex set C it is additionally supposed that $0 \in \mathrm{int}(C)$ (the interior of C).

Let be $C \subset X$, where X is a normed space (or more general a linear topological locally convex space). Then, for all $x \in X$ we have $\gamma_C(x) < \infty$. Otherwise there are $x \in X$ with $\gamma_C(x) = +\infty$ (cf. Figure 1.5, where there is no $\lambda \geq 0$ such that $x \in \lambda C$ and Figure 1.6).

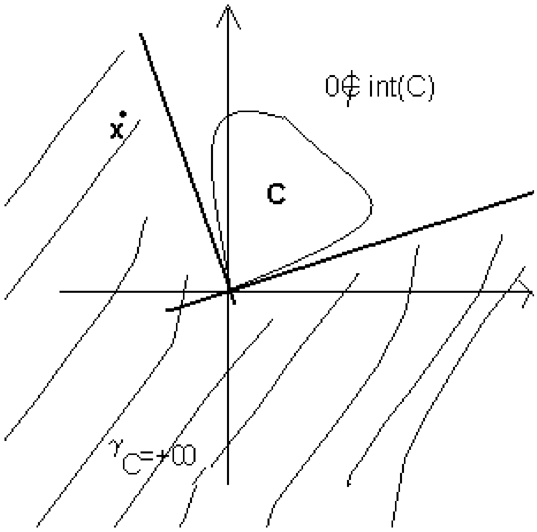

Figure 1.5

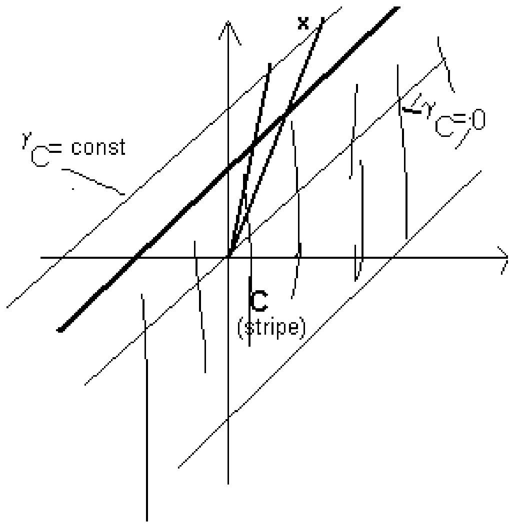

Figure 1.6

Exercise 1.4. Prove that γ_C is sublinear.

Exercise 1.5. If f is a positively homogeneous proper convex function, then the following statements are true.

(a) $f(\lambda_1 x_1 + ... + \lambda_m x_m) \leq \lambda_1 f(x_1) + ... + \lambda_m f(x_m)$, whenever $\lambda_1 > 0, ..., \lambda_m > 0$.

(b) $f(-x) \geq -f(x)$ for every $x \in X$.

The closure of a convex set C, $0 \in \mathrm{int}\, C$, can be described using its gauge

$$\bar{C} = \{x \in X : \gamma_C(x) \leq 1, \text{ i.e. } x \in \lambda_n C \text{ for a sequence } \lambda_n \to 1, \lambda_n > 1 \ \ \forall n \in \mathbb{N}\},$$

that coincides to C when it is closed, while the interior of C is

$$\mathrm{int}\, C = \{x \in X : \gamma_C(x) < 1\}$$

and its boundary

$$\partial C = \{x \in X : \gamma_C(x) = 1\}.$$

Thus C plays the role of the **unit ball** as for the case when the **gauge is a norm**

$$C = B_{0,1} = \{x \in X : \|x\| \leq 1\}$$

If from $x \in C$ follows $-x \in C$ for every $x \in C$ (C is symmetric), then $\gamma_C(-x) = \gamma_C(x)$ and therefore $\gamma_C(x) = \|x\|_C$ defines a norm if $0 \in \text{int } C$ and C is bounded because $\|\lambda x\|_C = |\lambda|\gamma_C(x)$ also for $\lambda < 0$. Of course $0 \in \text{int } C$ is supposed and C is bounded so $\gamma_C(0) = \|0\| = 0$ and $\gamma_C(x) > 0$, $\gamma_C(x) \neq \infty$ for $x \neq 0$ (cf. Figure 1.7).

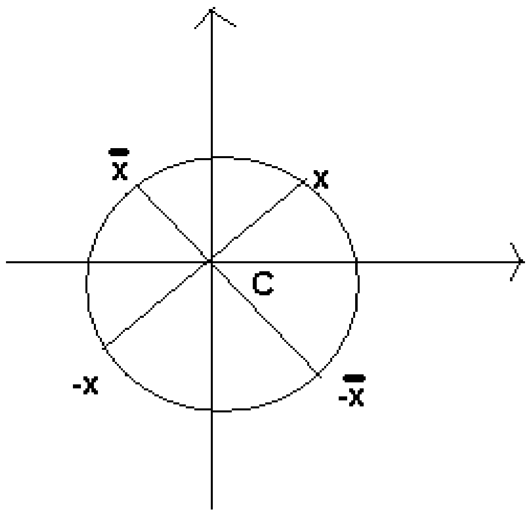

Figure 1.7

Let X be a **normed space** and $C \subset X$ **convex**, $0 \in \text{int } C$. Then there exists $M > 0$ such that

$$\gamma_C(x) \leq M\|x\| \quad \forall x \in X,$$

i.e. γ_C is a bounded functional. To point out that, take a ball with radius r centered at 0 and contained in C, $B_{0,r} = \{x \in X : \|x\| \leq r\} \subset C$ (this is possible because $0 \in \text{int } C$). Since $B_{0,r} \subset C$ it follows

$$\begin{aligned} \gamma_C(x) &= \inf\{\lambda : \lambda \geq 0, x \in \lambda C\} \\ &\leq \gamma_{B_{0,r}} = \inf\{\lambda : \lambda \geq 0, x \in \lambda B_{0,r}\} = \frac{\|x\|}{r}. \end{aligned}$$

If we take $M = \frac{1}{r}$ we obtain $\gamma_C(x) \leq M\|x\| \ \forall x \in X$. From that follows that $\gamma_C(x)$ is **continuous** (see Figure 1.8).

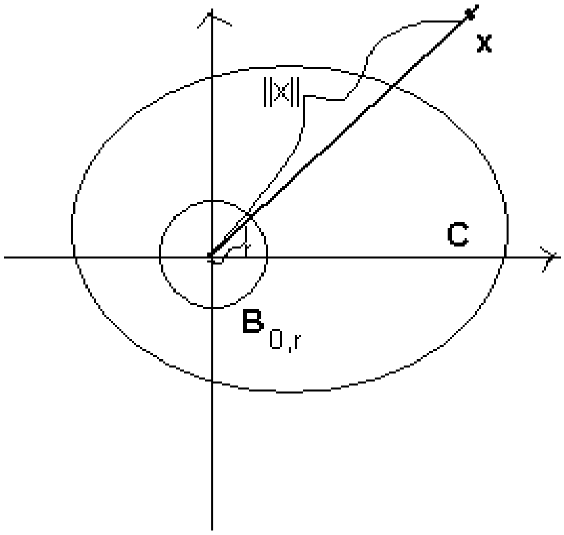

Figure 1.8

Let be the sequence $x_n \to x$ and we have $\gamma_C(x_n - x) \leq M\|x_n - x\| \to 0$ when $n \to \infty$. This does not prove yet the continuity, as we need $|\gamma_C(x_n) - \gamma_C(x)| \to 0$ when $n \to \infty$.

But as for a norm, we can also for a gauge (or more general for a sublinear function) estimate in an analogous manner

$$\gamma_C(x_n) = \gamma_C(x_n - x + x) \leq \gamma_C(x_n - x) + \gamma_C(x_n)$$

and from here

$$\gamma_C(x_n) - \gamma_C(x) \leq \gamma_C(x_n - x) \leq M\|x_n - x\|, n \in \mathbb{N}.$$

Moreover,

$$\gamma_C(x) = \gamma_C(x - x_n + x_n) \leq \gamma_C(x - x_n) + \gamma_C(x_n),$$

implying

$$\gamma_C(x) - \gamma_C(x_n) \leq \gamma_C(x - x_n) \leq M\|x - x_n\| = M\|x_n - x\|, n \in \mathbb{N}.$$

From the above inequalities we obtain

$$-M\|x_n - x\| \leq \gamma_C(x_n) - \gamma_C(x) \leq M\|x_n - x\|$$

which is nothing but the continuity of $\gamma_C(x)$,

$$|\gamma_C(x_n) - \gamma_C(x)| \leq M\|x_n - x\|, n \in \mathbb{N}.$$

Exercise 1.6. Conversely, any positively homogeneous, subadditive (i.e. sublinear), non-negative and continuous function p on X is of the form γ_C, i.e. is a gauge.

Lemma 1.1. *Let be X a real normed space, f a convex function over an open set D (f is considered on* $\operatorname{int}(\operatorname{dom} f)$*). If $x_0 \in D$, then for each $h \in X$ the "right hand" directional derivative*

$$d^+ f(x_0)(h) = \lim_{t \to 0^+} \frac{f(x_0 + th) - f(x_0)}{t}$$

exists and is a sublinear function on X.

Proof. We point out that the difference quotient is nonincreasing as $t \to 0^+$ and bounded below by the corresponding difference quotient from the left. Thus the limit exists.

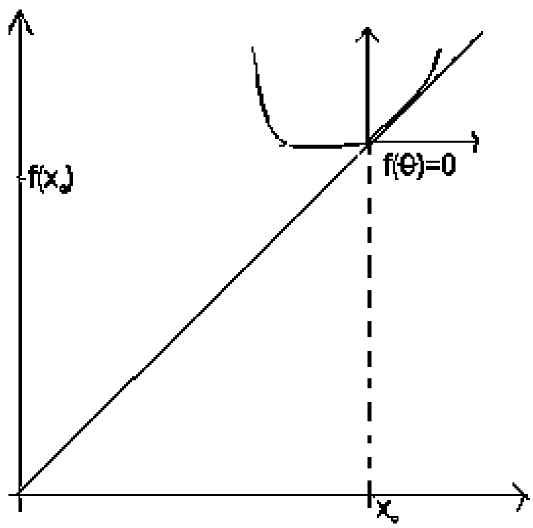

Figure 1.9

To prove this, we assume that $x_0 = 0$ and $f(x_0) = 0$ (possible by a translation of X and f, cf. Figure 1.9). Let be $0 < t < s$. By convexity, with $th = \frac{t}{s}(sh) + \left(1 - \frac{t}{s}\right) 0$, we have

$$f(th) \leq \frac{t}{s} f(sh) + \frac{s-t}{s} f(0) = \frac{t}{s} f(sh)$$

and so $\frac{1}{t} f(th) \leq \frac{1}{s} f(sh)$. This proves the monotonicity of the difference quotient

$$\frac{f(x_0 + th) - f(x_0)}{t} = \frac{f(th)}{t} \leq \frac{f(sh)}{s} = \frac{f(x_0 + sh) - f(x_0)}{s}.$$

We apply this to $(-h)$ in place of h (we see that $f(x_0 + t(-h)) = f(x_0 - th)$). It follows that $-\frac{f(x_0+th)-f(x_0)}{t}$ is nondecreasing as $t \to 0^+$. By convexity we have for $t > 0$

$$f(x_0) = f\left(\frac{1}{2}(x_0 - 2th) + \frac{1}{2}(x_0 + 2th)\right) \leq \frac{1}{2} f(x_0 - 2th) + \frac{1}{2} f(x_0 + 2th).$$

Therefore

$$2f(x_0) \leq f(x_0 - 2th) + f(x_0 + 2th).$$

Transforming this inequality and dividing by $2t$, we obtain

$$\frac{-[f(x_0 - 2th) - f(x_0)]}{2t} \leq \frac{[f(x_0 - 2th) - f(x_0)]}{2t}.$$

Because the left hand side is nondecreasing as $t \to 0^+$ and the right hand side is nonincreasing as $t \to 0^+$, this shows that the right hand side (the difference quotient) is bounded below and the left hand side is bounded above. So both limits exist.

The left limit is denoted by $-d^+ f(x_0)(-h)$ and it holds

$$-d^+ f(x_0)(-h) \leq d^+ f(x_0)(h) \quad \forall h \in X.$$

Obviously $d^+ f(h)$ is positively homogeneous (cf. definition of $d^+ f(h)$)

$$\begin{aligned}
d^+ f(x_0)(\lambda h) &= \lim_{t \to 0^+} \frac{f(x_0 + t\lambda h) - f(x_0)}{t} \\
&= \lim_{t \to 0^+} \lambda \frac{f(x_0 + (t\lambda) h) - f(x_0)}{t\lambda} \\
&= \lambda \lim_{\bar{t} \to 0^+} \frac{f(x_0 + \bar{t} h) - f(x_0)}{\bar{t}} \\
&= \lambda d^+ f(x_0)(h),
\end{aligned}$$

where $\lambda > 0$ and $\bar{t} = t\lambda$. The fact that $d^+f(h)$ is subadditive follows by convexity

$$\begin{aligned} f(x_0 + t(h_1 + h_2)) &= f\left(\frac{1}{2}(x_0 + 2th_1) + \frac{1}{2}(x_0 + 2th_2)\right) \\ &\leq \frac{1}{2}f(x_0 + 2th_1) + \frac{1}{2}f(x_0 + 2th_2). \end{aligned}$$

By subtracting from both hand sides $-f(x_0)$ and dividing by $t > 0$, we obtain

$$\frac{f(x_0 + t(h_1 + h_2)) - f(x_0)}{t} \leq \frac{f(x_0 + 2th_1) - f(x_0)}{2t} + \frac{f(x_0 + 2t(h_2) - f(x_0)}{2t}.$$

Taking limits as $t \to 0^+$, follows

$$d^+f(x_0)(h_1 + h_2) \leq d^+f(x_0)(h_1) + d^+f(x_0)(h_2).\Box$$

Definition 1.4. *If the functional $h \longmapsto d^+f(x_0)(h)$ is linear (instead of only sublinear) then the convex function is said to be* **Gateaux differentiable** *at $x_0 \in D$. The functional is in this case denoted by $df(x_0)(h)$ and $d(f(x_0))$ (as a linear functional, continuity is not assumed) is called the* **Gateaux derivative (differential)** *of f at x_0.*

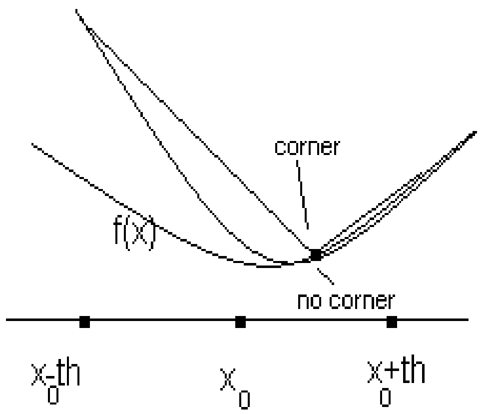

Figure 1.10

Remark: It is clear that f is Gateaux differentiable at x_0 if and only if

$$-d^+t(x_0)(-h) = d^+t(x_0)(h).$$

What does this equality mean? Let us see (cf. Figure 1.10)

$$\begin{aligned}
-d^+ f(x_0)(-h) &= - \lim_{t\to 0^+} \frac{f(x_0 - th) - f(x_0)}{-t} \\
&= \lim_{\bar{t}\to 0^-} \frac{f(x_0 + \bar{t}h) - f(x_0)}{\bar{t}} \\
&= \lim_{t\to 0^+} \frac{f(x_0 + th) - f(x_0)}{t} \\
&= d^+ f(x_0)(h),
\end{aligned}$$

i.e. the two-sided limit exists and so does the limit itself. Therefore we can say that the convex function f is **Gateaux differentiable** at $x_0 \in D$ if and only if

$$df(x_0)(h) = \lim_{t\to 0^+} \frac{f(x_0 + th) - f(x_0)}{t} = \frac{d}{df} f(x_0 + th)|_{t=0}$$

and the limit exists!

Exercise 1.7. Prove that a sublinear functional p is linear if and only if for all $x \in X$ it holds $p(-x) = -p(x)$.

2 Continuity of Convex Functions

Lemma 2.1. *If a convex function $f : X \to \overline{\mathbb{R}}$, with X a normed space (more general locally convex space), has a neighborhood of a point $x \in X$ where f is bounded above by a finite constant, then f is continuous at x.*

Proof. We reduce again the problem by translation to the case where $x = 0$ and $f(x) = 0$, too.

Let U be a neighborhood of the origin, such that $f(x) \leq c < \infty$ for all $x \in U$. It follows that $V = U \cap (-U)$ is a symmetric neighborhood of the origin.

Consider further $\varepsilon \in (0, 1)$ and $x \in \varepsilon\ V = \{y : \ y = \varepsilon v, \ v \in V\}$. Due to the convexity of f and because x is a convex combination of 0 and $\frac{1}{\varepsilon}x$, i.e.

$$x = (1-\varepsilon)0 + \varepsilon\left(\frac{1}{\varepsilon}x\right),$$

one has

$$f(x) \leq (1-\varepsilon)f(0) + \varepsilon f\left(\frac{1}{\varepsilon}x\right) \leq \varepsilon c, \tag{2. 1}$$

the last inequality taking place because $\frac{1}{\varepsilon}x \in V \subseteq U$ and so $f(\frac{1}{\varepsilon}x) \leq \varepsilon c$.

Analogously we obtain from

$$0 = \frac{1}{1+\varepsilon}x + \frac{\varepsilon}{1+\varepsilon}\left(-\frac{1}{\varepsilon}x\right),$$

i.e. 0 is a convex combination of x and $\left(-\frac{1}{\varepsilon}x\right) \in V \subseteq U$, because from $x \in \varepsilon V$ follows that $-x \in \varepsilon V$, V being symmetric, the following

$$f(0) \leq \frac{1}{1+\varepsilon}f(x) + \frac{\varepsilon}{1+\varepsilon}f\left(-\frac{1}{\varepsilon}x\right).$$

Multiplying the last relation by $(1+\varepsilon)$ we obtain

$$f(x) \geq (1+\varepsilon)f(0) - \varepsilon f\left(-\frac{1}{\varepsilon}x\right) \geq -\varepsilon c \tag{2. 2}$$

From (2.1) and (2.2) we have that $\mid f(x) \mid \leq \varepsilon c \ \ \forall x \in \varepsilon V$, and this means, by the definition of continuity, that f is continuous at $x = 0$. □

Proposition 2.1. *Let X be a normed space and $f : X \to \overline{\mathbb{R}}$ be a convex function. The following statements are equivalent to each other.*

(i) There is a non-empty open set M on which f is not everywhere equal to $-\infty$ and is bounded above by a constant $c < +\infty$ (i.e. $M \subset \operatorname{int}(\operatorname{dom} f)$ since M is open).

(ii) f is a proper function and it is continuous over the interior of its effective domain $\operatorname{int}(\operatorname{dom} f)$ which is non-empty.

Proof. It is clear that (ii) implies (i). Let us conclude (ii) from (i).

Let $M \subset \operatorname{int}(\operatorname{dom} f)$ and $x \in M$ such that $f(x) > -\infty$ (cf.(i)). Lemma 2.1 shows that f is continuous at x (observe that from M open follows that x is an inner point of M). Therefore, f is finite on a neighborhood U of x and hence proper.

As U we can choose a ball $B(x, \rho)$ around x with radius ρ. The reason is that a convex function which takes the value $-\infty$ at $\overline{x}$, $f(\overline{x}) = -\infty$, has the property that on every half-line starting from $\overline{x}$ either f is identically equal to $-\infty$ or f has the value $-\infty$ between $\overline{x}$ and a point $\widehat{x}$, any value at $\widehat{x}$, and $+\infty$ beyond $\widehat{x}$.

Thus, a function which is finite in a neighborhood of a point x can take nowhere the value $-\infty$. If it would be $-\infty$ at $\overline{x}$, i.e. $f(\overline{x}) = -\infty$, then connect $\overline{x}$ with x by a straight line and so arises a contradiction to the fact that f is finite on a neighborhood of x.

Let be $\overline{x} \in \operatorname{int}(\operatorname{dom} f)$ an arbitrary point. It follows that there exists $\delta > 1$ such that $y = x + \delta(\overline{x} - x) \in \operatorname{int}(\operatorname{dom} f)$. (cf. Figure 2.1)

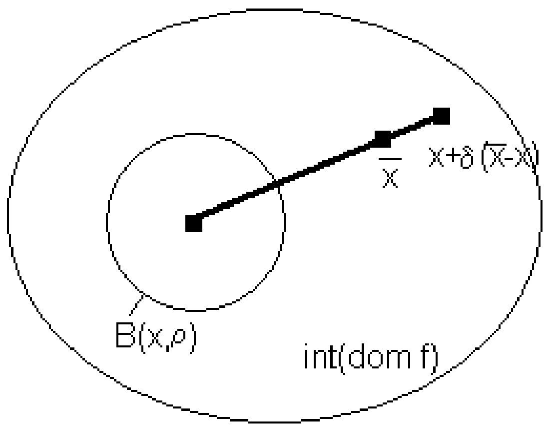

Figure 2.1

Then the closed ball $B(\overline{x}, (1-\frac{1}{\delta})\rho)$ is included in $\mathrm{int}(\mathrm{dom}\, f)$. A geometrical explanation follows.

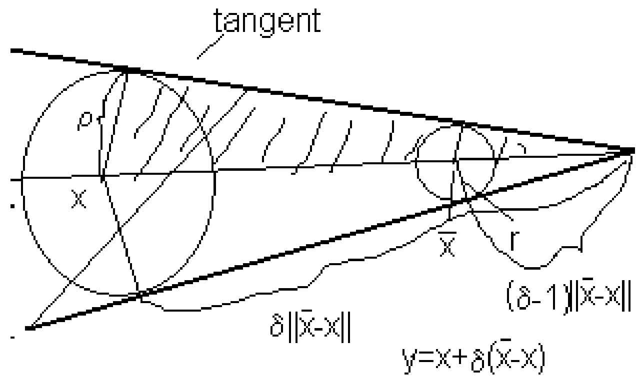

Figure 2.2

In Figure 2.2 one may notice some similar triangles, so it follows

$$\frac{\delta\|\bar{x}-x\|}{\rho} = \frac{(\delta-1)\|\bar{x}-x\|}{r}.$$

The value of the radius of the small ball is

$$r = \frac{\delta-1}{\delta}\rho = \Big(1-\frac{1}{\delta}\Big)\rho,$$

so $y = x + \delta\|\bar{x}-x\|$ belongs to $\mathrm{int}(\mathrm{dom}(f))$.

These considerations can be modified in order to hold also for locally convex topological vector spaces.

By convexity, let $x^0 \in B(\overline{x}, (1-\frac{1}{\delta}\rho))$. Then there exists a $z \in B(0_X, \rho)$ such that $x^0 = \overline{x} + \left(1-\frac{1}{\delta}\right) z$. It follows that

$$\begin{aligned} f(x^0) &= f\left(\overline{x} + \left(1-\frac{1}{\delta}\right) z\right) = f\left(\overline{x} + \left(1-\frac{1}{\delta}\right) x + \left(1-\frac{1}{\delta}\right)(x+z)\right) \\ &= f\left(\frac{1}{\delta}(x+\delta(\overline{x}-x)) + \left(1-\frac{1}{\delta}\right)(x+z)\right) \\ &\leq \frac{1}{\delta} f(x+\delta(\overline{x}-x)) + \left(1-\frac{1}{\delta}\right) f(x+z) \leq \frac{1}{\delta} f(y) + \left(1-\frac{1}{\delta}\right) c = \alpha. \end{aligned}$$

To obtain the last inequality we used the following facts $x + \delta(\overline{x}-x) = y$, $f(y) < \infty$ (because $y \in \mathrm{int}(\mathrm{dom}\, f)$) and $f(x+z) \leq c$.

It is clear that α does not depend on $x^0 \in B\left(\overline{x}, \left(1 - \frac{1}{\delta}\right)\rho\right)$. It follows that f is bounded on the neighborhood $B\left(\overline{x}, \left(1 - \frac{1}{\delta}\right)\rho\right)$ of $\overline{x} \in \operatorname{int}(\operatorname{dom} f)$. Again by Lemma 2.1, it follows that f is continuous at $\overline{x}$. But $\overline{x}$ is an arbitrary point of $\operatorname{int}(\operatorname{dom} f)$, hence f is continuous on the whole $\operatorname{int}(\operatorname{dom} f)$. $\square$

Proposition 2.2. *Let X be a normed space, f convex and continuous at $x_0 \in \operatorname{int}(\operatorname{dom} f)$. Then f is locally Lipschitzian at x_0, that is, there exist $L > 0$ and $\delta > 0$ such that*

$$B(x_0, \delta) = \{x \in X : \|x - x_0\| \leq \delta\} \subset \operatorname{int}(\operatorname{dom} f)$$

and $\mid f(x) - f(y) \mid \leq L\|x - y\|$, whenever $x, y \in B(x_0, \delta)$.

Proof. The continuity of f at x_0 implies that f is locally bounded there, i.e. there exist $L_1 > 0$ and $\delta > 0$ satisfying $\mid f \mid \leq L_1$ on $B(x_0, 2\delta) \subset \operatorname{int}(\operatorname{dom} f)$.

Now, let be x, y distinct points in $B(x_0, \delta)$, $\alpha := \|x - y\|$ and let

$$z = y + \frac{\delta}{\alpha}(y - x) = y + \delta\frac{y - x}{\|y - x\|} = \left(1 + \frac{\delta}{\alpha}\right) y - \frac{\delta}{\alpha}x.$$

We have that

$$\|z - x_0\| = \left\| y - x_0 + \delta\frac{y - x}{\|y - x\|} \right\| \leq \|y - x_0\| + \delta \leq 2\delta,$$

and hence $z \in B(x_0, 2\delta)$.

Because

$$y = \frac{\alpha}{\alpha + \delta}z + \frac{\delta}{\alpha + \delta}x$$

is a convex combination lying in $B(x_0, 2\delta)$, it follows that

$$f(y) \leq \frac{\alpha}{\alpha + \delta}f(z) + \frac{\delta}{\alpha + \delta}f(x).$$

Subtracting from both left and right hand side $f(x)$, we obtain

$$f(y) - f(x) \leq \frac{\alpha}{\alpha + \delta}\left[f(z) - f(x)\right] \leq \frac{\alpha}{\delta}2L_1 \leq \frac{2L_1}{\delta}\|x - y\|$$

(since $\mid f \mid \leq L_1$). Interchanging x and y, yields

$$f(x) - f(y) \leq \frac{2L_1}{\delta}\|x - y\|.$$

This is equivalent to

$$f(y) - f(x) \geq -\frac{2L_1}{\delta} \|x - y\| .$$

Hence we obtained $| f(y) - f(x) | \leq L\|y - x\|$, with $L = \frac{2L_1}{\delta}$. $\square$

Corollary 2.1. *Let f be a convex and continuous function at $x_0 \in \mathrm{int}(\mathrm{dom}\, f)$. Then $d^+ f(x_0)$ is a continuous sublinear function on X and therefore $df(x_0)$ (if it exists) is a continuous linear function ($df(x_0)$ is the Gateaux-differential or Gateaux-derivative of f at the point x_0).*

Proof. With Proposition 2.2 follows for $x_0 \in \mathrm{int}(\mathrm{dom}\, f)$ and $h \in X$

$$f(x_0 + th) - f(x_0) \leq L\|x_0 + th - x_0\| = Lt\|h\|,$$

provided that $t > 0$ is sufficiently small such that $x_0 + th \in B = B(x_0, \delta)$ (cf. Proposition 2.2). It follows that

$$d^+ f(x_0)(h) = \lim_{t \to 0^+} \frac{f(x_0 + th) - f(x_0)}{t} \leq L\|h\| \ \forall h \in X.$$

Hence $d^+ f(x_0)$ is continuous. $\square$

If the Gateaux-differential (Gateaux-derivative) $df(x_0)$ (which is by definition linear) is even continuous (which is true e.g. under the assumptions of Corollary 2.1), then the following representation formula is valid

$$\lim_{t \to 0} \frac{f(x_0 + th) - f(x_0)}{t} = \frac{d}{df} f(x_0 + th) \mid_{t=0} = df(x_0)(h) = \langle f'(x_0), h \rangle,$$

i.e. $df(x_0)$ is now also denoted by $f'(x_0) \in X^*$.

Thus $f' : X \to X^*$ defines a (in general nonlinear) mapping from X into X^*. f' is also said to be the **gradient of** f, and $f'(x_0)$ is said to be the **gradient of** $f(x_0)$ (f at x_0).

We give now another Corollary, namely to Proposition 2.1.

Corollary 2.2. *A finite convex function f on an open convex set $D \subset \mathbb{R}^n$ ($n \geq 1$) is continuous.*

Proof. We set $f(x) = +\infty$ for $x \notin D$. Then f is a convex function (as a function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$. Let $x \in D$ and consider for example $n = 2$.

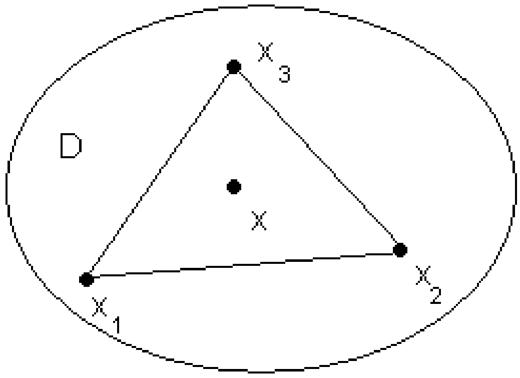

Figure 2.3

Choosing a triangle Δ (Figure 2.3) spanned by $x_1, x_2, x_3 \in D$ with $x \in \operatorname{int} \Delta \subseteq D$, we see that x may be represented as convex linear combination

$$x = \lambda_1 x_1 + \lambda_2 x_2 + \lambda_3 x_3,\ 0 \leq \lambda_i \leq 1,\ i = 1, \ldots, 3,\ \lambda_1 + \lambda_2 + \lambda_3 = 1.$$

Due to the convexity of f there holds

$$f(x) \leq \lambda_1 f(x_1) + \lambda_2 f(x_2) + \lambda_3 f(x_3).$$

This means that f is bounded from above on the triangle D which is a neighborhood of x and thus the assertion follows from Proposition 2.1.

For $n > 2$ ($n = 1$) the proof is analogous: x may be represented as a convex linear combination of $(n + 1)$ points $x_1, \ldots, x_{n+1}$ spanning a simplex S which contains x in is interior. $\square$

3 Lower Semicontinuity of (Convex) Functions

Let X be a Banach space (more generally it may be a topological locally convex space).

Definition 3.1. *The function $f : X \to \overline{\mathbb{R}}$ is said to be* **lower (upper) semicontinuous** *at $x^0 \in X$ if for $\varepsilon > 0$ there exists a neighborhood $U(x^0)$ of x^0 such that $\varepsilon < (>) f(x) - f(x^0)$ for all $x \in U(x^0)$.*

Remark: Of course, if f is both lower and upper semicontinuous at x^0 then f is continuous at x^0, i.e. $|f(x) - f(x^0)| < \varepsilon$ for all $x \in U(x^0)$.

Consideration of continuous functions turns out to be an unnecessary restriction (limitation) in particular within optimization theory and convex analysis. For example, regarding minimization of a continuous function f over a compact set D (the minimum exists due to the Weierstrass theorem), we can define

$$\tilde{f}(x) = \begin{cases} f(x), & \text{if} \quad x \in D, \\ +\infty, & \text{otherwise.} \end{cases}$$

Then $\min\limits_{x \in D} f(x) = \min\limits_{x \in X} \tilde{f}(x)$ (minimizing over the entire space X). Of course $\tilde{f}$ is not necessarily continuous on the whole X, but lower semicontinuous. Moreover, for continuous functions, there are a lot of properties and assertions in convex analysis and optimization which can be generalized to semicontinuous (lower or upper and weakly semicontinuous respectively) functions. As an example consider the **Weierstrass theorem**. It is true also for lower semicontinuous functions in the sense that a **lower (upper) semicontinuous function f on a compact set B attains its minimum (maximum)**. The proof can remain as an exercise for the reader.

Remarks: The following assertions result immediately from the definition.

(i) If f is continuous, then f is lower (upper) semicontinuous.

(ii) If f, g are lower semicontinuous and $\lambda > 0$, then λf and $f + g$ are lower semicontinuous.

(iii) If f is upper semicontinuous, then $-f$ is lower semicontinuous.

Exercise 3.1. The function f is lower semicontinuous at x^0 if and only if

$$\underline{\lim}_{x \to x^0} f(x) \geq f(x^0).$$

Definition 3.2. *The function $f : X \to \overline{\mathbb{R}}$ is said to be* **sequentially lower semicontinuous** *at $x^0 \in X$ if and only if $f(x^0) \leq \underline{\lim}_{n \to +\infty} f(x_n)$ holds for each sequence (x_n), $x_n \in X$, satisfying $\underline{\lim}_{n \to +\infty} x_n = x^0$.*

Again, trivially by definition, for X Banach space (as supposed above) a sequentially lower semicontinuous function is lower semicontinuous at x^0 and viceversa (as for continuous functions at x^0, where the inferior limit has to be replaced by the limit)(e.g. for $X = \mathbb{R}^n$). The lower semicontinuity can be characterized in different ways as the following theorem says.

Theorem 3.1. *Let be X a Banach space and $f : X \to \overline{\mathbb{R}}$. The following conditions are equivalent.*

(i) The function f is lower semicontinuous on X (for all $x \in X$).

(ii) The set $\{x \in X : f(x) > k\}$ is an open set for each $k \in \mathbb{R}$.

(iii) The set $E_k := \{x \in X : f(x) \leq k\}$ is an closed set (level set) for each $k \in \mathbb{R}$.

(iv) The set $\operatorname{epi} f$ is closed (as subset of $X \times \mathbb{R}$).

Proof. "$(i) \Rightarrow (ii)$" Let be x^0 such that $f(x^0) > k$, $k \in \mathbb{R}$. Because f is lower semicontinuous at x^0, to a given $\varepsilon = \frac{1}{2}(f(x^0) - k)$ there exists a neighborhood $W(x^0)$ of x^0 such that (by definition)

$$f(x) > f(x^0) - \varepsilon = f(x^0) - \frac{1}{2}(f(x^0) - k) = \frac{1}{2}(f(x^0) + k) > k \ \forall x \in W(x^0).$$

This implies that the set $\{x \in X : f(x) > k\}$ is open.
"$(ii) \Rightarrow (i)$" Consider any x^0 and any $\varepsilon > 0$. Because of (ii) the set

$$M = \left\{x \in X : f(x) > k = f(x^0) - \varepsilon\right\}$$

is open and $f(x^0) \in M$ since $f(x^0) > k = f(x^0) - \varepsilon$. From here follows that there exists a neighborhood $U(x^0)$ of x^0 contained in M: $U(x^0) \subset M$. That means that for all $x \in U(x^0)$ holds $f(x) > f(x^0) - \varepsilon$, which is indeed the definition of lower semicontinuity for f at x^0.

"$(ii) \iff (iii)$" Trivial, because E_k is the complement set of the set from (ii).

"$(i) \iff (iv)$" We define a new function $F : X \times \mathbb{R} \to \overline{\mathbb{R}} :\ F(x, \lambda) := f(x) - \lambda$. In order to continue the proof the following result is required.

Lemma 3.1. *F is lower semicontinuous if and only if f is lower semicontinuous (on $X \times \mathbb{R}$ and X respectively).*

Proof. *Sufficiency.* Let f be lower semicontinuous on X. Assume F is not lower semicontinuous. Because of $(ii) \iff (iii)$ from the previous theorem (for F) the set

$$M := \big\{(x, \lambda) \in X \times \mathbb{R} : F(x, \lambda) = f(x) - \lambda \le L\big\}$$

is not closed for at least one $L \in \mathbb{R}$ and that means that there exists a limit point $(x^0, \lambda^0) \in \partial M \backslash M$ fulfilling $f(x^0) - \lambda^0 > L$.

Thus for any neighborhood $U(x^0, \lambda^0)$ of the point (x^0, λ^0) there exists a point $(\overline{x}, \overline{\lambda}) \in U(x^0, \lambda^0) \cap M$, i.e. $f(\overline{x}) - \overline{\lambda} \le L$. Especially let $|\lambda - \lambda^0| < \delta$ which implies $\lambda < \lambda^0 + \delta$ for a sufficiently small $\delta > 0$.

Because f is lower semicontinuous at x^0 for $\varepsilon := \frac{1}{2}(f(x^0) - (L + \lambda^0)) - \delta > 0$ there follows the existence of a neighborhood $V(x^0)$ of x^0 such that

$$\begin{aligned} f(x) &> f(x^0) - \varepsilon \\ &= f(x^0) - \left[\frac{1}{2}\Big(f(x^0) - \big(L + \lambda^0\big)\Big) - \delta\right] \\ &= \frac{1}{2}\Big(f(x^0) + L + \lambda^0\Big) + \delta \\ &> L + \lambda^0 + \delta \ \ \forall x \in V(x^0). \end{aligned}$$

Thus we can choose as $U(x^0, \lambda^0)$ the set $U(x^0, \lambda^0) = V(x^0) \times (\lambda^0 - \delta, \lambda^0 + \delta)$. Then for all $(x, \lambda) \in U(x^0, \lambda^0)$ there is $f(x) > L + \lambda^0 + \delta > L + \lambda$. But for $(\overline{x}, \overline{\lambda}) \in U(x^0, \lambda^0)$ holds $f(\overline{x}) \le L + \overline{\lambda}$ which is a contradiction.

Necessity. Let f be lower semicontinuous on X for all $x \in X$. Assume that

F is not lower semicontinuous. Because of $(ii) \Leftrightarrow (iii)$ (for F) the set

$$M = \{(x, \lambda) \in X \times \mathbb{R} : F(x, \lambda) = f(x) - \lambda \leq L\}$$

is not closed for at least an $L \in \mathbb{R}$ and that means that there exists a limit point (x^0, λ^0), i.e. boundary point of M, but not belonging to M, that yields the existence of a sequence $(x_n, \lambda_n) \in M$ with $(x_n, \lambda_n) \to (x^0, \lambda^0)$ $\forall n \in \mathbb{N}$, i.e. $f(x_n) - \lambda_n \leq L \quad \forall n \in \mathbb{N}$ and $f(x^0) - \lambda^0 > L$. This is a contradiction since $\underline{\lim}_{x_n \to x^0} f(x_n) \geq f(x^0)$ (because f is lower semicontinuous), but from the inequalities above we get

$$\underline{\lim}_{n\to\infty} (f(x_n) - \lambda_n) = \underline{\lim}_{n\to\infty} f(x_n) - \lambda^0 \leq L,$$

followed by $f(x^0) - \lambda^0 \leq L$, which contradicts $f(x^0) - \lambda^0 > L$. $\square$

Continuation of the Proof of Theorem 3.1. "$(i) \Longleftrightarrow (iv)$" Let f be lower semicontinuous, which means nothing but that F is lower semicontinuous (from Lemma 3.1). Because of (iii) F lower semicontinuous implies that the level set

$$\{(x, \lambda) : F(x, \lambda) \leq \mu\}$$

is closed for all $\mu \in \mathbb{R}$. But $F(x, \lambda) = f(x) - \lambda \leq \mu$ implies $f(x) \leq \mu + \lambda$ and we can write

$$\{(x, \lambda) : F(x, \lambda) \leq \mu\} = \{(x, \lambda) : (x, \lambda + \mu) \in \operatorname{epi} f\} = \operatorname{epi} f - (0, \mu),$$

which is a translation of $\operatorname{epi} f$ and so this is equivalent with $\operatorname{epi} f$ is closed. $\square$

Example 3.1. The indicator function

$$\chi_A(x) = \begin{cases} 0, & \text{if} \quad x \in A, \\ +\infty, & \text{otherwise,} \end{cases}$$

is lower semicontinuous if and only if A is closed (follows from (iii) in the theorem).

Remark: Let $f_i(x)$ be lower semicontinuous functions on X, $i \in I$. Then

$$f(x) = \sup_{i\in I} f_i(x)$$

(the pointwise supremum of a family of lower semicontinuous functions) is a lower semicontinuous function.

Proof. Since f_i, $i \in I$ are lower semicontinuous functions then $\operatorname{epi} f_i$ is a closed set for all $i \in I$. But because the intersection of infinitely many closed sets is closed we have that $\operatorname{epi} f = \bigcap\limits_{i \in I} \operatorname{epi} f_i$ is a closed set. $\square$

Definition 3.3. *The largest lower semicontinuous minorant of the function* $f : X \to \overline{\mathbb{R}}$ *is called the* **lower semicontinuous regularization** *of* f *and is denoted by* $\overline{f}$.

Remark: It exists as the pointwise supremum of those lower semicontinuous functions everywhere less than f (cf. Remark above). It can be characterized by the following statement.

Proposition 3.1. *Let be* $f : X \to \overline{\mathbb{R}}$ *and* $\overline{f}$ *its lower semicontinuous regularization. Then* $\operatorname{epi} \overline{f} = \overline{\operatorname{epi} f}$ *and* $\overline{f}(x) = \varliminf\limits_{y \to x} f(y) \quad \forall x \in X$.

Proof. Because $\overline{f}(x) \leq f(x)$, as $\overline{f}$ is lower semicontinuous, we have $\operatorname{epi} \overline{f} \supseteq \operatorname{epi} f$.

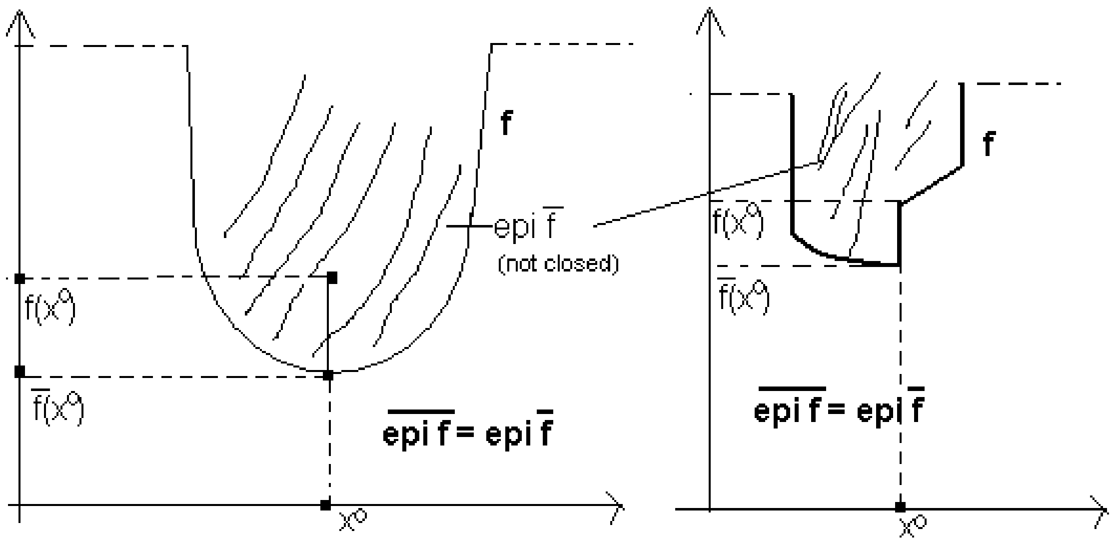

Figure 3.1

But epi $\overline{f}$ is closed and this implies that epi $f \supseteq \overline{\text{epi}\, f}$.

Conversely, $\overline{\text{epi}\, f}$ can be considered as epigraph of a function g, epi $g = \overline{\text{epi}\, f}$ (see Figure 3.1).

Since epi g is closed, then g is a lower semicontinuous function with $g(x) \leq f(x)$ and this implies $g(x) \leq \overline{f}(x)$ because the definition of $\overline{f}$ is as the largest lower semicontinuous minorant of f and from here follows epi $g = \overline{\text{epi}\, f} \supseteq$ epi $\overline{f}$. Getting these two inclusions together, we obtain that $\overline{\text{epi}\, f} =$ epi $\overline{f}$. Consequently $\overline{f} = \underline{\lim}_{y \to x} f(y) \;\; \forall\, x \in X$. □

We need an even more general definition, so-called **weak semicontinuity**.

Consider X a real normed space (not necessary a Banach space). We introduce the notion of **weak convergence.**

Definition 3.4. *A sequence $(x_n)_{n\in\mathbb{N}}$, $x_n \in X$ is said to be* **weakly convergent** *to $x \in X$ if for all continuous linear functionals $x^* \in X^*$ we have $\lim\limits_{n\to\infty} \langle x^*, x_n\rangle = \langle x^*, x\rangle$. Moreover, x is called* **the weak limit** *of the sequence (x_n) and the following notations are usually used $x_n \rightharpoonup x$ or $w - \lim\limits_{n\to\infty} x_n = x$.*

Example 3.2. Consider the Hilbert space l_2 of all real sequences $x = (x^i)_{i\in\mathbb{N}}$ with $\sum\limits_{i=1}^{\infty} |x^i|^2 < \infty$, scalar product is $\langle x, y\rangle = \sum\limits_{i=1}^{\infty} x^i y^i$, $x, y \in l_2$.

We study the special sequence in l_2 (i.e. a sequence of l_2-sequences) $x_1 = (1, 0, 0, 0, ...)$, $x_2 = (0, 1, 0, 0, ...)$, $x_3 = (0, 0, 1, 0, ...)$ etc. Then $x_n \rightharpoonup 0_{l_2} = (0, 0, 0, ...)$ because a continuous linear functional $x^* \in l_2^*$ can be identified with an element $y = (y^i) \in l_2$, $i \in \mathbb{N}$ ($H^* \cong H$ for Hilbert spaces) and $\langle x^*, x\rangle$ can be considered by the scalar product $\langle x^*, x\rangle = \langle y, x\rangle = \sum\limits_{i=1}^{\infty} y^i x^i$. Since $\sum\limits_{i=1}^{\infty} |y^i|^2 < \infty$, we have $\lim\limits_{n\to\infty} \langle x^*, x_n\rangle = \lim\limits_{n\to\infty} \langle y, x_n\rangle = \lim\limits_{n\to\infty} y^n = 0$. On the other hand, $\langle x^*, 0_{l^2}\rangle = \langle y, 0_{l^2}\rangle = \sum\limits_{i=1}^{\infty} y^i \cdot 0 = 0$ and hence $\lim\limits_{n\to\infty} \langle x^*, x_n\rangle = \langle x^*, 0_{l^2}\rangle = 0$ and so $x_n \rightharpoonup 0_{l^2}$.

But $x_n \to 0_{l^2}$ does not stand because $\|x_n\|_{l^2} = \sqrt{\langle x_n, x_n\rangle} = \sqrt{\sum\limits_{i=1}^{\infty} (x_i)^2} = 1$ $\forall\, n \in \mathbb{N}$.

Definition 3.5. *Let X be a real normed space $(X, \|\cdot\|)$. The function f :*

$X \to \overline{\mathbb{R}}$ *is called* **weak (sequentially) lower semicontinuous** *at* x^0 *if for every sequence* $(x_n)_{n\in\mathbb{N}}$ *converging weakly to* x^0*, i.e.* $x_n \rightharpoonup x^0$*, holds* $f(x^0) \leq \varliminf_{n\to\infty} f(x_n)$.

Comparing this to the definition of sequentially lower semicontinuous functions one may notice that the only difference consists in that there was $x_n \to x^0$ instead of $x_n \rightharpoonup x^0$.

Remark: Instead of "weak sequentially lower semicontinuous" we simply say "weak-lower semicontinuous".

Definition 3.6. *Let* $(X, \|\cdot\|)$ *be a normed space. A non-empty subset* D *of* X *is called (cf. Figure 3.2)*

(i) **weakly (sequentially) closed** *if for every weakly convergent sequence* $x_n \rightharpoonup x$, $x_n \in D$, *follows* $x \in D$,

(ii) **weak sequentially compact** *if every sequence in* D *contains a weakly convergent subsequence whose weak limit belongs to* D.

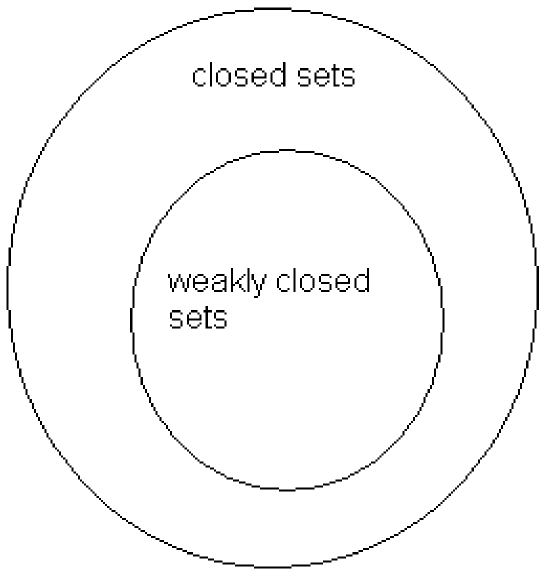

Figure 3.2

Remark: D weak (sequentially) closed implies that D is closed. To prove this, let be $x_n \to x$, $x_n \in D$ which assures that $x_n \rightharpoonup x$, so $x_n \in D$ and this means that D is closed. The converse is, in general, not true. This leads to a modification (generalization) of the Weierstrass Theorem as follows.

Theorem 3.2. *Let be X a normed space, D a non-empty weak sequentially compact set and f weak-lower semicontinuous on D. Then there exists at least one $\overline{x} \in D$ with $f(\overline{x}) \leq f(x)\ \forall\ x \in D$, i.e. the optimization problem $\min\limits_{x \in D} f(x)$ has at least one solution.*

Proof. First, $(x_n)_{n \in \mathbb{N}}$ is assumed to be an infimal sequence in D,

$$\lim_{n \to \infty} f(x_n) = \inf_{x \in S} f(x).$$

Because D is weak sequentially compact, there exists a subsequence $(x_{n_i}),\ i \in \mathbb{N}$ with $x_{n_i} \rightharpoonup \overline{x}$, where $\overline{x}$ is some $\overline{x} \in D$.

Since f is weak-lower semicontinuous we have

$$f(\overline{x}) \leq \varliminf_{i \to \infty} f(x_{n_i}) = \inf_{x \in D} f(x)$$

and our proof is complete. □

Remark: Obviously, f weak-lower semicontinuous (weak (sequentially) lower semicontinuous) in a Banach space implies f is lower semicontinuous, but not viceversa.

Proof. In a Banach space a sequentially lower semicontinuous function is lower semicontinuous and viceversa. Let f be weak-lower semicontinuous and $x_n \to x^0$ any sequence. Then $x_n \rightharpoonup x^0$ and $f(x^0) \leq \varliminf\limits_{n \to \infty} f(x_n)$, which implies that f is lower semicontinuous. □

Example 3.3. Let be $(X, \|\cdot\|)$ a Banach space. Then $f(x) = \|x\|$ is weak-lower semicontinuous.

To prove this, we need a conclusion of the famous Hahn-Banach Theorem.

Theorem 3.3. *Let X be a real linear space and $f : X \to \overline{\mathbb{R}}$ a sublinear functional. Then there exists a linear functional l on X such that $l(x) \leq f(x)\ \forall\ x \in X$.*

This is a so-called basic version of the Hahn-Banach Theorem. There exist some further versions of it. We give another formulation of the Hahn-Banach Theorem a so-called **Continuation Theorem**.

Theorem 3.4. (Hahn-Banach Continuation) *Let X be a real (or even complex) linear space, p a seminorm on X and $L \subset X$ a linear subspace of X. Further, let f be a linear functional defined on L, fulfilling the estimation*

$$|f(x)| \leq p(x) \quad \forall\ x \in X.$$

Then there exists a linear functional $\tilde{f}$ defined on whole X (i.e. $\tilde{f} \in X'$, where X' is the algebraic dual space to X) such that

$$\tilde{f} = f(x)\ \forall x \in L \text{ and } |\tilde{f}(x)| \leq p(x) \qquad \forall x \in X.$$

Therefore $\tilde{f}$ turns out to be a continuation of f from L to X which satisfies the estimation by the semi-norm p on the whole X, too.

A proof is available in [2].

From this general version (the space X is only supposed to be a linear space etc.) one can deduce the so-called Hahn-Banach- Continuation Theorem for normed spaces.

Theorem 3.5. (Hahn-Banach-Continuation for normed spaces) *Let $(X, \|\cdot\|)$ be a normed space and f a linear continuous functional on a linear subspace $L \subset X$. Then there exists a linear continuous functional $\tilde{f}$ on X (i.e. $\tilde{f} \in X^*$) which is a continuation of f from L to X keeping the norm*

$$\tilde{f}(x) = f(x)\ \forall x \in L \text{ and } \|\tilde{f}\|_{X^*} = \|f\|_L.$$

Proof. The function $p(x) := \|f\|_L\|x\|$ defines a semi-norm on X. Indeed we have $p(x) \geq 0 \ \ \forall\ x \in X$, $p(\lambda x) = \|f\|_L\|\lambda x\| = |\lambda|p(x)$, $\lambda \in \mathbb{R}$ and $p(x+y) = \|f\|_L\|x+y\| \leq \|f\|_L(\|x\|+\|y\|) = p(x)+p(y)$. From the definition of $\|\cdot\|_L$ we obtain $|f(x)| \leq \|f\|_L\|x\| = p(x) \ \ \forall\ x \in L$. With the Hahn-Banach Continuation Theorem (Theorem 3.4) follows the existence of a continuation $\tilde{f}$ of f, $\tilde{f}$ linear functional on X, satisfying the estimation

$$\tilde{f}(x) = f(x)\ \forall x \in L \text{ and } \|\tilde{f}\| \leq p(x) = \|f\|_L\|x\|\ \forall x \in X.$$

This means that $\tilde{f}$ is continuous, i.e. $\tilde{f} \in X^*$ and $\|\tilde{f}\|_{X^*} \leq \|f\|_L$ (by definition). But we also have $|f(x)| = |\tilde{f}(x)| \leq \|\tilde{f}\|_{X^*}\|x\|$ for all $x \in X$ (in particular even for all $x \in L$) which means $\|\tilde{f}\|_{X^*} \leq \|f\|_L$. Consequently, $\|\tilde{f}\|_{X^*} = \|f\|_L$. □

There is also an interesting conclusion of this statement (as announced above).

Conclusion 3.1. *Let $(X, \|\cdot\|)$ be a linear normed space, $0 \neq x_0 \in X$ any element. Then there exists a linear continuous functional $\tilde{f} = x^* \in X$ on X such that $\|x^*\|_* = 1$ and $\langle x^*, x_0\rangle = \|x_0\|$.*

Proof. Let $L = \{\alpha x_0\}$ be a one dimensional linear subspace of X, $\alpha \in \mathbb{R}$ (i.e. L is spanned by x_0). Define on L the linear continuous functional $f(x) = f(\alpha x_0) := \alpha\|x_0\|$, $x = \alpha x_0 \in L$. Consequently $f(x) = \|x\| \ \ \forall x \in X$, especially $f(x_0) = \|x_0\|$. The functional f is linear because $f(\lambda x) = f(\lambda\alpha x_0) = \lambda\alpha\|x_0\| = \lambda f(x)$, $\lambda \in \mathbb{R}$ and $f(x+y) = f(\alpha_1 x_0 + \alpha_2 x_0) = f((\alpha_1+\alpha_2)x_0) = (\alpha_1+\alpha_2)\|x_0\| = \alpha_1\|x_0\| + \alpha_2\|x_0\| = f(x)+f(y)$.

Because $\|f(x)\| = \|x\|\ \forall x \in L$, we have $\|f\|_L = 1$. Using Theorem 3.5 (Hahn-Banach for normed spaces) follows the existence of a linear continuous functional denoted $\tilde{f} = x^*$ fulfilling for each $x \in L$

$$\langle x^*, x\rangle = f(x) = \|x\|,$$

in particular $\langle x^*, x_0\rangle = \|x_0\|$, since $x_0 \in L$ and

$$\|x^*\|_* = \|\tilde{f}\|_* = \|f\|_L = 1.$$

□

Now we return to Example 3.3. We need to prove that the norm attached to a Banach space is a weak-lower semicontinuous function.

Proof. Considering a sequence $x_n \rightharpoonup x_0$, we show $\|x_0\| \le \underline{\lim}_{n\to\infty} \|x_n\|$.

Let us suppose the contrary, i.e.

$$\|x_0\| > \underline{\lim}_{n\to\infty} \|x_n\|.$$

Consequently there exists $c \in \mathbb{R}$ such that

$$\|x_0\| > c > \underline{\lim}_{n\to\infty} \|x_n\|.$$

By the definition of the limit there exists a subsequence $(x_{n_k})_{k\in\mathbb{N}}$ such that $\|x_0\| > c > \|x_{n_k}\|$. From Conclusion 3.1 follows the existence of a functional $x^* \in X^*$, $\|x^*\|_* = 1$ and $\langle x^*, x_0\rangle = \|x_0\| > c$. On the other hand,

$$\|x^*, x_{n_k}\| \le \|x^*\|_* \|x_{n_k}\| = \|x_{n_k}\| < c,$$

so

$$\langle x^*, x_0\rangle \lim_{n_k \rightharpoonup \infty} \langle x^*, x_{n_k}\rangle \le c \text{ (weak convergence)}.$$

This contradicts the result obtained above, so the norm is weak-lower semicontinuous indeed. □

Remark: Previously, we have remarked that a weak (sequentially) closed set is also closed. The converse assertion is not true in general. A closed set is also weak (sequentially) closed if and only if the set is also convex. The proof is not trivial, but we omit it here, being available in [1].

The following conclusion follows from the last remark.

Conclusion 3.2. *The indicator function $\chi_D(x)$ of a convex closed set $D \subset X$ (X normed space or, more general, locally convex topological vector space - in this case weak-lower semicontinuity is defined by means of the weak topology) is weak-lower (sequentially) semi-continuous .*

Proof. Let $x_n \rightharpoonup x_0$.

(i) If $x_0 \in D$, we have $\chi_D(x_0) = 0$, so

$$0 = \chi_D(x_0) \leq \varliminf_{n\to\infty} \chi_D(x_n) = \begin{cases} +\infty, & \text{if } x_n \notin D, \\ 0, & \text{if } x_n \in D. \end{cases}$$

(ii) If $x_0 \notin D$, it follows that there is no n_0 such that for any $n > n_0,\ x_n \notin D$. Otherwise there would exist a subsequence $x_{n_k} \rightharpoonup x_0$ such that $x_{n_k} \in D$, implying $x_0 \in D$ (D is weak sequential closed, since is closed and convex) and this is a contradiction. Hence

$$+\infty = \chi_D(x_0) = \lim_{n\to\infty} \chi_D(x_n) = +\infty,$$

so $\chi_D(x_0) \leq \varliminf_{n\to\infty} \chi_D(x_n)$. □

Now we prove that a convex function $f : X \to \overline{\mathbb{R}}$, which has a continuous Gateaux-differential (gradient) is weak-lower semicontinuous (X is a normed space). To show this we need the following statement first.

Proposition 3.2. *Let X be a real linear normed space, $f : X \to \mathbb{R}$ a Gateaux-differentiable convex function, $f' : X \to X^*$ a linear continuous function. Then $f(y) - f(x) \geq \langle f'(x), y - x\rangle\ \forall x, y \in X$.*

Proof. Let $0 < \lambda \leq 1$. Because f is convex, we have

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

It follows that

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

We build the limit $\lambda \to 0$. By the definition of Gateaux-derivative, we obtain $\langle f'(x), y - x\rangle \leq f(y) - f(x)$. □

Proposition 3.3. *Let X be a real linear normed space, $f : X \to \mathbb{R}$ a Gateaux-differentiable convex function, $f' : X \to X^*$ a linear continuous function. Then f is weak-lower semicontinuous.*

Proof. Let be $x_0 \in X$ an arbitrary element, (x_n) a sequence such that $x_n \rightharpoonup x_0$. From Proposition 3.2 follows $f(x_n) \geq f(x_0) + \langle f'(x_0), x_n - x_0 \rangle$ and

$$\varliminf_{n \to +\infty} f(x_n) \geq f(x_0) + \varliminf_{n \to +\infty} \langle f'(x_0), x_n - x_0 \rangle = f(x_0),$$

as $x_n \rightharpoonup x_0$ implies $\langle f'(x_0), x_n - x_0 \rangle \to f(x_0)$. $\square$

Exercise 3.2. Let X be a reflexive Banach space, $B : X \to X^*$ a linear, bounded, non-negative operator (i.e. $\langle Bx, x \rangle \geq 0 \ \forall x \in X$), $B^* : X \to X^*$ its adjoint operator and the function $f : X \to \mathbb{R}$, $f(x) = \langle Bx, x \rangle$. Then $f'(x) = Bx + B^*x$.

Exercise 3.3. Let be X a Hilbert space, $f(x) = \|x\| = \langle x, x \rangle^{\frac{1}{2}}$. Then

$$f'(x) = \frac{x}{\|x\|}, \quad x \neq 0.$$

Example 3.4. Let be X a reflexive Banach space, $B : X \to X^*$ a linear, bounded, non-negative and self-adjoint operator ($B = B^*$). Then $f(x) = \langle Bx, x \rangle$ is weak-lower semicontinuous.

To prove this we apply Proposition 3.3., as f is convex (cf. above). From Exercise 3.2 above, follows that $f'(x) = Bx + B^*x = 2Bx$ is gradient. Now, from Proposition 3.3 it follows that f is weak-lower semicontinuous.

Now we are going to prove some properties (monotony) of the gradient f'.

Definition 3.7. *Let X be a linear normed space. $A : X \to X^*$ is said to be a*

(i) **monotone operator** *(mapping) if $\langle Ax - Ay, x - y \rangle \geq 0 \ \forall x, y \in X$,*

(ii) **strictly monotone** *if $\langle Ax - Ay, x - y \rangle > 0 \ \forall x, y \in X, \ x \neq y$,*

(iii) **strongly monotone** *if $\langle Ax - Ay, x - y \rangle \geq \gamma \|x - y\|_X^2, \quad \gamma > 0$.*

Theorem 3.6. *Let be X a linear normed space, $f : X \to \mathbb{R}$ a convex and Gateaux-differentiable function, $f' : X \to X^*$. Then the gradient f' is a monotone operator.*

Remark: This is a generalization of the fact that for $f : \mathbb{R} \to \mathbb{R}$ convex and differentiable, $f' : \mathbb{R} \to \mathbb{R}$ is monotonously increasing.

Proof. Because of Proposition 3.2, we have

$$\langle f'(x), y - x \rangle \leq f(y) - f(x) \text{ and } \langle f'(y), x - y \rangle \leq f(x) - f(y).$$

It follows $\langle f'(x), y - x \rangle \leq -(f(x) - f(y)) \leq \langle f'(y), y - x \rangle$. Therefore $\langle f'(y) - f'(x), y - x \rangle \geq 0 \ \forall x, y \in X$. Hence, f' is monotone. $\square$

The reverse assertion stands, too.

Theorem 3.7. *Let be X a linear normed space, $f : X \to \mathbb{R}$ Gateaux-differentiable and $f' : X \to X^*$ monotone. Then f is convex.*

In order to prove the assertion, we need the following intermediate result.

Lemma 3.2. (Lagrange formula, mean value theorem) *Let X be a linear normed space and $f : X \to \mathbb{R}$ have a gradient (Gateaux derivative) at each point $x \in X$: $f' : X \to X^*$. Then for $x, y \in X$ there exists a $\delta \in (0,1)$ such that*

$$f(x + y) - f(x) = \langle f'(x + \delta y), y \rangle.$$

Proof. Define $\varphi : \mathbb{R} \to \mathbb{R}$, $\varphi(t) := f(x + ty)$. Then

$$\varphi'(t) = \frac{d}{dt} f(x + ty) = \lim_{\tau \to 0} \frac{f(x + ty + \tau y) - f(x + ty)}{\tau} = \langle f'(x + ty), y \rangle.$$

Applying the mean value theorem for function φ, we get a $\delta \in (0,1)$ such that

$$f(x + y) - f(x) = \varphi(1) - \varphi(0) = \varphi'(\delta) = \langle f'(x + \delta y), y \rangle. \square$$

Remark: When $X = \mathbb{R}$, the mean value theorem says

$$f(x + y) - f(x) = f'(x + \delta y) y,$$

where $x + \delta y$ is a point between x and y.

Proof of Theorem 3.7. Let be given any $x, y \in X$. We verify Jensen's inequality. First set

$$d := \lambda f(x) + (1-\lambda)f(y) - f(\lambda x + (1-\lambda)y).$$

We want to show that $d \geq 0$. We have

$$d = \lambda[f(x) - f(\lambda x + (1-\lambda)y)] + (1-\lambda)[f(y) - f(\lambda x + (1-\lambda)y)]. \qquad (3.\ 1)$$

There is $x = [\lambda x + (1-\lambda)y] + (1-\lambda)(x-y)$, as well as $y = [\lambda x + (1-\lambda)y] + \lambda(y-x)$. Using Lemma 3.2, we obtain

$$\begin{aligned} d &= \lambda[f(\lambda x + (1-\lambda)y + (1-\lambda)(x-y)) - f(\lambda x + (1-\lambda)y)] \\ &+ (1-\lambda)[f(\lambda x + (1-\lambda)y + \lambda(y-x))] - f(\lambda x + (1-\lambda)y) \\ &= \lambda\langle f'(\lambda x + (1-\lambda)y + \delta_1(1-\lambda)(x-y)), (1-\lambda)(x-y)\rangle \\ &+ (1-\lambda)\langle f'(\lambda x + (1-\lambda)y + \delta_2\lambda(y-x)), \lambda(y-x)\rangle, \end{aligned}$$

where $0 < \delta_1, \delta_2 < 1$. Denote $V := \lambda x + (1-\lambda)y + \delta_1(1-\lambda)(x-y)$ and $W := \lambda x + (1-\lambda)y + \delta_2\lambda(y-x)$. It follows

$$V - W = \delta_1(1-\lambda)(x-y) - \delta_2\lambda(y-x) = (\delta_1(1-\lambda) + \delta_2\lambda)(x-y).$$

To ease the calculations we introduce another variable, $\mu := \delta_1(1-\lambda) + \delta_2\lambda$ and d can be written as

$$d = \lambda(1-\lambda)[\langle f'(V), x-y\rangle - \langle f'(W), x-y\rangle = \frac{\lambda(1-\lambda)}{\mu}\langle f'(V) - f'(W), V-W\rangle \geq 0,$$

since f is monotone. $\square$

We come back to weak-lower semi-continuity.

Theorem 3.8. *Let be X a linear normed space, $f : X \to \overline{\mathbb{R}}$. Then the following assertions are equivalent to each other*

(i) f is weak-lower semicontinuous.

(ii) $E_k = \{x \in X : \ f(x) \leq k\}$ is weakly (sequentially) closed for each $k \in \overline{\mathbb{R}}$.

(iii) epi f *is weakly (sequentially) closed.*

Proof. We omit the proof here because is a slightly modified version of the proof of Theorem 3.1. It can remain as an exercise. $\square$

Remark: Comparing Theorem 3.1 to the last one, it is not difficult to notice that that was the analogous assertion for lower semicontinuous functions.

4 Subdifferential

For convex Gateaux-differentiable functions, due to Proposition 3.2, one has

$$f(y) - f(x) \geq \langle f'(x), y - x\rangle \ \forall x, y \in X.$$

But this can be generalized by the so-called subdifferential (subgradient).

Definition 4.1. *Let X be a linear normed space, f a proper function on X (i.e. $\operatorname{dom} f \neq \emptyset,\ f > -\infty$) . Then $x^* \in X^*$ is said to be* **subgradient** *of f at $x \in \operatorname{dom} f$ if*

$$f(y) - f(x) \geq \langle x^*, y - x\rangle \ \forall y \in X.$$

The set of all subgradients of f at x is called **subdifferential** *and is denoted by $\partial f(x)$,*

$$\partial f(x) := \{x^* \in X^* : \ f(y) - f(x) \geq \langle x^*, y - x\rangle \ \forall y \in X\}.$$

The function f is called subdifferentiable at x if $\partial f(x) \neq \emptyset$.

Remark: Here f does not need to be convex.

Geometrical interpretation. (cf. Figure 4.1)

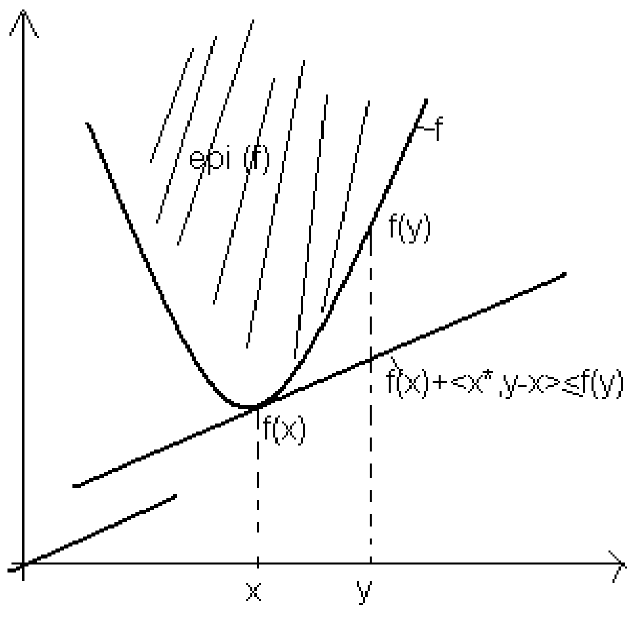

Figure 4.1

$f(x)+\langle x^*, y-x\rangle$ is supporting f (and epi f) at $(x, f(x))$. Thus the subdifferential generalizes the classical concept of derivative.

Example 4.1. Let X be a Banach space equipped with the norm $\|\cdot\|$ and $f : X \to \mathbb{R}_+$, $f(x) = \|x\|$. Then

$$\partial f(x) = \begin{cases} \{x^* \in X^* : \|x^*\|_{X^*} \leq 1\}, & \text{if } x = 0, \\ \{x^* \in X^* : \|x^*\|_{X^*} = 1, \langle x^*, x\rangle = \|x\|\}, & \text{if } x \neq 0. \end{cases}$$

To prove this we slit into two cases. Let us consider first that $x = 0$. Then $x^* \in \partial f(0)$ if and only if

$$\|y\| - \|0\| \geq \langle x^*, y - 0\rangle \ \forall y \in X,$$

that is equivalent to $\|y\| \geq \langle x^*, y\rangle$. Further

$$\|y^*\|_{X^*} := \sup_{y \neq 0} \frac{\langle x^*, y\rangle}{\|y\|} \leq 1.$$

Take now $x \neq 0$. Let be $x^* \in X^*$ such that

$$\langle x^*, x\rangle = \|x\|_X \text{ and } 1 = \|x^*\|_{x^*} := \sup_{y \neq 0} \frac{\langle x^*, y\rangle}{y\|_X}.$$

It follows $\frac{\langle x^*, y\rangle}{\|y\|_X} \leq 1 \ \forall y \in X$, i.e. $\langle x^*, y\rangle \leq \|y\|_X \forall y \in X$. Consequently

$$\|y\|_X - \|x\|_X \geq \langle x^*, y\rangle - \langle x^*, x\rangle = \langle x^*, y - x\rangle \ \forall y \in X,$$

i.e. $x^* \in \partial f(x)$.

Conversely, for an x^* in $\partial f(x)$ we have

$$-\|x\|_X = \|0\|_X - \|x\|_X \geq \langle x^*, 0 - x\rangle = -\langle x^*, x\rangle,$$

so $\langle x^*, x\rangle \geq \|x\|$. But on the other hand we have

$$\|x\|_X = \|2x\|_X - \|x\|_X \geq \langle x^*, 2x - x\rangle = \langle x^*, x\rangle,$$

followed by $\|x\|_X = \langle x^*, x\rangle$. For some $y \in X$ and $\lambda > 0$ it holds

$$lay + x\|_X - \|x\|_X \geq \langle x^*, \lambda y,$$

i.e.

$$\left\|y + \frac{x}{\lambda}\right\|_X - \frac{1}{\lambda}\|x\|_X \geq \langle x^*, y\rangle \;\forall y \in X.$$

Setting λ to tend towards $+\infty$ we get for all $y \in X$ $\|y\| \geq \langle x^*, y\rangle$, i.e. $\|x^*\|_{X^*} \leq 1$. According to a result obtained above the last expression turns into equality.

Especially for $X = \mathbb{R}$ it holds $\|x\| = |x|$ and, as the scalar product in $\mathbb{R}$ is the usual product and the dual of $\mathbb{R}$ is also $\mathbb{R}$, we can distinguish three cases (see also Figure 4.2)

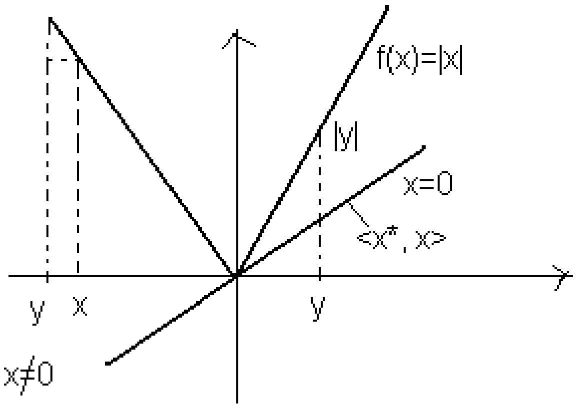

Figure 4.2

1. When $x = 0$, we have $f(y) - f(0) \geq \langle x^*, y - 0\rangle \;\Rightarrow\; |y| - 0 \geq x^*y \;\Rightarrow\; x^*y \leq |y| \;\Rightarrow\; -1 \leq x^* \leq 1 \;\Rightarrow\; |x^*| \leq 1 \;\Rightarrow\; \|x^*\| \leq 1$.

2. When $x < 0$, we have $f(y) - f(x) = |y| - |x| \geq (-1)(y - x) = x - y \;\Rightarrow$ $x^* = -1$ is subgradient. Moreover $\langle x^*, x\rangle = (-1)x = |x| = \|x\|$ and $\|x^*\|_{X^*} = |-1| = 1$.

3. The case $x > 0$ gives analogously $x^* = 1$ as subgradient.

Example 4.2. Let be X a reflexive Banach space, $B : X \to X^*$ a linear, bounded, non-negative, self-adjoint operator. Let be $f(x) = \langle Bx, x\rangle = x^T Bx, x \in X = \mathbb{R}^n$. In the first chapter, we have shown that f is convex. Now, taking in consideration the previous result

$$f(y) - f(x) \geq \langle 2Bx, y - x\rangle \;\forall y \in X,$$

it follows that $2Bx \in \partial f(x)$.

Theorem 4.1. *Let X be a real linear normed space and $f : X \to \overline{\mathbb{R}}$ a proper functional having at each point of X a subgradient (f subdifferentiable on X). Then f is convex and weak-lower semicontinuous (over the whole X).*

Proof. Let $x \in X$ and $x^* \in \partial f(x)$. Then for x_1 and $x_2 \in X$ we have

$$f(x_1) \geq f(x) + \langle x^*, x_1 - x\rangle$$

and

$$f(x_2) \geq f(x) + \langle x^*, x_2 - x\rangle.$$

After multiplying the first expression by $\lambda \in (0, 1)$ and the second by $(1 - \lambda)$, we sum up the two resulting relations. Hence

$$\lambda f(x_1) + (1 - \lambda)f(x_2) \geq f(x) + \langle x^*, \lambda x_1 + (1 - \lambda)x_2 - x\rangle.$$

Setting $x = \lambda x_1 + (1 - \lambda)x_2 - x$, we obtain

$$\lambda f(x_1) + (1 - \lambda)f(x_2) \geq f(\lambda x_1 + (1 - \lambda)x_2 - x),$$

so f is a convex function.

To prove that f is also weak-lower semicontinuous, consider a weakly convergent sequence $x_n \rightharpoonup x$ and $x^* \in \partial f(x)$. Then $f(x_n) \geq f(x) + \langle x^*, x_n - x\rangle$ and consequently

$$\varliminf_{n\to\infty} f(x_n) \geq f(x) + \varliminf_{n\to\infty} \langle x^*, x_n - x\rangle.$$

Hence f is weak-lower semicontinuous. $\square$

Now we will give a generalization to Theorem 3.6 (monotony of the gradient).

Theorem 4.2. *Let X be a linear normed space and $f : X \to \overline{\mathbb{R}}$ a proper function having at each $x \in X$ the subgradient $x^* = Ax \in X^*$ (A is in general a nonlinear operator). Then $A : X \to X^*$ is a monotone operator.*

Remark: Because of Theorem 4.1 f is convex.

Proof. We have

$$f(y) - f(x) \geq \langle Ax, y - x \rangle \ \forall x, y \in X$$

and

$$f(x) - f(y) \geq \langle Ay, x - y \rangle \ \forall x, y \in X.$$

By summing these two inequalities, we obtain

$$0 \geq \langle Ax - Ay, y - x \rangle \ \forall x, y \in X,$$

equivalent to

$$\langle Ay - Ax, y - x \rangle \geq 0 \ \forall x, y \in X,$$

therefore A is monotone. $\square$

Theorem 4.3. *Let X be a linear normed space, $A : X \to X^*$ a monotone operator and Ax the gradient of the proper function $f : X \to \overline{\mathbb{R}}$. Then $Ax \in X^*$ is a subgradient of f at x and $\partial f(x) = \{Ax\}$.*

Proof. Because of Theorem 3.7, it follows that f is convex. So when x is any fixed element of X we have for all $\lambda \in (0, 1)$ and all $y \in X$

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

Hence

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

For $\lambda \to 0$ follows

$$\langle f'(x), y - x \rangle = \langle Ax, y - x \rangle \leq f(y) - f(x),$$

therefore $Ax \in \partial f(x)$.

Let us consider a subgradient $x^* \in \partial f(x)$. It follows

$$f(x + \lambda y) - f(x) \geq \langle x^*, x + \lambda y - x \rangle = \lambda \langle x^*, y \rangle.$$

Dividing by λ and letting $\lambda \to 0$, we get $\langle f'(x), y \rangle \geq \langle x^*, y \rangle \ \forall y \in X$, i.e. $\langle f'(x) - x^*, y \rangle \geq 0 \ \forall y \in X$, hence $x^* = f'(x) = Ax$. $\square$

Theorem 4.4. *Let X be a linear normed space and $f : X \to \overline{\mathbb{R}}$ a convex function finite and continuous at $x_0 \in X$. Then $\partial f(x_0) \neq \emptyset$, i.e. f is subdifferentiable at x_0.*

Remarks:

(i) This a central (main) theorem in Convex Analysis and Optimization, with various applications.

(ii) With our former result (Proposition 2.1), we have $\partial f(x) \neq 0$ for all $x \in \operatorname{int}(\operatorname{dom} f)$, if f is bounded above in a neighborhood of x.

Proof. We divide the proof into five steps because of its complexity.

(i) From f convex and continuous at x_0 follows that $\operatorname{int}(\operatorname{epi} f) \neq \emptyset$, because there exists an open neighborhood $\mathcal{U}(x_0)$ of x_0 to any given $\varepsilon > 0$ such that $|f(x) - f(x_0)| < \varepsilon \ \forall x \in \mathcal{U}(x_0)$. Then the set

$$D := \{(x, \alpha) \in X \times \mathbb{R} : \alpha > f(x_0) + \varepsilon,\ x \in \mathcal{U}(x_0)\}$$

is open and

$$D \subset \text{epi } f \Rightarrow \operatorname{int}(\operatorname{epi} f) \neq \emptyset.$$

Especially, there is $x_0 \in \operatorname{int}(\operatorname{dom} f)$.

(ii) Further, $(x_0, f(x_0)) \in \partial\, \text{epi } f$ (by the definition of epi f).

(iii) The next step is to use a **Separation Theorem**, whose proof is available in the literature ([3]).

Theorem 4.5. *Let be X a linear normed space and V and W convex sets such that* int $V \neq \emptyset$ *and* (int $V) \cap W = \emptyset$. *Then there exists $x^* \in X^*$ $(x^* \neq 0)$ separating V and W, i.e. there exists a $c \in \mathbb{R} :\ \langle x^*, y\rangle \leq c \leq \langle x^*, x\rangle \ \forall x \in V \ \forall y \in W$* (so-called **weak separation**).

To apply this Separation Theorem, we set $V = \operatorname{epi} f$, that is a convex set, so int $V = \text{int(epi } f) \neq \emptyset$ (cf. above), $W = (x_0, f(x_0))$ and X is replaced by

$X \times \mathbb{R}$. Follows the existence of a functional $(x^*, \alpha^*) \in (X \times \mathbb{R})^* = X^* \times \mathbb{R}$ such that $\forall (x, \alpha) \in \operatorname{epi} f$ we have

$$\langle (x^*, \alpha^*), (x_0, f(x_0)) \rangle = \langle x^*, x_0 \rangle + \alpha^* f(x_0) \leq \langle x^*, x \rangle + \alpha^* \alpha. \qquad (4.\ 1)$$

(iv) Now we conclude that $\alpha^* > 0$. It is true because $\alpha^* = 0$ means $\langle x^*, x_0 \rangle \leq \langle x^*, x \rangle$ $\forall x \in \operatorname{epi} f$, i.e x^* separates x_0 and $\operatorname{dom} f$, which is a contradiction since $x_0 \in \operatorname{int}(\operatorname{dom}(f))$. For $\alpha^* \leq 0$ similar calculations guide us to another contradiction. By (4.1) follows

$$\langle x^*, x - x_0 \rangle \geq \alpha^* (f(x_0) - \alpha) \ \forall (x, \alpha) \in \operatorname{epi} f. \qquad (4.\ 2)$$

Consider a neighborhood $\mathcal{U}(x_0)$ of x_0 where for given $\varepsilon > 0$ holds $|f(x) - f(x_0)| < \varepsilon$, so in particular also $f(x) < f(x_0) + \varepsilon$. Choose $\overline{\alpha}$ such that $\overline{\alpha} > f(x_0) + \varepsilon > f(x)$ $\forall x \in \mathcal{U}(x_0)$. So $(x, \overline{\alpha}) \in \operatorname{epi} f$ $\forall x \in \mathcal{U}(x_0)$ (cf. Figure 4.3).

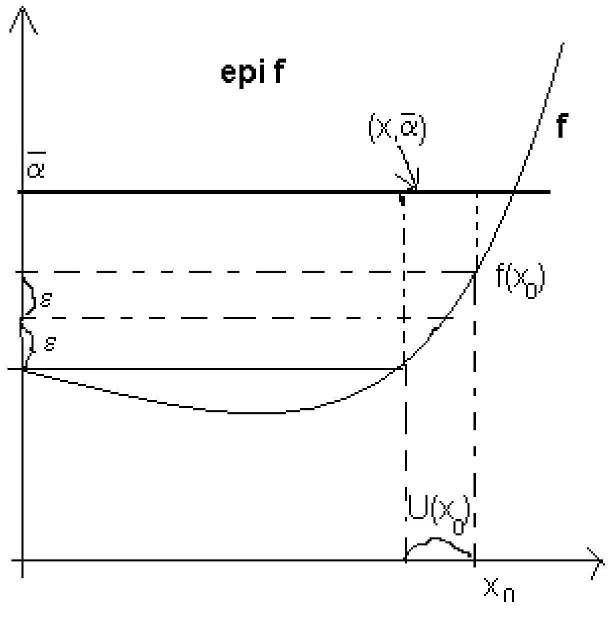

Figure 4.3

Inserting this $\overline{\alpha}$ (or any $\alpha \geq \overline{\alpha}$) into (4.2) leads to $f(x_0) - \overline{\alpha} < 0$. For the supposed $\alpha^* < 0$ results $\alpha^*(f(x_0) - \overline{\alpha}) > 0$, i.e. (cf. (4.2)) $\langle x^*, x - x_0 \rangle > 0$ $\forall x \in \mathcal{U}(x_0)$. But this is a contradiction (e.g. set $x = x_0$).

(v) We divide in (4.1) by $\alpha^*(> 0)$ and set $\alpha = f(x)$ (since $(x, f(x)) \in \text{epi } f$). It follows

$$\left\langle \frac{x^*}{\alpha^*}, x_0 \right\rangle + f(x_0) \leq \left\langle \frac{x^*}{\alpha^*}, x \right\rangle + f(x)\ \forall x \in \operatorname{dom} f.$$

Hence,

$$\left\langle -\frac{x^*}{\alpha^*}, x - x_0 \right\rangle \leq f(x) - f(x_0)\ \forall x \in X.$$

For $x \notin \operatorname{dom} f$ this is trivially fulfilled, because $f(x) = +\infty$ there. The last inequality means actually that $-\frac{x^*}{\alpha^*} \in \partial f(x_0)$, so f is subdifferentiable at x_0. □

Now, we present some important properties of the subgradient (subdifferential), partially without proof.

Theorem 4.6. *Let X be a linear normed space and $f : X \to \overline{\mathbb{R}}$ a subdifferentiable function. Then $\partial f(x)$ is convex and weak (sequentially) closed.*

Proof.

(i) Let be $x_1^*, x_2^* \in \partial f(x)$ and $\lambda \in (0, 1)$. It follows

$$\lambda f(y) - \lambda f(x) \geq \langle \lambda x_1^*, y - x \rangle\ \forall y \in X$$

and

$$(1 - \lambda) f(y) - (1 - \lambda) f(x) \geq \langle (1 - \lambda) x_2^*, y - x \rangle\ \forall y \in X.$$

By summing up these two inequalities, we obtain

$$f(y) - f(x) \geq \langle \lambda x_1^* + (1 - \lambda) x_2^*, y - x \rangle\ \forall y \in X,$$

i.e. $\lambda x_1^* + (1 - \lambda) x_2^* \in \partial f(x)$, so $\partial f(x)$ is a convex set.

(ii) Let $\{x_n^*\} \in \partial f(x)$ be a sequence with the property that $x_n^* \rightharpoonup x^*$. It follows

$$f(y) - f(x) \geq \langle x_n^*, y - x \rangle\ \forall y \in X.$$

Taking $n \to \infty$ we have

$$f(y) - f(x) \geq \lim_{n \to \infty} \langle x_n^*, y - x \rangle = \langle x^*, y - x \rangle\ \forall y \in X.$$

So $x^* \in \partial f(x)$, i.e. the set $\partial f(x)$ is weak (sequentially) closed. □

What about the usual well-known rules for differentiation for (classical) differentiable functions? These rules may be generalized to subdifferentials (subgradients), partially with some modifications and additional assumptions. This is called the **subdifferential calculus**.

For example, let be $f : X \to \overline{\mathbb{R}}$ and $\lambda > 0$. Then we have (obviously, by the definition of the subdifferential)

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

By the definition of the subgradient also immediately follows

$$\partial(f_1 + f_2)(x) \geq \partial f_1(x) + \partial f_2(x).$$

Proof. Let be $x_1^* \in \partial f_1(x)$ and $x_2^* \in \partial f_2(x)$. By the definition of subgradient, follows

$$f_1(y) - f_1(x) \geq \langle x_1^*, y - x\rangle \ \forall y \in X$$

and

$$f_2(y) - f_2(x) \geq \langle x_2^*, y - x\rangle \ \forall y \in X.$$

By summing this two inequalities, one gets

$$(f_1 + f_2)(y) - (f_1 + f_2)(x) \geq \langle x_1^* + x_2^*, y - x\rangle \ \forall y \in X,$$

so $x_1^* + x_2^* \in \partial(f_1 + f_2)(x)$. □

The following question arises, as a generalization of the same rule of the classical differential calculus: When does equality $\partial(f_1+f_2)(x) = \partial f_1(x)+\partial f_2(x)$ hold?

We use the algebraic sum of sets

$$A, B \subset X : \ A + B = \{a + b : \ a \in A, \ b \in B\}$$

and $A + \emptyset = A$.

Theorem 4.7. (Moreau, Rockafellar) *The sum rule*

$$\partial(f_1 + \ldots + f_n)(x) = \partial f_1(x) + \ldots + \partial f_n(x), \ n \geq 2$$

holds for all $x \in X$ (where X is a normed space or, more general, real locally convex space) (cf. [4], p. 389), when the following assumptions are fulfilled simultaneously

(i) $f_1, \ldots, f_n : X \to \overline{\mathbb{R}}$ *are proper and convex functions.*

(ii) There exists an $x_0 \in X$ such that all $f_i(x_0)$, $i = 1, \ldots, n$, are finite and all f_i's, except at most one f_j, $j \in \{1, \ldots, n\}$ are continuous at x_0 (in particular $x_0 \in \bigcap_{i=1}^{n} \mathrm{dom}\ f_i$).

The proof is available in [4], so we omit it here.

Remark: The main idea and part of the proof consists in the usage of the Separation Theorem 4.5 again!

5 Conjugate Functionals

Conjugate functionals play an important role in **Convex Analysis**, in particular in the **duality theory**. They have useful and interesting properties and important connections to subdifferentials.

Let X be a linear normed space and X^* its topological dual space.

Definition 5.1. *Consider the function $f : X \to \overline{\mathbb{R}}$. The function $f^* : X^* \to \overline{\mathbb{R}}$*

$$f^*(x^*) := \sup_{x \in X}\{\langle x^*, x\rangle - f(x)\}$$

is called **conjugate functional** *to f (sometimes also denoted as* **polar functional***).*

Geometrical interpretation. (cf. Figure 5.1)

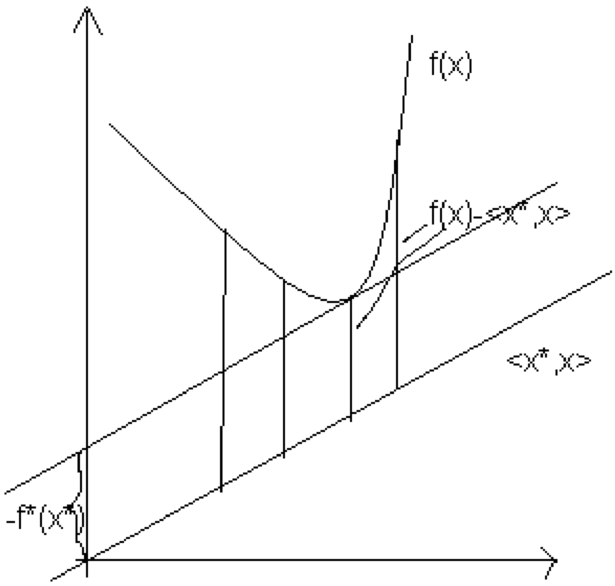

Figure 5.1

According to the definition we have

$$\inf_{x \in X}\{f(x) - \langle x^*, x\rangle\} = -\sup_{x \in X}\{\langle x^*, x\rangle - f(x)\} = -f^*(x^*).$$

There is also **another geometrical interpretation**. Consider affine functions less than $f(x)$ (affine minorants) $\langle x^*, x\rangle - \alpha \le f(x)\ \forall x \in X$. This is equivalent to

$$\langle x^*, x\rangle - f(x) \leq -\alpha \ \ \forall x \in X.$$

The **smallest** such α exists for $\sup\limits_{x\in X}\{\langle x^*, x\rangle - f(x)\}$ which is $f^*(x^*)$ and this holds for the **greatest** $-\alpha$ which defines by $\langle x^*, x\rangle - \alpha$ the **greatest affine function less than** f.

Examples 5.1.

(i) Let $f(x) = x^2, \ \ x \in \mathbb{R}$. Its conjugate function is $f^*(x^*) = \sup\limits_{x\in\mathbb{R}}\{x^*x - x^2\}$. To determine a simplified formula to it, denote $h(x) = x^*x - x^2$, that is a derivable function. Then $h'(x) = x^* - 2x = 0$ implies $x = \frac{1}{2}x^*$. The second derivative is $h''(x) = -2 < 0 \ \forall x \in \mathbb{R}$, therefore the function h has a maximum (global, since $\lim\limits_{x\to\pm\infty} h(x) = -\infty$) point at $x = \frac{1}{2}x^*$. It follows that $f^*(x^*) = \max\limits_{x\in\mathbb{R}} h(x) = h(\frac{1}{2}x^*) = x^*\frac{1}{2}x^* - (\frac{1}{2}x^*)^2 = \frac{1}{4}x^{*2}$. Moreover, for $f(x) = \frac{1}{2}x^2$, we have $f^*(x^*) = \frac{1}{2}x^{*2}$ hence $f^* \equiv f$, as functions defined on $\mathbb{R}$.

(ii) Let $f(x) = e^x, \ x \in \mathbb{R}$. Then $f^*(x^*) = \sup\limits_{x\in\mathbb{R}}\{x^*x - e^x\}$. Again we denote the function used to compute the conjugate function by $h(x) = x^*x - e^x$, a derivable function. Consequently, $h'(x) = x^* - e^x = 0$ yields $x^* > 0$ and $x = \ln x^*$. Since $h''(x) = -e^x < 0 \ \forall x \in \mathbb{R}$, h has a maximum (global, since $\lim\limits_{x\to\pm\infty} h(x) = -\infty$) at $x = \ln x^*$. Hence

$$f^*(x^*) = \begin{cases} x^*(\ln x^* - 1), & \text{if } x^* > 0, \\ 0, & \text{if } x^* = 0, \\ \infty, & \text{if } x^* < 0. \end{cases}$$

(iii) Let $f(x) = \begin{cases} x(\ln x - 1), & \text{if } x > 0, \\ 0, & \text{if } x = 0, \\ +\infty, & \text{if } x < 0. \end{cases}$

Then $f^*(x^*) = \sup\limits_{x>0}\{x^*x - x(\ln x - 1)\}$. Let us introduce the function h defined by $h(x) = x^*x - f(x)$ for $x > 0$, $h(0) = 0$ and $h(x) = -\infty$ if $x < 0$. It is derivable on $(0, +\infty)$. For $x > 0$ we have $h'(x) = x^* - (\ln x - 1) - x\frac{1}{x} =$

$x^* - \ln x = 0$ that implies $x = e^{x*}$. Because $h''(e^{x*}) = -\frac{1}{e^{x*}} < 0$ it follows that e^{x*} is a point where h attains its maximum over $(0, +\infty)$. But this is a global maximum, because

$$\begin{aligned}\lim_{x\to 0^+} h(x) &= \lim_{x\to 0^+} x(x^* - (\ln x - 1)) = \lim_{x\to 0^+} \frac{x^* - \ln x + 1}{\frac{1}{x}} \\ &= \lim_{x\to 0^+} \frac{-\frac{1}{x}}{-\frac{1}{x^2}} = 0 = h(0)\end{aligned}$$

and $\lim\limits_{x\to\infty} h(x) = -\infty = h(x)\ \forall x < 0 \quad \forall x^* \in \mathbb{R}$. It follows $f^*(x^*) = x^* e^{x*} - e^{x*}(\ln e^{x*} - 1) = e^{x*} \quad \forall x^* \in \mathbb{R}$.

Remark: From (ii) and (iii) follows for $f(x) = e^x$, $f^{**}(x) = (f^*)^*(x) = f(x)$ (i.e. $f^{**} \equiv f$).

Later we will see that this property ($f^{**} \equiv f$) is fulfilled for each weak-lower semicontinuous and proper function on X.

(iv) Let us consider an affine function on $\mathbb{R}$ $f(x) = mx + n$. Then $f^*(x^*) = \sup\limits_{x\in\mathbb{R}} \{x^*x - (mx + n)\}$. As before, denote $h(x) = x^*x - (mx + n)$. Because it is a derivable function, we can proceed to determine its maximum point by the well-known method.

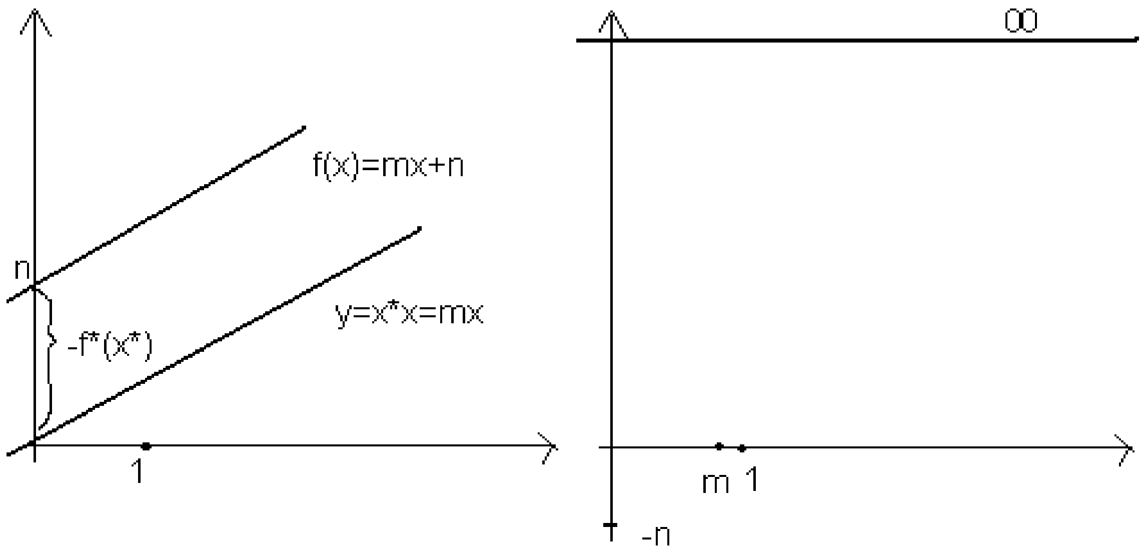

Figure 5.2

Therefore $h'(x) = x^* - m = 0$ implies $x^* = m$, so $f^*(m) = -n$. Consider $x^* \neq m$. If $x^* > m$ then $h(x)$ tends to $+\infty$ when $x \to +\infty$, while for $x^* < m$ one has $h(x) \to +\infty$ when $x \to -\infty$.

Hence (see also Figure 5.2)

$$f^*(x^*) = \begin{cases} -n, & \text{if } x^* = m, \\ +\infty, & \text{otherwise .} \end{cases}$$

(v) Let $F : \mathbb{R} \to \overline{\mathbb{R}},\ f(x) = \begin{cases} -n, & \text{if } x = m, \\ +\infty, & \text{otherwise.} \end{cases}$

It follows $f^*(x^*) = \sup\limits_{x \in \mathbb{R}} \{x^* x - f(x)\} = x^* m + n$, hence $f^{**} \equiv f$.

(vi) Let $f(x) = \frac{1}{3}x^3, x \in \mathbb{R}$. This is a non-convex function. Then $f^*(x^*) = \sup\limits_{x \in \mathbb{R}} \left\{x^* x - \frac{1}{3}x^3\right\}$. Denote $h(x) = x^* x - \frac{1}{3}x^3$, that is derivable. Further $h'(x) = x^* - x^2 = 0$ implies $x = \sqrt{x^*}$ for $x^* \geq 0$ supposed. Regarding the second derivative one has $h''(x) = -2x = -2\sqrt{x^*} < 0$ if $x^* > 0$. Thus $x = \sqrt{x^*}$, where $x^* \geq 0$ is a maximum point, but it is only local, because $\lim\limits_{x \to -\infty} h(x) = \lim\limits_{x \to -\infty} x(x^* - \frac{1}{3}x^2) = \infty\ \forall x^* \in \mathbb{R}$ hence $f^*(x^*) \equiv +\infty\ \forall x^* \in \mathbb{R}$, i.e. f^* is not a proper function.

(vii) If $f(x) = \frac{1}{3}|x|^3$ (a convex and weak-lower semicontinuous function), then $f^{**} \equiv f$. Moreover, $f^*(x^*) = \frac{2}{3}\|x^*\|^{\frac{3}{2}}$.

(viii) More general: let be $f(x) = \frac{1}{p}|x|^p,\ p > 1$. Then (cf [3]) $f^*(x^*) = \frac{1}{q}|x^*|^q, 1 < q < +\infty, \frac{1}{p} + \frac{1}{q} = 1$.

(ix) Let be A a convex set and consider its indicator function,

$$\chi_A = \begin{cases} 0, & x \in A, \\ +\infty, & \text{otherwise.} \end{cases}$$

The conjugate function to χ_A is

$$\chi_A^*(x^*) = \sup_{x \in X}\{\langle x^*, x\rangle - \chi_A(x)\} = \sup_{x \in A}\langle x^*, x\rangle = S_A(x^*),$$

and this is actually the **support functional** of A (cf. Chapter 1).

(x) Let $(X, \|\cdot\|)$ be a normed space and $f(x) = \|x\|, x \in X$. To calculate the conjugate function to f,

$$f^*(x^*) = \sup_{x\in X}\{\langle x^*, x\rangle - \|x\|\},$$

two cases are distinguished during calculations

(a) If $\|x\| \leq 1$, then $\langle x^*, x\rangle \leq \|x^*\|\|x\| \leq \|x\|\ \forall x \in X$. Therefore $\langle x^*, x\rangle - \|x\| \leq 0$, equality being attained for example at $x = 0$. Hence $f^*(x^*) = 0$ when $\|x\| \leq 1$.

(b) If $\|x\| > 1$, then

$$\|x^*\|_{X^*} := \sup_{x\neq 0, x\in X}\left\{\frac{\langle x^*, x\rangle}{\|x\|}\right\} > 1,$$

so there is and $x_0 \in X\backslash\{0\}$ such that $\frac{\langle x^*, x_0\rangle}{\|x_0\|} > 1$, i.e. $\langle x^*, x_0\rangle - \|x_0\| > 0$. Then

$$\begin{aligned} f^*(x^*) &= \sup_{x\in X}\{\langle x^*, x\rangle - \|x\|\} \\ &\geq \sup_{\lambda>0}\{\langle x^*, \lambda x_0\rangle - \|\lambda x_0\|\} \\ &= \sup_{\lambda>0}\lambda(\langle x^*, x_0\rangle - \|x_0\|) = +\infty \end{aligned}$$

Therefore $f^*(x^*) = \begin{cases} 0, & \text{if } \|x^*\|_* \leq 1, \\ +\infty, & \text{if } \|x^*\|_* > 1. \end{cases}$

Hence $f^*(x^*) = \chi_{B^*}(x^*)$ the indicator functional of the dual unit ball B^* in X^*, $B^* = \{x^* \in X^* :\ \|x^*\|_* \leq 1\}$.

(xi) Let be $\varphi : \mathbb{R} \to \overline{\mathbb{R}}$ a proper even function, φ^* its conjugate and X a normed space. Let be $f(x) = \varphi(\|x\|), x \in X$. Then $f^*(x^*) = \varphi^*(\|x^*\|_*)$ because

$$\begin{aligned} f^*(x^*) &= \sup_{x\in X}\{\langle x^*, x\rangle - \varphi(\|x\|)\} \\ &= \sup_{t\geq 0}\sup_{\|x\|=t}\{\langle x^*, x\rangle - \varphi(t)\} \\ &= \sup_{t\geq 0}\left[\sup_{\|x\|=t}\{\langle x^*, x\rangle - \varphi(t)\}\right] \\ &= \sup_{t\geq 0}\{t\|x^*\|_* - \varphi(t)\}. \end{aligned}$$

Since $\varphi(t) = \varphi(-t)$, we have

$$f^*(x^*) = \sup_{t \in \mathbb{R}} \{t\|x^*\|_* - \varphi(t)\} = \varphi^*(\|x^*\|_*).$$

For example, let be $\varphi(t) = \frac{1}{2}t^2$. Then $\varphi^*(t^*) = \frac{1}{2}t^{*2}$. Consequently,

$$f(x) = \varphi(\|x\|) = \frac{1}{2}\|x\|^2 \text{ and } f^*(x^*) = \varphi^*(\|x^*\|_*^2).$$

Some Elementary Properties of Conjugate Functionals

Let X be a linear normed space and X^* its topological dual space. The functions that appear below are defined on X and have real values.

(1) (Young's inequality) $f(x) + f^*(x^*) \geq \langle x^*, x\rangle \; \forall x \in X \; \forall x^* \in X^*$.
This result follows immediately from the definition of f^*.

(2) $f^*(0) = \sup_{x \in X} \langle 0, x\rangle - f(x)\} = \sup_{x \in X}(-f(x)) = -\inf_{x \in X} f(x)$. Many applications in optimization use the equivalent formulation $\inf_{x \in X} f(x) = -f^*(0)$.

(3) $f \leq g$ implies $f^* \geq g^*$.

(4) $(\sup_{i \in I} f_i)^* \leq \inf_{i \in I} f_i^*$, where I is any index set, because

$$\begin{aligned}(\sup_{i \in I} f_i)^*(x^*) &= \sup_{x \in X}\{\langle x^*, x\rangle - (\sup_{i \in I} f_i(x)\} \\ &\leq \sup_{x \in X}\{\langle x^*, x\rangle - f_i(x)\} \\ &= f_i^*(x^*) \; \forall i \in I \; \forall x^* \in X^*.\end{aligned}$$

Analogously, there holds $(\inf_{i \in I} f_i)^* \geq \sup_{i \in I} f_i^*$.

(5) $(\lambda f)^*(x^*) = \lambda f^*(\frac{1}{\lambda}x^*) \; \forall \lambda > 0$, because

$$\begin{aligned}(\lambda f)^*(x^*) &= \sup_{x \in X}\{\langle x^*, x\rangle - (\lambda f(x))\} \\ &= \sup_{x \in X}\left\{\lambda\left[\left\langle \frac{1}{\lambda}x^*, x\right\rangle - f(x)\right]\right\} = \lambda f^*\left(\frac{1}{\lambda}x^*\right).\end{aligned}$$

(6) Consider the translation function $f_\alpha(x) = f(x-\alpha), \alpha \in X,\ x \in X$. One has

$$\begin{aligned} (f_\alpha^*(x^*) &= \sup_{x\in X}\{\langle x^*, x\rangle - f(x-\alpha)\} \\ &= \sup_{x\in X}\{\langle x^*, x-\alpha\rangle - f(x-\alpha)\} + \langle x^*, \alpha\rangle \\ &= f^*(x^*) + \langle x^*, \alpha\rangle. \end{aligned}$$

Proposition 5.1. *Let X be a linear normed space and $f : X \to \overline{\mathbb{R}}$. Then f^* is convex and weak-lower semicontinuous.*

Proof.

(a) Let be $x_1^*, x_2^* \in X^*$ and $\lambda \in [0,1]$. Concerning the convexity of f^* we have

$$\begin{aligned} f^*(\lambda x_1^* + (1-\lambda)x_2^* &= \sup_{x\in X}\{\langle \lambda x_1^* + (1-\lambda)x_2^*, x\rangle - f(x)\} \\ &\leq \lambda \sup_{x\in X}\{\langle x_1^*, x\rangle - f(x)\} + \sup_{x\in X}\{\langle x_2^*, x\rangle - f(x)\} \\ &= \lambda f^*(x_1^*) + (1-\lambda) f^*(x_2^*). \end{aligned}$$

(b) Let be $x_n^* \rightharpoonup x^*$ (in X^*). Applying Young's inequality, we obtain

$$f^*(x_n^*) \geq \langle x_n^*, x\rangle - f(x)\ \forall x \in X.$$

Then follows

$$\varliminf_{n\to\infty} f^*(x_n^*) \geq \varliminf_{n\to\infty}\{\langle x_n^*, x\rangle - f(x)\} = \langle x^*, x\rangle - f(x) \ \ \forall x \in X.$$

Hence

$$\varliminf_{n\to\infty} f^*(x_n^*) \geq \sup_{x\in X}\{\langle x^*, x\rangle - f(x)\} = f^*(x^*),$$

therefore f is weak-lower semicontinuous. □

Proposition 5.2. *Let X be a linear normed space and $f : X \to \overline{\mathbb{R}}$ a proper convex and weak-lower semicontinuous function. Then f^* is a proper functional.*

Proof. To help the reader acquire a better understanding of the proof we divide it into four steps. A separation theorem is also being used.

- Because f is proper, it follows the existence of an $x \in X$ such that $f(x) < \infty$ and so

$$f^*(x^*) = \sup_{y \in X}\{\langle x^*, y\rangle - f(y)\} \geq \langle x^*, x\rangle - f(x) > -\infty \ \forall x^* \in X^*.$$

- Since $\operatorname{epi} f$ is (sequentially) weakly closed, $\operatorname{epi} f$ is closed (cf. Theorem 3.8). For any $d > 0$ and $x \in X$ such that $f(x) < +\infty$ holds $(x, f(x) - d) \notin \operatorname{epi} f$.

Now we prove a Separation Theorem, which is a conclusion of the Separation Theorem 4.5.

Separation Theorem 5.1. *Let X be a linear normed space, $W \subset X$ closed and convex and $x_0 \notin W$. Then there exists $x^* \in X^*$, $x^* \neq 0$, strictly separating W and x_0, i.e. there is a $c \in \mathbb{R}$ such that $\langle x^*, x\rangle < c \ \forall x \in W$ and $\langle x^*, x_0\rangle > c$.*

Proof. Since W is closed and x_0 is not in W, it follows that there exists an open convex neighborhood of x_0 $V = \mathcal{U}(x_0)$ such that $W \cap V = W \cap \operatorname{int} V = \emptyset$. Thus we can use the Separation Theorem 4.5. It follows that there exist $x^* \in X^*$, $x^* \neq 0$ and $d \in \mathbb{R}$ such that V and W are weakly separated, i.e. (see also Figure 5.3) $\langle x^*, x\rangle \leq d \leq \langle x^*, y\rangle \ \forall x \in W \ \forall y \in \mathcal{U}(x_0) = V$.

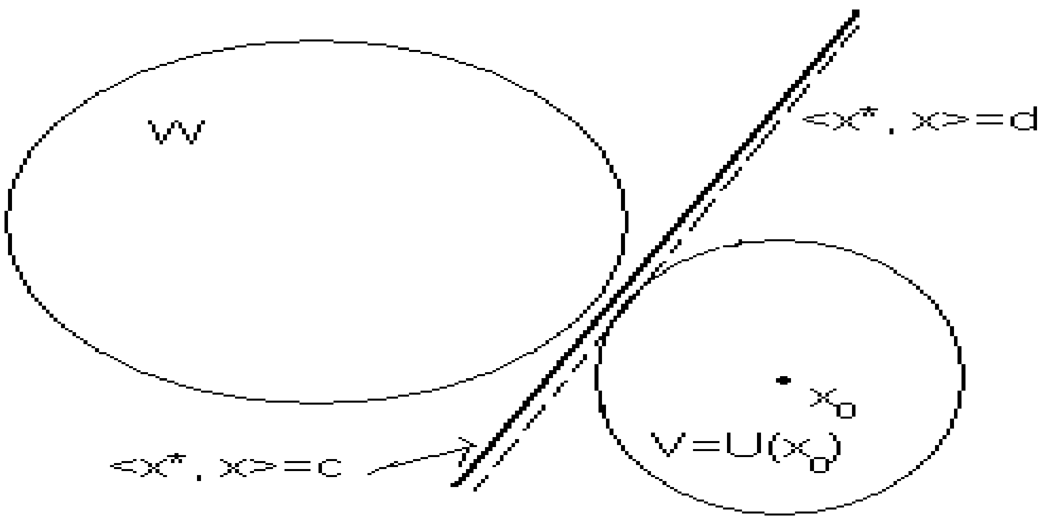

Figure 5.3

This supporting hyperplane may be translated a little bit in the direction of x_0 (because separates $\mathcal{U}(x_0)$) such that strict separation holds

$$\exists c \in \mathbb{R} : \ \langle x^*, x \rangle < c < \langle x^*, x_0 \rangle \ \forall x \in W. \square$$

Now we continue with the proof of Proposition 5.2.

- Applying Separation Theorem 5.1 to $W = \text{epi } f$ ($\text{epi}\, f$ is a closed convex set), $x_0 = (x, f(x) - d) \notin W$ and X replaced by $X \times \mathbb{R}$, we obtain some $x^* \in X^*$ and $\alpha \in \mathbb{R}$ such that

$$\begin{aligned} \langle (x^*, \alpha^*), (x, f(x) - d) \rangle &= \langle x^*, x \rangle + \alpha^*(f(x) - d) \\ &> \langle (x^*, \alpha^*), (y, \alpha) \rangle \\ &= \langle x^*, y \rangle + \alpha^* \alpha \ \forall (y, \alpha) \in \text{epi } f. \quad (5.\ 1) \end{aligned}$$

  We want to divide by $-\alpha^*$.

- We show that $\alpha^* < 0$.

  To the contrary, let be $\alpha^* > 0$. It follows that the right-hand side of the above relation tends to $+\infty$ (with $\alpha \to \infty$) and this is a contradiction to the same relation.

  For $\alpha^* = 0$ holds $\langle x^*, x \rangle > \langle x^*, y \rangle \ \ \forall y \in \text{dom}\, f$. Since $x \in \text{dom}\, f$ this is a contradiction, too.

- Setting $x_1^* := -\frac{x^*}{\alpha^*}$ and $\alpha := f(y)$ for $y \in \text{dom}\, f((y, \alpha) \in \text{epi}\, f)$, (5.1) leads to $\langle x_1^*, x \rangle - f(x) + d > \langle x_1^*, y \rangle - f(y) \ \forall y \in \text{dom}\, f$ and even $y \in X$, since $f(y) = +\infty$ if $y \notin \text{dom}\, f$. Therefore $+\infty > \langle x_1^*, x \rangle - f(x) + d \geq \sup\limits_{y \in X} \{\langle x_1^*, y \rangle - f(y)\} = f^*(x_1^*)$, i.e. f is proper. $\square$

There is possible to introduce also the so-called **biconjugate** of a function (functional) f.

Definition 5.2. *Let $f^* : X^* \to \overline{\mathbb{R}}$ be the conjugate function to $f : X \to \overline{\mathbb{R}}$, with X a linear normed space and X^* its dual. Then the function*

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

is called the **biconjugate** *function to* f.

Remark: From Young's inequality it follows

$$f^{**}(x) = \sup_{x^* \in X^*} \{\langle x^*, x\rangle - f^*(x^*)\} \leq \sup_{x^* \in X^*} \{\langle x^*, x\rangle - \langle x^*, x\rangle + f(x)\} = f(x),$$

therefore $f^{**}(x) \leq f(x)$ always stands.

The following statement gives an answer to the natural question regarding the equality case between f and f^{**}.

Theorem 5.2. *Let* $f : X \to \overline{\mathbb{R}}$ *be a proper function and* X *a linear normed space. Then* f *is convex and weak-lower semicontinuous if and only if* $f^{**} \equiv f$ *(i.e.* $f^{**}(x) = f(x) \ \ \forall x \in X$*).*

Proof. *Necessity.* Let the proper functional f be convex and weak-lower semicontinuous. Then $f^{**}(x) \leq f(x) \ \ \forall x \in X$ (cf. remark above).

We show the reverse inequality $f^{**}(x) \geq f(x) \ \ \forall x \in X$. Let be $x \notin \operatorname{dom} f^{**}$, i.e. $f^{**}(x) = \infty$. Because of Proposition 5.1 f^* is convex and weak-lower semicontinuous and so f^{**} enjoys the same properties, too. Because of Proposition 5.2 f^* is a proper function and again from Proposition 5.2 (replacing f by f^* and f^* by f^{**}) follows that f^{**} is proper. Consequently $f^{**}(x) > -\infty \ \ \forall x \in X$. Therefore for $x \notin \operatorname{dom} f^{**}$, $\infty = f^{**} \geq f(x)$ is fulfilled.

Now, let be $x \in \operatorname{dom} f^{**}$ and suppose $f(x) > f^{**}(x)$. Let us introduce also $d := \frac{1}{2}(f(x) - f^{**}(x)) > 0$. We apply the considerations from the proof of Proposition 5.2, i.e. we use the Separation Theorem 5.1 to separate $W = \operatorname{epi} f$ (closed and convex set) and $(x, f(x) - d) \notin W$. Then there exists $x_1^* \in X^*,\ x_1^* \neq 0$, such that

$$\langle x_1^*, x\rangle - f(x) + \frac{1}{2}(f(x) - f^{**}(x)) \geq f^*(x_1^*)$$

and therefore

$$\langle x_1^*, x\rangle - \frac{1}{2}f(x) - \frac{1}{2}f^{**}(x) \geq f^*(x_1^*) \geq \langle x_1^*, x\rangle - f^*(x_1).$$

So $f^{**}(x) \geq f(x)$. But this contradicts our assumption $\Rightarrow f^{**} \equiv f$.

Sufficiency. Now consider $f^{**} \equiv f$ and f proper. Then f^{**}, as the conjugate of

f^*, is convex and weak-lower semicontinuous (cf. Proposition 5.1) and therefore f is convex and weak-lower semicontinuous, too. $\square$

Remark: It is not necessary for the function f to be proper in order to prove the sufficiency in Theorem 5.2.

Conclusion 5.1. *For any function f there is $f^{***} \equiv f^*$. (If f^* is proper, then the assertion is straightforward, but if this is not the case the assertion is not trivial).*

Proof. Because of Proposition 5.1, f^* is convex and weak-lower semicontinuous. Using Theorem 5.2 with f replaced by f^*, we see that $f^{***} \equiv f^*$, if f^* is proper. When f^* is not proper, we have to consider two cases.

(i) If there is an x_0^* such that $f^*(x_0^*) = -\infty$ one has

$$f^{**}(x) = \sup_{x^* \in X^*} \{\langle x^*, x\rangle - f^*(x^*)\} = \infty \ \forall x \in X.$$

Moreover

$$f^*(x^*) = \sup_{x \in X}\{\langle x^*, x\rangle - f(x)\} \geq -\infty = f^*(x_0^*) = \sup\{\langle x_0^*, x\rangle - f(x)\}$$

and so $f(x) = \infty \quad \forall x \in X$. Therefore $f^*(x^*) = \sup_{x\in X}\{\langle x^*, x\rangle - f(x)\}$ $\forall x^* \in X^*$, i.e. $f^* \equiv -\infty$.

Finally, $f^{***}(x^*) = \sup_{x\in X}\{\langle x^*, x\rangle - f^{**}(x)\} = -\infty \quad \forall x^* \in X^*$ leading to $f^{***}(x^*) = f^*(x^*) = -\infty \quad \forall x^* \in X^*$.

(ii) Suppose that there is no x_0^* such that $f^*(x_0^*) = -\infty$ (otherwise f^* would be proper). It follows

$$f^{**}(x) = \sup_{x^* \in X^*} \{\langle x^*, x\rangle - f^*(x^*)\} = -\infty \ \forall x \in X.$$

Further we have

$$f^{***}(x^*) = \sup_{x \in X}\{\langle x^*, x\rangle - f^{**}(x)\} = +\infty \ \forall x^* \in X^*$$

therefore

$$f^{***} \equiv f^* \equiv +\infty. \square$$

Now we are going to derive assertions concerning the connections between conjugate functionals and subdifferentials.

Theorem 5.3. *Let X be a linear normed space, the function $f : X \to \overline{\mathbb{R}}$ and its conjugate $f^* : X^* \to \overline{\mathbb{R}}$. Then for all $x \in X$, $x^* \in \partial f(x)$ if and only if $f(x) + f^*(x^*) = \langle x^*, x\rangle$.*

Remark: In general Young's inequality asserts

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

Therefore, it is fulfilled here as equality.

Proof. *Necessity.* Let be $x^* \in \partial f(x)$. Then we have

$$f(y) - f(x) \geq \langle x^*, y - x\rangle \ \forall y \in X,$$

i.e.

$$\langle x^*, x\rangle - f(x) \geq \langle x^*, y\rangle - f(y) \ \forall y \in X,$$

that implies

$$\langle x^*, x\rangle - f(x) \geq \sup_{y \in X}\{\langle x^*, y\rangle - f(y)\} = f^*(x^*).$$

On the other hand, because of Young's inequality, we have $\langle x^*, x\rangle - f(x) \leq f^*(x^*)$, therefore one can conclude

$$f(x) + f^*(x^*) = \langle x^*, x\rangle.$$

Sufficiency. Assume that for an $x \in X$ and $x^* \in X^*$ we have $f(x) + f^*(x^*) = \langle x^*, x\rangle$. Then

$$\langle x^*, x\rangle - f(x) = f^*(x^*) = \sup_{y \in X}\{\langle x^*, y\rangle - f(y)\} \geq \langle x^*, y\rangle - f(y) \ \forall y \in X.$$

Hence

$$f(y) - f(x) \geq \langle x^*, y - x\rangle \ \forall y \in X,$$

i.e. $x^* \in \partial f(x)$. $\square$

Conclusion 5.2. *From Theorem 5.3 immediately follows*

$$\partial f(x) = \{x^* \in X^* : \ f(x) + f^*(x^*) \leq \langle x^*, x\rangle\}.$$

This means $f(x) + f^*(x^*) = \langle x^*, x\rangle$, since $f(x) + f^*(x^*) \geq \langle x^*, x\rangle$ is true in any case.

Theorem 5.4. *Let be X a linear normed space.*

(a) Let be $f : X \to \overline{\mathbb{R}}$ a proper functional. If $x^* \in \partial f(x)$, then $x \in \partial f^*(x^*)$.

(b) Let be $f : X \to \overline{\mathbb{R}}$ a proper, convex and weak-lower semicontinuous function. Then $x^* \in \partial f(x)$ if and only if $x \in \partial f^*(x^*)$.

Proof.

(a) Let be $x^* \in \partial f(x)$. For any $y^* \in X^*$ by Young's inequality follows

$$f^*(y^*) = \sup_{y \in X}\{\langle y^*, y\rangle - f(y)\}) \geq \langle y^*, x\rangle - f(x).$$

Theorem 5.3 yields $f(x) = \langle x^*, x\rangle - f^*(x^*)$, therefore $f^*(y^*) \geq \langle y^* - x^*, x\rangle + f^*(x^*)$, i.e. $x \in \partial f^*(x^*)$.

(b) Let f be proper, convex and weak-lower semicontinuous. Because of (a) it remains to point out that from $x \in \partial f^*(x^*)$ follows $x^* \in \partial f(x)$. Therefore, let be $x \in \partial f^*(x^*)$. Theorem 5.2 implies $f^{**} \equiv f$. Applying assertion (a) in this theorem to f^* instead of f yields that from $x \in \partial f^*(x^*)$ follows $x^* \in \partial f^{**}(x) = \partial f(x)$. □

Theorem 5.5. *Let X be a linear normed space and $f : x \to \overline{\mathbb{R}}$ a function. If $\partial f(x_0) \neq \emptyset$, then $f(x_0) = f^{**}(x_0)$ (i.e. f is subdifferentiable at x_0).*

Proof. From a previous remark, we know that $f^{**}(x_0) \leq f(x_0)$. Therefore it is enough to show $f^{**}(x_0) \geq f(x_0)$. By assumption there exists an $x_0^* \in \partial f(x_0)$. With Theorem 5.3 we conclude $f(x_0) + f^*(x_0^*) = \langle x_0^*, x_0\rangle$. Hence

$$f^{**}(x_0) = \sup_{x^* \in X^*}\{\langle x^*, x_0\rangle - f^*(x^*)\} = \langle x_0^*, x_0\rangle - f^*(x_0^*) = f(x_0).\square$$

Theorem 5.6. *Let be X a linear normed space. If for $x_0 \in X$ applies $f(x_0) = f^{**}(x_0)$, then $\partial f(x_0) = \partial f^{**}(x_0)$.*

Proof. Let $x_0 \in \partial f(x_0)$. Using Theorem 5.3 and Conclusion 5.1, as well as $f(x_0) = f^{**}(x_0)$, we have

$$\langle x_0^*, x_0 \rangle = f(x_0) + f^*(x_0^*) = f^{**}(x_0) + f^{***}(x_0^*).$$

Again, by Theorem 5.3 (opposite direction and f^{**} instead of f) follows $x_0^* \in \partial f^{**}(x_0)$, implying $\partial f(x_0) \subseteq \partial f^{**}(x_0)$. Now, let $x_0^* \in \partial f^{**}(x_0)$. As above, (Theorem 5.3, applied to f^{**} instead of f), follows

$$\langle x_0^*, x_0 \rangle = f^{**}(x_0) + f^{***}(x_0^*) = f(x_0) + f^*(x_0^*).$$

Again, Theorem 5.3 provides $x_0^* \in \partial f(x_0)$, that yields $\partial f^{**}(x_0) \subseteq \partial f(x_0)$. Further this prompts $\partial f^{**}(x_0) = \partial f(x_0)$. □

Remark: If $\partial f(x_0) \neq \emptyset$ follows $f(x_0) = f^{**}(x_0)$ and by Theorem 5.6 one gets $\partial f^{**}(x_0) = \partial f(x_0)$.

6 Duality and Convex Analysis

6.1 Primal and Dual Optimization Problems

In this section we associate a so-called **dual problem** to a given **convex optimization problem**, in this context called **primal problem**. For both problems we will prove **duality assertions**. **Conjugate functionals** and **subdifferentials** (and therefore implicitly and explicitly **separation theorems**) play an **important role** for that together with the notion of **stability**.

Let us introduce first some general assumptions used within this chapter.

Let X and Y be linear normed spaces and $f : X \to \overline{\mathbb{R}}$ a proper function. We consider an optimization problem called **primal problem** $(\mathcal{P})$

$$(\mathcal{P}) \qquad \inf_{x \in X} f(x).$$

Definition 6.1. *A function $\Phi : X \times Y \to \overline{\mathbb{R}}$ is said to be a* **perturbation function** *of f if $\Phi(x, 0) = f(x) \quad \forall x \in X$.*

For all $y \in Y$ the problem

$$(\mathcal{P}_y) \qquad \inf_{x \in X} \Phi(x, y)$$

is called a **perturbed problem** *to $(\mathcal{P})$. The variable y is termed the* **perturbation variable** *(or parameter). For $y = 0$, clearly,*

$$(\mathcal{P}_0) \cong (\mathcal{P}),$$

i.e. $(\mathcal{P}_0)$ is nothing but $(\mathcal{P})$.

Now we are going to define a dual problem $(\mathcal{P}^*)$ to $(\mathcal{P})$. This is defined by means of the conjugate function $\Phi^* : (X \times Y)^* \cong X^* \times Y^* \to \overline{\mathbb{R}}$,

$$\Phi^*(x^*, y^*) = \sup_{x \in X, y \in Y} \{\langle x^*, x \rangle + \langle y^*, y \rangle - \Phi(x, y)\}.$$

Then we define the **dual problem** of $(\mathcal{P})$ by

$$(\mathcal{P}^*) \qquad \sup_{y^* \in Y^*} \{-\Phi^*(0, y^*)\}.$$

This is a dual problem with respect to the perturbation function Φ.

Remark: We denote the infimum for problem $(\mathcal{P})$ by $\inf(\mathcal{P})$ (analogously for $(\mathcal{P}^*)$ the supremum is termed $\sup(\mathcal{P}^*)$).

Between $(\mathcal{P})$ and $(\mathcal{P}^*)$ there is the relationship stated in the following assertion.

Proposition 6.1. (weak duality) *For problems* $(\mathcal{P})$ *and* $(\mathcal{P}^*)$ *we have*

$$-\infty \leq \sup(\mathcal{P}^*) \leq \inf(\mathcal{P}) \leq +\infty.$$

Remark: Convexity for $(\mathcal{P})$ is not assumed.

Proof. Let be $y^* \in Y^*$. It follows

$$\Phi^*(0, y^*) = \sup_{x \in X, y \in Y} \{\langle 0, x\rangle + \langle y^*, y\rangle - \Phi(x, y)\}.$$

Setting $y = 0$ we have $\Phi^*(0, y^*) \geq \langle y^*, 0\rangle - \Phi(x, 0) \quad \forall x \in X$, so $-\Phi^*(0, y^*) \leq \Phi(x, 0) \quad \forall x \in X \quad \forall y^* \in Y^*$. Thus

$$\sup_{y^* \in Y^*} \{-\Phi^*(0, y^*)\} \leq \inf_{x \in X} \Phi(x, 0) = \inf(\mathcal{P})$$

hence $\sup(\mathcal{P}^*) \leq \inf(\mathcal{P})$. $\square$

Of course, $(\mathcal{P}^*)$ may be rewritten as infimum problem and interpreted as primal problem. Then a **dual problem** $(\mathcal{P}^{**})$ to it and its relationship to the original $(\mathcal{P})$ are of interest, e.g. the **question**: are there conditions under which $(\mathcal{P}^{**})$ may be identified with $(\mathcal{P})$, i.e. $(\mathcal{P}^{**}) \cong (\mathcal{P})$?

Proceeding with this idea we get

$$(\mathcal{P}^*) \qquad \sup_{y^* \in Y^*} \{-\Phi^*(0, y^*)\} = -\inf_{y^* \in Y^*} \{\Phi^*(0, y^*)\}.$$

We consider the problem

$$(-\mathcal{P}^*) \qquad \inf_{y^* \in Y^*} \{\Phi^*(0, y^*)\}$$

and **use** $\Phi^*(x^*, y^*)$ **as perturbation function** with the **perturbation variable** x^*. Following the above construction of the dual problem we obtain as dual to $(-\mathcal{P}^*)$ the following problem

$$\sup_{x \in X} \{-\Phi^{**}(x, 0)\}.$$

Therefore we consider as **bidual** problem $(\mathcal{P}^{**})$ to $(\mathcal{P})$

$$(\mathcal{P}^{**}) \qquad -\sup_{x\in X}\{-\Phi^{**}(x,0)\} = \inf_{x\in X} \Phi^{**}(x,0).$$

As perturbation to $(\mathcal{P}^{**})$ one can in a natural way consider the problem

$$(\mathcal{P}_y^{**}) \qquad \inf_{x\in X} \Phi^{**}(x,y),$$

with the perturbation variable $y \in Y$. The corresponding dual problem reads as

$$(\mathcal{P}^{***}) \qquad \sup_{y^*\in Y^*} \{-\Phi^{***}(0,y^*)\}.$$

But because $\Phi^{***} \equiv \Phi^*$ follows $(\mathcal{P}^{***}) \equiv (\mathcal{P}^*)$. Thus we see the following.

Remark: When the perturbation function $\Phi(x,y)$ is proper, convex and weak-lower semicontinuous, then $\Phi^{**} = \Phi$ (cf. Theorem 5.2), therefore $\inf\limits_{x\in X}$ $\Phi(x,0) = \inf\limits_{x\in X} \Phi^{**}\ (x,0)$, i.e. $(\mathcal{P}) = (\mathcal{P}^{**})$: the bidual problem coincides with the primal problem, thus there is **symmetry**.

Remark: As we have remarked before, for X and Y Banach spaces and f proper convex lower semicontinuous function, follows f weak-lower semicontinuous. The reversed assertion holds in any case. If $\Phi(x,y)$ is proper, convex and lower semicontinuous we have $\Phi^{**} = \Phi$ and $(\mathcal{P}^{**}) = (\mathcal{P})$. Of course, in this case the original function f in $(\mathcal{P})$ has to be convex $(\Phi(x,0) = f(x))$ and lower semicontinuous.

6.2 Stability

We define the infimal value function

$$h(y) = \inf(\mathcal{P}_y) = \inf_{x\in X} \Phi(x,y).$$

It follows $h(0) = \inf(\mathcal{P}) = \inf\limits_{x\in X} \Phi(x,0) = \inf\limits_{x\in X} f(x)$.

Proposition 6.2. *Let $\Phi : X \times Y \to \overline{\mathbb{R}}$ be convex. Then $h : Y \to \mathbb{R}$ is convex.*

Proof. We show that Jensen's inequality holds. Consider a and b such that $h(y) < a < +\infty$ and $h(z) < b < +\infty$. Then there exist some points x and ξ such that $h(y) \leq \Phi(x, y) \leq a$ and $h(z) \leq \Phi(\xi, z) < b$. Because Φ is convex, it follows that for $\alpha \in (0,1)$ we have

$$\begin{array}{rcl} h(\alpha y + (1-\alpha)z) & = & \inf\limits_{w \in X} \Phi(w, \alpha y + (1-\alpha)z) \\ & \leq & \Phi(\alpha x + (1-\alpha)\xi, \alpha y + (1-\alpha)z) \\ & \leq & \alpha \Phi(x, y) + (1-\alpha)\Phi(\xi, z) \\ & \leq & \alpha a + (1-\alpha) b. \end{array}$$

Letting converge a towards $h(y)$ and b to $h(z)$ we obtain

$$h(\alpha y + (1-\alpha)z \leq \alpha h(y) + (1-\alpha)h(z).$$

If $h(y) = +\infty$ or $h(z) = +\infty$ the proof is trivial. $\square$

Proposition 6.3. *We have* $h^*(y^*) = \Phi^*(0, y^*) \quad \forall y^* \in Y^*$.

Proof. The following calculations go naturally

$$\begin{array}{rcl} h^*(y^*) & = & \sup\limits_{y \in Y} \{\langle y^*, y\rangle - h(y)\} \\ & = & \sup\limits_{y \in Y} \{\langle y^*, y\rangle - \inf\limits_{x \in X} \Phi(x, y)\} \\ & = & \sup\limits_{y \in Y} \sup\limits_{x \in X} \{\langle y^*, y\rangle - \Phi(x, y)\} = \Phi^*(0, y^*). \square \end{array}$$

Lemma 6.1. *Using the previous notations we have* $\sup(\mathcal{P}^*) = h^{**}(0)$.

Proof. Using the definitions we have

$$\begin{array}{rcl} \sup(\mathcal{P}^*) & = & \sup\limits_{y^* \in Y^*} \{-\Phi(0, y^*)\} \\ & = & \sup\limits_{y^* \in Y^*} \{-h^*(y^*)\} \\ & = & \sup\limits_{y^* \in Y^*} \{\langle 0, y^*\rangle - h^*(y^*)\} \\ & = & h^{**}(0). \square \end{array}$$

Remark: The well-known inequality $h^{**}(0) \leq h(0)$ means but $\sup(\mathcal{P}^*) \leq \inf(\mathcal{P})$.

Definition 6.2. *The problem*

$$(\mathcal{P}) \qquad \inf_{x \in X} f(x)$$

is said to be **stable** *if $h(0)$ is finite and $\partial h(0) \neq \emptyset$ (subdifferentiable).*

Proposition 6.4. *The set of solutions to problem $(\mathcal{P}^*)$ is identical to $\partial h^{**}(0)$.*

Proof. Let $z^* \in Y^*$ be a solution to $(\mathcal{P}^*)$. It follows that

$$-\Phi^*(0, z^*) \geq -\Phi^*(0, y^*) \; \forall y^* \in Y^*.$$

Because of Proposition 6.3, we have

$$\begin{aligned} -h^*(z^*) &\geq -h^*(y^*) \; \forall y^* \in Y^* \\ &\geq \sup_{y^* \in Y^*} \{\langle 0, y^* \rangle - h^*(y^*)\} = h^{**}(0). \end{aligned}$$

Therefore

$$h^{**}(0) + h^*(z^*) \leq 0 = \langle z^*, 0 \rangle.$$

On the other hand, from Young's inequality, we have

$$h^{**}(0) + h^*(z^*) \geq \langle z^*, 0 \rangle.$$

Hence $h^{**}(0) + h^*(z^*) = 0$. Theorem 5.3 implies $z^* \in \partial h^{**}(0)$. If $z^* \in \partial h^{**}(0)$ then all steps may be done in the reversed direction, yielding that z^* solves $(\mathcal{P}^*)$. □

Theorem 6.1. (strong duality) *The problem $(\mathcal{P})$ is stable if and only if the following two conditions are simultaneously fufilled.*

(i) $(\mathcal{P}^*)$ *has a solution.*

(ii) $\inf(\mathcal{P}) = \max(\mathcal{P}^*) < \infty$.

Remark: That means there is no duality gap. Therefore this situation is called **strong duality**.

Proof. *Necessity.* Let $(\mathcal{P})$ be stable. This means that $h(0)$ is finite and $\partial h(0) \neq \emptyset$. Let be $z^* \in \partial h(0)$. From Theorem 5.3 we have $h(0) + h^*(z^*) = \langle z^*, 0 \rangle = 0$, i.e. $h(0) = -h^*(z^*)$. Therefore

$$h^{**}(0) = \sup_{y^* \in Y^*} \{-h^*(y^*)\} \geq -h^*(z^*) = h(0).$$

On the other hand we have $h^{**}(0) \leq h(0)$, so $h^{**}(0) = h(0)$. We apply now Theorem 5.6 and we get $\partial h^{**}(0) = \partial h(0)$. This implies $z^* \in \partial h^{**}(0)$ and by Proposition 6.4 we obtain that z^* solves $(\mathcal{P}^*)$, i.e. (i) is true.

Moreover, Lemma 6.1 says

$$\sup(\mathcal{P}^*) = \max(\mathcal{P}^*) = h^{**}(0) = h(0) = \inf(\mathcal{P}) < \infty,$$

so also (ii) is true.

Sufficiency. Let (i) and (ii) be fulfilled and z^* a solution of $(\mathcal{P}^*)$. Proposition 6.4, again, implies $z^* \in \partial h^{**}(0)$ and (ii) with Lemma 6.1 tells us $h^{**}(0) = \max(\mathcal{P}^*) = \inf(\mathcal{P}) = h(0) < \infty$. Using Theorem 5.6 we obtain $\partial h(0) = \partial h^{**}(0)$, so $z^* \in \partial h(0)$ and thus $(\mathcal{P})$ is stable. □

Remark: As one may notice within the proof

$$\sup(\mathcal{P}^*) = h^{**}(0) = h(0) = \inf(\mathcal{P})$$

means strong duality (no duality gap). Another strong duality assertion is the following.

Theorem 6.2. (strong duality) *Let $\Phi : X \times Y \to \mathbb{R}$ be proper, convex and weak-lower semicontinuous. Then the following assertions are equivalent to each other.*

(i) Each of $(\mathcal{P})$ and $(\mathcal{P}^*)$ admits a solution and $\min(\mathcal{P}) = \max(\mathcal{P}^*) < \infty$.

(ii) $(\mathcal{P})$ and $(\mathcal{P}^*)$ are stable.

(iii) $(\mathcal{P})$ is stable and has a solution.

Remarks:

(a) For X Banach space a convex the lower semicontinuous functional f is weak-lower semicontinuous (Satz 2.5 in [1], p. 91). The converse implication always holds.

(b) Stability for $(\mathcal{P}^*)$ is defined analogously as for $(\mathcal{P})$.

Proof. "$(i) \Longleftrightarrow (ii)$" *Necessity.* From Theorem 6.1 (sufficiency) follows that $(\mathcal{P})$ is stable. We show the stability of $(\mathcal{P}^*)$.

Because Φ is proper, convex and weak-lower semicontinuous there is $(\mathcal{P}^{**}) = (\mathcal{P})$ since $\Phi^{**} = \Phi$. We apply Theorem 6.1 to $(\mathcal{P}^*)$ instead of $(\mathcal{P})$ taking into consideration that $(\mathcal{P}^{**}) = (\mathcal{P})$. Thus we obtain as Corollary to Theorem 6.1 the following result.

Corollary 6.1. (strong duality) *Let $\Phi : X \times Y \to \mathbb{R}$ be proper, convex and weak-lower semicontinuous. Then $(\mathcal{P}^*)$ is stable if and only if the following two conditions are fulfilled concomitantly.*

(i) Problem $(\mathcal{P})$ has a solution.

(ii) $\sup(\mathcal{P}^*) = \min(\mathcal{P}) < \infty$.

From Corollary 6.1 follows by Theorem 6.2(i) that $(\mathcal{P}^*)$ is stable.

Sufficiency. If $(\mathcal{P})$ and $(\mathcal{P}^*)$ are stable, then Theorem 6.1 and Corollary 6.1 imply that $(\mathcal{P})$ and $(\mathcal{P}^*)$ have solutions, moreover

$$\min(\mathcal{P}) = \max(\mathcal{P}^*) < \infty.$$

"$(ii) \Longleftrightarrow (iii)$" *Necessity.* If $(\mathcal{P})$ and $(\mathcal{P}^*)$ are stable, then with Corollary 6.1 follows the existence of a solution to $(\mathcal{P})$, i.e. (iii) is satisfied.

Sufficiency. Let $(\mathcal{P})$ be stable and suppose it has a solution. From Theorem 6.1 we have $\min(\mathcal{P}) = \max(\mathcal{P}^*) < \infty$. Because of Corollary 6.1 we obtain that $(\mathcal{P}^*)$ is stable. $\square$

To prove the stability directly by means of the definition is not so easy in many cases. Thus, so-called stability to more verifiable conditions are very useful. We give now such a stability criterion.

Theorem 6.3. (stability criterion) *Let Φ be convex, f proper and $-\infty < \inf_{x\in X} f(x) < +\infty$. Further, let us suppose that there exists an $x_0 \in X$ such that the functional $\Phi(x_0, \cdot) : Y \to \overline{\mathbb{R}}$ is finite and continuous at $0 \in Y$. Then $(\mathcal{P})$ is stable.*

Proof. The function $h(y) = \inf(\mathcal{P}_y)$ is convex and finite at $y = 0$. Since $\Phi(x_0, \cdot)$ is convex and continuous at $0 \in Y$, then there exists a neighborhood U of $0 \in Y$ such that $\Phi(x_0, y) \le c < +\infty$ for all $y \in U$. From here follows

$$h(y) = \inf_{x\in X} \Phi(x, y) \le \Phi(x_0, y) \le c,$$

i.e. $h(y)$ is convex and locally bounded above on a neighborhood of 0.

By Lemma 2.1 follows that h is continuous at $0 \in Y$. Theorem 4.4 assures $\partial h(0) \neq \emptyset$, moreover $h(0)$ is finite (by assumption) which implies that $(\mathcal{P})$ is stable by definition. $\square$

6.3 Optimality Conditions

Theorem 6.4. (optimality conditions)

(i) Assume $(\mathcal{P})$ and $(\mathcal{P}^*)$ have as solutions x^0 and y^{0*}, respectively, where strong duality is fulfilled

$$-\infty < f(x^0) = \inf(\mathcal{P}) = \sup(\mathcal{P}^*) = -\Phi^*(0, y^{0*}) < +\infty. \tag{6. 1}$$

Then the following optimality conditions hold

$$\Phi(x^0, 0) + \Phi^*(0, y^{0*}) = 0 \tag{6. 2}$$

and

$$(0, y^{0*}) \in \partial\Phi(x^0, 0). \tag{6. 3}$$

(ii) Let $x^0 \in X$ and $y^{0*} \in Y^*$ satisfy (6.2) or (6.3). Then x^0 is a solution to $(\mathcal{P})$ and y^{0*} is a solution to $(\mathcal{P}^*)$. Moreover, (6.1) holds.

Proof.

(i) At first we show that (6.2) and (6.3) are equivalent. From (6.1) we have

$$\begin{aligned}\Phi(x^0,0)+\Phi^*(0,y^{0*}) &= \langle 0,x^0\rangle_X+\langle y^{0*},0\rangle_Y\\ &= \langle (0,y^{0*}),(x^0,0)\rangle_{X\times Y}=0.\end{aligned}$$

But by Theorem 5.3 this is equivalent to $(0,y^{0*})\in\partial\Phi(x^0,0)$, i.e. (6.3) stands.

Because of Proposition 6.1 we have $\sup(\mathcal{P}^*)\le\inf(\mathcal{P})$. With

$$(\mathcal{P})\qquad \inf_{x\in X}\Phi(x,0)=\inf_{x\in X}f(x)$$

and

$$(\mathcal{P}^*)\qquad \sup_{y^*\in Y^*}\{-\Phi^*(0,y^*)\}$$

follows

$$-\Phi^*(0,y^*)\le\Phi(x,0)\ \forall x\in X\ y^*\in Y^*. \tag{6. 4}$$

From (6.1) holds with x^0 solution to $(\mathcal{P})$ and y^{0*} solution to $(\mathcal{P}^*)$ $-\Phi^*$ $(0,y^{0,*})=\Phi(x^0,0)$, i.e. (6.2) stands.

(ii) Let x^0,y^{0*} satisfy (6.2) (equivalent to (6.3)), i.e. $-\Phi^*(0,y^{0,*})=\Phi(x^0,0)$. Because of (6.4) we have that x^0 solves $(\mathcal{P})$, y^{0*} solves $(\mathcal{P}^*)$ and obviously strong duality holds, i.e. (6.1). □

7 Lagrangians and Saddle Points

In this chapter we will define the so-called Lagrange functional in a general matter and then show the relation between the conjugate duality (Fenchel-Rockafellar duality) and the well-known Lagrange duality.

Definition 7.1. *The function $L : X \times Y^* \to \overline{\mathbb{R}}$ defined by*

$$L(x, y^*) = -\sup_{y \in Y} \{\langle y^*, y\rangle - \Phi(x, y)\}$$

is said to be the **Lagrangean** *of the problem $(\mathcal{P})$ relative to the given perturbation.*

Obviously, one can write

$$L(x, y) = -\Phi_x^*(y^*),$$

where Φ_x denotes the function $y \to \Phi(x, y)$ for a fixed $x \in X$ and Φ_x^* denotes the conjugate function of Φ_x. From Proposition 5.1 we have that $\Phi_x^* : Y^* \to \overline{\mathbb{R}}$ is convex and weak-lower semicontinuous.

Lemma 7.1. *The function $L_x : y^* \to L(x, y^*)$, $x \in X$, is concave and weak-upper semicontinuous mapping Y^* into $\overline{\mathbb{R}}$. If Φ is convex, then for all $y^* \in Y^*$, the function $L_{y^*} : x \to L(x, y^*)$ is convex mapping X into $\overline{\mathbb{R}}$.*

Proof. Because $L_x(y^*) = L(x, y^*) = -\Phi_x^*(y^*)$, L_x is convex and weak-upper semicontinuous. We know $L(x, y^*) = \inf\limits_{y \in Y}\{\Phi(x, y)\langle y^*, y\rangle\}$. Let be $x_1, x_2 \in X$ and $\lambda \in (0, 1)$. If $L(x_1, y^*) = +\infty$ or $L(x_2, y^*) = +\infty$ it holds

$$L(\lambda x_1 + (1 - \lambda)x_2, y^*) \leq \lambda L(x_1, y^*) + (1 - \lambda)L(x_2, y^*).$$

Therefore assume $L(x_1, y^*) < \infty$ and $L(x_2, y^*) < \infty$ and choose a, b such that $a > L(x_1, y^*)$, $b > L(x_2, y^*)$ (a and b are fixed).

Because of the definition of L there exist some $y_1, y_2 \in Y$ such that

$$L(x_1, y^*) \leq \Phi(x_1, y_1) - \langle y^*, y_1\rangle \leq a$$

and

$$L(x_2, y^*) \leq \Phi(x_2, y_2) - \langle y^*, y_2\rangle \leq b.$$

In the same time we have

$$\begin{aligned}
L(\lambda x_1 + (1-\lambda)x_2, y^*) &= \inf_{y\in Y}\{\Phi(\lambda x_1 + (1-\lambda)x_2, y) - \langle y^*, y\rangle\} \\
&\leq \Phi(\lambda x_1 + (1-\lambda)x_2, \lambda y_1 + (1-\lambda)y_2) \\
&- \langle y^*, \lambda y_1 + (1-\lambda)y_2\rangle\rangle \\
&\leq \lambda[\Phi(x_1, y_1) - \langle y^*, y_1\rangle] + (1-\lambda)[\Phi(x_2, y_2) - \langle y^*, y_2\rangle] \\
&\leq \lambda a + (1-\lambda)b.
\end{aligned}$$

Setting a to tend to $L(x_1, y^*)$ and b towards $L(x_2, y^*)$, we get Jensen's inequality for L, so L is convex. $\square$

Now we can represent $(\mathcal{P})$ and $(\mathcal{P}^*)$ in terms of the Lagrangian L

$$\begin{aligned}
\Phi^*(x^*, y^*) &= \sup_{\substack{x\in X\\ y\in Y}}\{\langle x^*, x\rangle + \langle y^*, y\rangle - \Phi(x, y)\} \\
&= \sup_{x\in X}\{\langle x^*, x\rangle + \sup_{y\in Y}[\langle y^*, y\rangle - \Phi(x, y)]\} \\
&= \sup_{x\in X}\{\langle x^*, x\rangle - L(x, y^*)\},
\end{aligned}$$

hence (setting $x^* = 0$)

$$-\Phi^*(0, y^*) = \inf_{x\in X} L(x, y^*).$$

So, without assuming anything about Φ, we have

$$(\mathcal{P}^*) \qquad \sup_{y^*\in Y^*}\{-\Phi^*(0, y^*)\} = \sup_{y^*\in Y^*}\inf_{x\in X} L(x, y^*).$$

The assumption for Φ to be convex and weak-lower semicontinuous on $X \times Y$ yields that for all $x \in X$ the function $\Phi_x : y \to \Phi(x, y)$ is convex and weak-lower semicontinuous and therefore (cf. Theorem 5.2) $\Phi_x^{**} = \Phi_x$. Hence

$$\begin{aligned}
\Phi(x, y) &= \Phi_x^{**}(y) \\
&= \sup_{y^*\in Y^*}\{\langle y^*, y\rangle - \Phi_x^*(y^*)\} \\
&= \sup_{y^*\in Y^*}\{\langle y^*, y\rangle + L(x, y^*)\},
\end{aligned}$$

and setting $y = 0$ we obtain

$$\Phi(x, 0) = \sup_{y^*\in Y^*} L(x, y^*).$$

Consequently (if $\Phi_x : y \mapsto \Phi(x,y)$ is convex and weak-lower semicontinuous) we can derive

$$(\mathcal{P}) \qquad \inf_{x\in X} f(x) = \inf_{x\in X} \Phi(x,0) = \inf_{x\in X} \sup_{y^*\in Y^*} L(x,y^*).$$

Remark: Introducing the Lagrangian we see that $(\mathcal{P})$ and $(\mathcal{P}^*)$ are related to min-max problems and the weak duality relation

$$\sup(\mathcal{P}^*) \leq \inf(\mathcal{P})$$

is actually the inequality

$$\sup_{y^*\in Y^*} \inf_{x\in X} L(x,y^*) \leq \inf_{x\in X} \sup_{y^*\in Y^*} L(x,y^*)$$

well known in game theory.

Definition 7.2. $(\bar{x},\bar{y}^*) \in X \times Y^*$ *is said to be a* **saddle point** *of* L *if*

$$L(\bar{x},y^*) \leq L(\bar{x},\bar{y}^*) \leq L(x,\bar{y}^*) \ \forall x \in X \ \forall y^* \in Y^*.$$

Theorem 7.1. (saddle point theorem) *Let* $\Phi : X \times Y \to \mathbb{R}$ *be convex and weak-lower semicontinuous. Then the following conditions are equivalent to each other.*

(i) $(\bar{x},\bar{y}^*)$ *is a saddle point of* L.

(ii) $\bar{x}$ *solves* $(\mathcal{P})$, $\bar{y}^*$ *solves* $(\mathcal{P}^*)$ *and* $\min(\mathcal{P}) = \max(\mathcal{P}^*)$ *(i.e. strong duality).*

Proof.
"*(i)* $\Rightarrow$ *(ii)*" We have

$$L(\bar{x},\bar{y}^*) = \inf_{x\in X} L(x,\bar{y}^*) = -\Phi^*(0,\bar{y}^*)$$

and

$$L(\bar{x},\bar{y}^*) = \sup_{y^*\in Y^*} L(\bar{x},y^*) = \Phi(\bar{x},0).$$

Therefore

$$\Phi(\bar{x},0) + \Phi^*(0,\bar{y}^*) = 0.$$

Thus Theorem 6.4 demonstrates that $\bar{x}$ solves $(\mathcal{P})$, $\bar{y}^*$ solves $(\mathcal{P}^*)$ and strong duality holds.
"$(ii) \Rightarrow (i)$" We know that

$$-\Phi^*(0, \bar{y}^*) = \inf_{x \in X} L(x, \bar{y}^*) \leq L(\bar{x}, \bar{y}^*)$$

and

$$\Phi(\bar{x}, 0) = \sup_{y^* \in Y^*} L(\bar{x}, y^*) \geq L(\bar{x}, \bar{y}^*).$$

But since strong duality holds, $\bar{x}$ solves $(\mathcal{P})$, $\bar{y}^*$ solves $(\mathcal{P}^*)$ and we have

$$\Phi(\bar{x}, 0) + \Phi^*(0, \bar{y}^*) = 0.$$

Therefore we can write

$$\begin{aligned} L(\bar{x}, \bar{y}^*) &\leq \sup_{y^* \in Y^*} L(\bar{x}, y^*) = \Phi(\bar{x}, 0) \\ &= -\Phi^*(0, \bar{y}^*) = \inf_{x \in X} L(x, \bar{y}^*) \\ &\leq L(\bar{x}, \bar{y}^*), \end{aligned}$$

that implies

$$\sup_{y^* \in Y^*} L(\bar{x}, y^*) = \inf_{x \in X} L(x, \bar{y}^*) = L(\bar{x}, \bar{y}^*).$$

So we can conclude

$$L(\bar{x}, y^*) \leq L(\bar{x}, \bar{y}^*) \leq L(x, \bar{y}^*) \ \forall y^* \in Y^* \ \forall x \in X. \square$$

Theorem 7.2. (saddle point theorem) *Let $\Phi : X \times Y \to \bar{\mathbb{R}}$ ($\Phi \neq \pm\infty$) be convex and weak-lower semicontinuous and $(\mathcal{P})$ is assumed to be stable. Then $\bar{x} \in X$ is a solution to $(\mathcal{P})$ if and only if there exists an $\bar{y}^* \in Y^*$ such that $(\bar{x}, \bar{y}^*)$ is saddle point of L.*

Proof. *Necessity.* Let $\bar{x} \in X$ be a solution to $(\mathcal{P})$. Because $(\mathcal{P})$ is stable $(\mathcal{P}^*)$ has a solution $\bar{y}^*$ and $\min(\mathcal{P}) = \max(\mathcal{P}^*)$. Theorem 7.1 implies that $(\bar{x}, \bar{y}^*)$ is saddle point of L.
Sufficiency. This is a direct consequence of Theorem 7.1 $(i) \Rightarrow (ii)$. $\square$

8 Important Special Cases of Dual Optimization Problems

8.1 Case I

Given X, Y normed spaces, $A \in L(X, Y)$ (linear continuous operator mapping from X into Y), $A^* \in L(Y^*, X^*)$ transpose, f to be minimized has the form $f(x) := q(x, Ax)$ where $q : x \times Y \to \overline{\mathbb{R}}$ and our primal problem is

$$(\mathcal{P}) \qquad \inf_{x \in X} q(x, Ax).$$

To establish a dual problem we have to introduce a perturbation function $\Phi(x, y)$.

Thus we propose

$$\Phi(x, y) := q(x, Ax - y).$$

Now we deduce the dual problem according to the approach in Chapter 6. We need

$$\begin{aligned} \Phi^*(0, y^*) &= \sup_{\substack{x \in X \\ y \in Y}} \{\langle y^*, y \rangle - q(x, Ax - y)\} \\ &= \sup_{x \in X} \sup_{y \in Y} \{\langle y^*, y \rangle - q(x, Ax - y)\}. \end{aligned}$$

Introducing a new variable p instead of y setting $p := Ax - y$ (for a fixed x) we obtain

$$\begin{aligned} \Phi^*(0, y^*) &= \sup_{x \in X} \sup_{p \in Y} \{\langle y^*, Ax \rangle - \langle y^*, p \rangle - q(x, p)\} \\ &= \sup_{\substack{x \in X \\ p \in Y}} \{\langle A^* y^*, x \rangle + \langle -y^*, p \rangle - q(x, p)\} \\ &= q^*(A^* y^*, -y^*) \end{aligned}$$

and the dual problem is

$$(\mathcal{P}^*) \qquad \sup_{y^* \in Y^*} \{-q^*(A^* y^*, -y^*)\}.$$

Remarks:

(i) q convex implies Φ convex.

(ii) q weak-lower semicontinuous with $q \neq \pm\infty$ implies Φ weak-lower semicontinuous and $\Phi \neq \pm\infty$.

Theorem 8.1. (strong duality) *Assume q convex, $-\infty < \inf(\mathcal{P}) < +\infty$. Moreover, let there exist an $x_0 \in X$ such that $q(x_0, Ax_0) < +\infty$ and let the function $y \to q(x_0, y)$ be continuous at Ax_0. Then $(\mathcal{P})$ is stable, $(\mathcal{P}^*)$ has a solution $\bar{y}^*$ and $\inf(\mathcal{P}) = \max(\mathcal{P}^*)$.*

Proof. The assumption says that $y \to \Phi(x_0, y) = q(x_0, Ax_0 - y)$ is finite and continuous at $0 \in Y$. That means that Theorem 6.3 renders stability of $(\mathcal{P})$ and Theorem 6.1 implies the assertion. □

Theorem 8.2. (optimality conditions) *The following conditions are equivalent to each other.*

(i) $\bar{x}$ solves $(\mathcal{P})$, $\bar{y}^*$ solves $(\mathcal{P}^*)$ and $\min(\mathcal{P}) = \max(\mathcal{P}^*)$.

(ii) $\bar{x} \in X$ and $\bar{y}^* \in Y^*$ satisfy the optimality conditions (extremality relation) $q(\bar{x}, A\bar{y}) + q^*(A^*\bar{y}^*, -\bar{y}^*) = 0$ which is equivalent to $(A^*\bar{y}^*, -\bar{y}^*) \in \partial q(\bar{x}, A\bar{x})$.

Proof. Applying Theorem 6.4 yields that $(0, \bar{y}^*) \in \partial\Phi(\bar{x}, 0)$ is equivalent to

$$\begin{aligned} 0 &= \Phi(\bar{x}, 0) + \Phi^*(0, \bar{y}^*) \\ &= q(\bar{x}), A\bar{x}) + q^*(A^*\bar{y}^*, -\bar{y}^*) \\ &= \langle(\bar{x}), A\bar{x}), (A^*\bar{y}^*, -\bar{y}^*)\rangle, \end{aligned}$$

so Theorem 5.3 leads to $(A^*\bar{y}^*, -\bar{y}^*) \in \partial q(\bar{x}, A\bar{x})$. □

Remark: We consider the further specialization

$$q(x, Ax) := f(x) + g(Ax)$$

(q is decomposed). The primal problem is in this case

$$(\mathcal{P}) \qquad \inf_{x \in X}\{f(x) + g(Ax)\}.$$

The conjugate of the primal objective function is

$$\begin{aligned}
q^*(x^*, y^*) &= \sup_{\substack{x\in X,\\ y\in Y}} \{\langle x^*, x\rangle + \langle y^*, y\rangle - q(x,y)\} \\
&= \sup_{x\in X}\{\langle x^*, x\rangle - f(x)\} + \sup_{y\in Y}\{\langle y^*, y\rangle - g(y)\} \\
&= f^*(x^*) + g^*(y^*).
\end{aligned}$$

So the dual problem reads as

$$(\mathcal{P}^*) \qquad \sup_{y^*\in Y^*} \{-q^*(A^*y^*, -y^*)\}$$

that is in this case

$$(\mathcal{P}^*) \qquad \sup_{y^*\in Y^*} \{-f^*(A^*y^*) - g^*(-y^*)\}.$$

Remarks: We note that if

(i) f and g are convex then q and hence Φ is convex.

(ii) f and g are convex and weak-lower semicontinuous and $f, g \neq \pm\infty$ then q and Φ are convex and weak-lower semicontinuous and $q,\ \Phi \neq \pm\infty$.

The stability criterion from Theorem 8.1 can be written in the following way.

There exists $x_0 \in X$ such that $f(x_0) < +\infty$, $g(Ax_0) < +\infty$, g being continuous at Ax_0, f and g convex and $\inf(\mathcal{P})$ finite. Then can be applied Theorem 8.1 (strong duality) and Theorem 8.2 (optimality conditions).

The optimality conditions of Theorem 8.2 can be decomposed

$$\begin{aligned}
0 &= q(\bar{x}), A\bar{x}) + q^*(A^*\bar{y}^*, -\bar{y}^*) \\
&= f(\bar{x}) + g(A\bar{x}) + f^*(A^*\bar{y}^*) + g^*(-\bar{y}^*) \\
&= f(\bar{x}) + f^*(A^*\bar{y}^*) + g(A\bar{x}) + g^*(-\bar{y}^*) \\
&= [f(\bar{x}) + f^*(A^*\bar{y}^*) - \langle A^*\bar{y}^*, \bar{x}\rangle] + [g(A\bar{x}) + g^*(-\bar{y}^*) - \langle -\bar{y}^*, A\bar{x}\rangle].
\end{aligned}$$

But because of Young's inequality both expression in square brackets are greater or equal to zero and the left hand side of the equation above is equal to zero, we have that $f(\bar{x})+f^*(A^*\bar{y}^*)-\langle A^*\bar{y}^*, \bar{x}\rangle = 0$ and $g(A\bar{x})+g^*(-\bar{y}^*)-\langle -\bar{y}^*, A\bar{x}\rangle = 0$.

Corollary 8.1. (strong duality) *Assume f and g are convex and $-\infty < \inf(\mathcal{P}) < +\infty$. Let there exist an $x_0 \in X$ such that $f(x_0) < +\infty$, $g(Ax_0) < +\infty$ and assume g continuous at Ax_0. Then $(\mathcal{P})$ is stable, $(\mathcal{P}^*)$ has a solution $\bar{y}^*$ and $\inf(\mathcal{P}) = \max(\mathcal{P}^*)$.*

Corollary 8.2. (optimality conditions) *The following conditions are equivalent to each other.*

(i) ($\bar{x}$ *solves* $(\mathcal{P})$, $\bar{y}^*$) *solves* $(\mathcal{P}^*)$ *and* $\min(\mathcal{P}) = \max(\mathcal{P}^*)$.

(ii) $\bar{x} \in X$ *and* $\bar{y}^* \in Y^*$ *satisfy the optimality conditions (extremality relations)*

$$f(\bar{x}) + f^*(A^*\bar{y}^*) = \langle A^*\bar{y}^*, \bar{x}\rangle$$

and

$$g(A\bar{x}) + g^*(-\bar{y}^*) = \langle -\bar{y}^*, A\bar{x}\rangle,$$

which are equivalent to $A^*\bar{y}^* \in \partial f(\bar{x})$ *and* $-\bar{y}^* \in \partial g(A\bar{x})$.

Exercise 8.1. Consider

$$(\mathcal{P}) \qquad \inf_{x \in X} \left\{ f(x) + \sum_{i=1}^{m} g_i(A_i x) \right\}.$$

Determine its dual problem and calculate the corresponding optimality conditions.

8.2 Case II

Let X be a linear space.

Definition 8.1. *A set $C \subseteq X$ is said to be a* **cone** *if $\lambda C \subseteq C \ \ \forall \lambda \geq 0$ (cone with vertex 0).*

Let C be convex, additionally. Define a partial ordering " $\leq$ " (or " $\geq$ ") by setting

$$x_1 \leq x_2 \Longleftrightarrow x_2 - x_1 \in C.$$

Obviously $x \leq x \ \ \forall x \in X$ and if $x_1 \leq x_2$ and $x_2 \leq x_3$ follows $x_1 \leq x_3$.

The partial ordering is compatible with the structure of the vector space X in the sense

- if $x \geq 0$ then $\lambda x \geq 0 \ \ \forall \lambda \geq 0$,
- if $x_1 \geq x_2$ then $x_1 + x \geq x_2 + x \ \ \forall x \in X$.

The partial ordering introduced above induces also the following sets

(i) $C = \{x \in X : x \geq 0\}$ is the **set of positive elements**

and

(ii) $-C = \{x \in X : x \leq 0\}$ is the **set of negative elements**.

Now let X be a linear normed space and X^* its topological dual space.

Definition 8.2. *Let $C \subseteq X$ be a cone. The set*

$$C^* = \{x^* \in X^* : \langle x^*, x\rangle \geq 0, \ \forall x \in C\}$$

is called the **dual cone** *to the cone C.*

In the dual cone the partial ordering is $x_1^* \leq x_2^* \Longleftrightarrow x_2^* - x_1^* \in C^*$. So one may say that C^* is the cone of positive elements in X^*.

The Primal Problem

Let X and Y be linear normed spaces, $D \subset X$ closed, convex, $D \neq \emptyset$, $f : D \to \mathbb{R}$ convex weak-lower semicontinuous, C closed convex cone in Y defining a partial ordering relation " $\leq$ ", $B : D \to Y$ a mapping (Possibly nonlinear) such that B is convex with respect to the relation " $\leq$ ", i.e.

$$B(\lambda x_1 + (1-\lambda)x_2) \leq \lambda B(x_1) + (1-\lambda)B(x_2) \ \forall x_1, x_2 \in D \ \forall \lambda \in (0,1).$$

Let further for each $y^* \in Y^*$, $y^* \geq 0$ the mapping $x \to \langle y^*, Bx\rangle$ of D into $\mathbb{R}$ be lower semicontinuous. Finally, let $\{x \in D : Bx \leq 0\} \neq \emptyset$. This set is convex.

Consider $x_1, x_2 \in D$ such that $Bx_1 \leq 0$ and $Bx_2 \leq 0$. We have for $\lambda \in (0,1)$ $\lambda Bx_1 \leq 0$ and $(1-\lambda)Bx_2 \leq 0$ thus $\lambda x_1 + (1-\lambda)x_2 \in D$. Then

$$B(\lambda x_1 + (1-\lambda)x_2) \leq \lambda Bx_1 + (1-\lambda)Bx_2 \leq 0.$$

Consider the optimization problem

$$(\mathcal{P}) \qquad \inf_{\substack{x\in D,\\ Bx\leq 0}} f(x).$$

We can rewrite this as

$$\inf_{x\in X} F(x),$$

setting

$$F : X \to \overline{\mathbb{R}},\ F(x) = \begin{cases} f(x), & \text{if } x \in D \text{ and } Bx \leq 0, \\ +\infty, & \text{otherwise.} \end{cases}$$

Define a perturbation function

$$\Phi(x,y) = \begin{cases} f(x), & \text{if } x \in D \text{ and } Bx \leq y, \\ +\infty, & \text{otherwise,} \end{cases}$$

that proves to be proper. We observe that Φ can be written as

$$\Phi : X \to \overline{\mathbb{R}},\ \Phi(x,y) = \tilde{f}(x) + \chi_{E_y}(x),$$

where

$$\tilde{f} : X \to \mathbb{R},\ \tilde{f}(x) = \begin{cases} f(x), & \text{if } x \in D, \\ +\infty, & \text{otherwise,} \end{cases}$$

and χ_{E_y} is the indicator functional of the set $E_y = \{x \in X : x \in D, Bx \leq y\}$.

To specify the properties of Φ we consider the following additional results.

Lemma 8.1.

(i) E_y is closed and convex in X $\forall y \in Y$ (and also weakly (sequentially) closed because we have that in topologocal Hausdorff space X if $E \subset X$ is convex and closed then E is weakly (sequentially) closed).

(ii) $E = \{(x,y) \in X \times Y : x \in D, Bx \leq y\}$ is closed and convex in $X \times Y$.

Proof.

(i) Let be $x \in E_y$. Then $Bx - y \leq 0$ and using Proposition 8.1 we obtain $\langle y^*, Bx - y\rangle \leq 0$ $\forall y^* \in C^*$, i.e. $\forall y^* \geq 0$.

The function $x \mapsto \langle y^*, Bx - y\rangle$, $x \in D$ is convex (since B is convex) for $y^* \geq 0$ fixed. Because for each $\lambda \in (0,1)$

$$\lambda B(x_1) + (1-\lambda)B(x_2) - B(\lambda x_1 + (1-\lambda)B(x_2)) \in C,$$

follows for $y^* \in C^*(y^* \geq 0)$

$$\langle y^*, \lambda b(x_1) + (1-\lambda)B(x_2) - B(\lambda x_1 + (1-\lambda)B(x_2))\rangle \geq 0,$$

equivalent to

$$\langle y^*, B(\lambda x_1 + (1-\lambda)B(x_2)) - y\rangle \leq \lambda\langle y^*, B(x_1) - y\rangle + (1-\lambda)\langle y^*, B(x_2) - y\rangle,$$

which proves the convexity of the function.

Moreover, by assumption, this function $x \mapsto \langle y^*, Bx - y\rangle$, $x \in D$ is weak-lower semicontinuous (even lower semicontinuous if X is a Banach space) and therefore the set $\{x \in D : \langle y^*, Bx - y\rangle \leq 0\}$ is closed (cf. Theorem 3.1) and convex for fixed $y^* \geq 0$. Then same holds for the intersection of all such sets which correspond to all $y^* \geq 0$.

(ii) Note that the function $(x, y) \mapsto \langle Bx - y, y^*\rangle$ is lower semicontinuous and convex on $D \times Y \to \mathbb{R}$ $\forall y^* \geq 0$ fixed. Using analogous arguments as for (i) yields the statement. □

Lemma 8.2. *The function $\Phi : X \times Y \to \overline{\mathbb{R}}$ is convex and lower semicontinuous if f is lower semicontinuous (weak-lower semicontinuous if f is weak-lower semicontinuous) and also $\Phi \neq \pm\infty$.*

Proof. Write Φ in the form $\Phi(x, y) = \tilde{f}(x) + \chi_E(x, y)$, where $\tilde{f}$ is convex and lower semicontinuous (weak-lower semicontinuous if f is weak-lower semicontinuous).

Moreover, $\chi_E(x, y)$ is convex and lower semicontinuous since E is a closed convex set in $X \times Y$ then $\{(x, y) : \chi_E(x, y) \leq k\}$ is closed for all constants $k \in \mathbb{R}$.

The dual problem

We compute for $y^* \in Y^*$

$$\begin{aligned}\Phi^*(0,y^*) &= \sup_{\substack{x\in X,\\ y\in Y}}\{\langle y^*,y\rangle - \Phi(x,y)\}\\ &= \sup_{\substack{x\in D, y\in Y,\\ Bx\le y}}\{\langle y^*,y\rangle - f(x)\}.\end{aligned}$$

Setting $p := y - Bx$ with $p \geq 0$ we obtain

$$\begin{aligned}\Phi^*(0,y^*) &= \sup_{x\in D}\sup_{\substack{p\in Y,\\ p\ge 0}}\{\langle y^*,Bx\rangle + \langle y^*,p\rangle - f(x)\}\\ &= \sup_{\substack{p\in Y,\\ p\ge 0}}\langle y^*,p\rangle + \sup_{x\in D}\{\langle y^*,Bx\rangle - f(x)\}.\end{aligned}$$

Because

$$\sup_{p\in Y, p\ge 0}\langle y^*,p\rangle = \left\{\begin{array}{ll} 0, & \text{if} \quad y^*\le 0,\\ +\infty, & \text{otherwise,}\end{array}\right\} = \chi_C^*(-y^*)$$

we can write

$$-\Phi^*(0,y^*) = -\chi_C^*(-y^*) + \inf_{x\in D}\{\langle -y^*,Bx\rangle + f(x)\}.$$

Therefore the **dual problem** is

$$(\mathcal{P}^*) \qquad \sup_{y^*\le 0}\inf_{x\in D}\{\langle -y^*,Bx\rangle + f(x)\},$$

equivalent to

$$(\mathcal{P}^*) \qquad \sup_{y^*\ge 0}\inf_{x\in D}\{\langle y^*,Bx\rangle + f(x)\},$$

but we prefer to use further the first formulation.

Theorem 8.3. *In addition to the above assumptions (concerning D, f and B (in particular convexity, closedness and $\{x \in D : Bx \leq 0\} \neq \emptyset$)), assume that $\inf(\mathcal{P})$ is finite and moreover let the following regularity condition (Slater condition) be satisfied, i.e. there exists $x_0 \in D$ such that $Bx_0 < 0$, i.e. $-Bx_0 \in intC$. Then $(\mathcal{P})$ is stable.*

Proof. From the regularity condition follows the existence of a neighborhood $U(0)$ of 0 in Y such that $-Bx_0 + y \in C \quad \forall y \in U(0)$, followed by $-Bx_0 + y \geq 0$ and even $Bx_0 \leq y \quad \forall y \in U(0)$.

Hence $\Phi(x_0, y) = f(x_0)\ \forall y \in U(0)$ and the functional $y \mapsto \Phi(x_0, y)$ is finite and continuous at $0 \in Y$. Therefore the assumption of Theorem 6.3 (stability criterion) is fulfilled and $(\mathcal{P})$ is stable. $\square$

Now we can apply Theorem 6.1 (strong duality).

Theorem 8.4. (strong duality) *Under the assumptions of Theorem 8.3 $(\mathcal{P}^*)$ has a solution and strong duality is fulfilled*

$$\inf(\mathcal{P}) = \inf_{\substack{x \in D,\\ Bx \le 0}} = \max_{y^* \le 0} \inf_{x \in D} \{f(x) + \langle -y^*, Bx \rangle\}.$$

Remarks:

(i) We may also apply the other strong duality assertions, like Theorem 6.2 and Corollary 6.1 where we need weak-lower semicontinuity of Φ which follows from that of f.

(ii) Moreover, by Theorem 6.4 (optimality conditions) one can infer optimality conditions for our present problem $(\mathcal{P})$ and $(\mathcal{P}^*)$.

If x^0 solves $(\mathcal{P})$ and y^{0*} solves $(\mathcal{P}^*)$ then $\Phi(x^*, 0) + \Phi^*(0, y^{0*}) = 0$ (optimality condition). In our case we have

$$0 = f(x^0) + \chi_C^*(-y^{0*}) - \inf_{x \in D} \big\{\langle -y^{0*}, Bx \rangle + f(x)\big\}.$$

Because $y^{0*} \le 0$ follows

$$\begin{aligned} f(x^0) &= \inf_{x \in D} \{\langle -y^{0*}, Bx \rangle + f(x)\} \\ &\le \langle -y^{0*}, Bx^0 \rangle + f(x^0) \end{aligned}$$

and we obtain

$$\langle y^{0*}, Bx^0 \rangle = 0$$

as an optimality condition.

Computation of the Lagrangian

By the definition of the Lagrangian we have

$$
\begin{aligned}
-L(x, y^*) &= \sup_{y\in Y}\{\langle y^*, y\rangle - \Phi(x, y)\} \\
&= \begin{cases} \sup\limits_{y\in Y, Bx\le y} \{\langle y^*, y\rangle - f(x)\}, & \text{if} \quad x \in D, \\ -\infty, & \text{otherwise.} \end{cases}
\end{aligned}
$$

Setting $p = y - Bx$ with $p \geq 0$ for $Bx \leq y$ gives us

$$
\begin{aligned}
-L(x, y^*) &= \sup_{p\in Y, p\ge 0} \{\langle y^*, p\rangle + \langle y^*, Bx\rangle - f(x)\} \\
&= \sup_{p\in Y, p\ge 0} \{\langle y^*, p\rangle\} + \langle y^*, Bx\rangle - f(x).
\end{aligned}
$$

The case $x \notin D$ can be included by replacing $f(x)$ by

$$
\tilde{f} : X \to \overline{\mathbb{R}},\ \tilde{f} = \begin{cases} f(x), & \text{if } x \in D, \\ +\infty, & \text{otherwise,} \end{cases}
$$

With

$$
\sup_{p\in Y, p\ge 0} \langle y^*, p\rangle = \left\{ \begin{array}{ll} 0, & \text{if } y^* \le 0, \\ +\infty, & \text{otherwise,} \end{array} \right\} = \chi_C^*(-y^*)
$$

the Lagrangian is

$$
L(x, y^*) = \tilde{f}(x) - \langle y^*, Bx\rangle - \chi_C^*(-y^*).
$$

From the definition of a saddle point we see that $(\bar{x}, \bar{y}^*)$ is a saddle point of $L(x, y^*)$ if and only if $\bar{x} \in D$ and $\bar{y}^* \leq 0$. For all $x \in D$ and all $y^* \leq 0$ we have

$$
\begin{aligned}
L(x, y^*) &= f(\bar{x}) - \langle y^*, B\bar{x}\rangle \\
&\le f(\bar{x}) - \langle \bar{y}^*, B\bar{x}\rangle = L(\bar{x}, \bar{y}^*) \\
&\le f(x) - \langle \bar{y}^*, Bx\rangle = L(x, \bar{y}^*).
\end{aligned}
$$

Remark: If $\bar{y}^* \leq 0$, $x \in D$, then also for x not in D we get $L(x, \bar{y}^*) = +\infty$ and for $y^* > 0$ we have $L(\bar{x}, y^*) = -\infty$. Thus the saddle point property for $(\bar{x}, \bar{y}^*)$ is also fulfilled.

Therefore we obtain from Theorem 7.2 the saddle point assertion (we replace y^* by $-y^*$ and we have $y^* \geq 0$ as usual in convex programming).

Theorem 8.4. *Consider X and Y normed linear spaces and*

- $f : D \to \mathbb{R}$ *is a convex and weak-lower semicontinuous functional, where* D *is a non-empty closed convex subset of* X*,*

- C *is a closed convex cone in* Y *(defining "$\leq$"),*

- $B : D \to Y$ *is convex with respect to "$\leq$",*

- *for each* $y^* \in Y^*$*,* $y^* \geq 0$ *the mapping* $x \mapsto \langle y^*, Bx \rangle$ *of* D *into* $\mathbb{R}$ *is lower semicontinuous,*

- $\{x \in D : Bx \leq 0\} \neq \emptyset$*,*

- *regularity condition from Theorem 8.3 is fulfilled,*

- $\inf\limits_{\substack{x \in D, \\ Bx \leq 0}} f(x)$ *is finite.*

(i) Then $\bar{x}$ *is a solution to* $(\mathcal{P})$ *if and only if there exists* $\bar{y}^* \in Y^*$*,* $\bar{y}^* \geq 0$ *such that* $(\bar{x}, \bar{y}^*)$ *is a saddle point for the Lagrangian* $L(xy^*) = \tilde{f}(x) + \langle y^*, Bx \rangle + \chi_C^*(y^*)$*, i.e. for all* $x \in D$ *and for all* $y^* \geq 0$ *one has*

$$f(\bar{x}) + \langle y^*, B\bar{x} \rangle \leq f(\bar{x}) + \langle \bar{y}^*, B\bar{x} \rangle \leq f(x) + \langle \bar{y}^*, Bx \rangle.$$

In this case holds

$$\langle \bar{y}^*, B\bar{x} \rangle = 0.$$

(ii) $\bar{y}^*$ *solves*

$$(\mathcal{P}^*) \qquad \max_{y^* \geq 0} \inf_{x \in D} \{f(x) + \langle y^*, Bx \rangle\}.$$

Proof. From Theorem 8.3 we know that $(\mathcal{P})$ is stable. Thus Theorem 7.2 proves the claimed saddle point property. Setting $y^* = 0$ in the saddle point inequality follows

$$f(\bar{x}) + \langle y^*, B\bar{x} \rangle = f(\bar{x}) \leq f(\bar{x}) + \langle \bar{y}^*, B\bar{x} \rangle \leq f(x) + \langle \bar{y}^*, Bx \rangle,$$

that implies

$$\langle \bar{y}^*, B\bar{x} \rangle = 0.$$

The fact that $\bar{y}^*$ solves $(\mathcal{P}^*)$ follows immediately from Theorem 7.1 (saddle point theorem). $\square$

Finite dimensional convex programming

Set $X = \mathbb{R}^n = X^*$, $Y = \mathbb{R}^m = Y^*$, $x = (x_1, ..., x_n)$, $y = (y_1, ..., y_m)$, $C = \{y = (y_1, ..., y_m) : y_i \geq 0, i = 1, ..., m\}$, $D \subseteq \mathbb{R}^n$ convex and closed and $By = (B_1 y, ..., B_m y)$ with $B_i : D \to \mathbb{R}$ convex and lower semicontinuous, $i = 1, ..., m$. The problem $(\mathcal{P})$ is

$$(\mathcal{P}) \qquad \inf_{x \in D, B_i(x) \leq 0, i=1,...,m} f(x),$$

where f is a lower semicontinuous convex function of D into $\mathbb{R}$. Let be given the regularity condition: there exists $x_0 \in D$ such that $B_1 x_0 < 0$, $i = 1, ..., m$. For $y_i \geq 0$, $i = 1, ..., m$, and $x \in D$ the Lagrangian is $L(x, y) = f(x) + \sum_{i=1}^{n} y_i B_i x$. Theorem 8.4 turns then into the following statement.

Theorem 8.5. (Kuhn-Tucker). *Let be fulfilled the above introduced hypotheses. Then $\bar{x} \in D$ is a solution to $(\mathcal{P})$ if and only if there exists $\bar{y} \in \mathbb{R}^m$, $\bar{y} \geq 0$ such that*

$$L(\bar{x}, y) \leq L(\bar{x}, \bar{y}) \leq L(x, \bar{y}) \ \forall x \in D \ \forall y_i \geq 0, \ i = 1, ..., m.$$

Then holds also

$$\langle \bar{y}, B\bar{x} \rangle = \sum_{i=1}^{m} \bar{y}_i B_i \bar{x} = 0.$$

For all i, $1 \leq i \leq m$, applies either $B_i \bar{x} < 0$ and $y_i = 0$ or $B_i \bar{x} = 0$ and $y_i \geq 0$.

9 Anexa

This section is dedicated to solutions for some exercises given in the lecture.

Examples 1.1.(i) Let X be a real normed space. Then $f : X \to \mathbb{R}, f(x) = \|x\|^n$, $n > 1$ is a convex function.

Solution. Let be $x, y \in X$ and $\lambda \in (0,1)$. From the properties of the norms we have

$$f(\lambda x + (1-\lambda)y) = \|\lambda x + (1-\lambda)y\|^n \leq \big(\lambda\|x\| + (1-\lambda)\|y\|\big)^n. \qquad (9.\ 1)$$

Consider now the function $g : (0, +\infty) \to \mathbb{R}$, $g(t) = t^n$. Because $g''(t) \geq 0\ \forall t > 0$, g is convex, so for all $a, b > 0$ and $\lambda \in (0,1)$ one has

$$\big(\lambda a + (1-\lambda)b\big)^n \leq \lambda a^n + (1-\lambda)b^n. \qquad (9.\ 2)$$

If $a = 0$ or $b = 0$ or both a and b are equal to 0 the inequality (9.2) is also true. From (9.1) and (9.2) we may conclude, taking $a = \|y\|$ and $b = \|y\|$,

$$\begin{aligned} f(\lambda x + (1-\lambda)y) \leq \big(\lambda\|x\| + (1-\lambda)\|y\|\big)^n &\leq \lambda\|x\|^n + (1-\lambda)\|y\|^n \\ &= \lambda f(x) + (1-\lambda)f(y), \end{aligned}$$

so f is convex. □

Exercise 1.1.

(i) Let X be a reflexive real Banach space and X^* its dual. Then the function $f : X \to \mathbb{R}, f(x) = \langle Bx, x\rangle$, where $B : X \to X^*$ is a linear bounded (continuous)(i.e. $B \in L(X, X^*)$) non-negative self-adjoint operator, is convex.

(ii) If we consider $X = \mathbb{R}^n$, $B = (b_{ij})_{i,j=1,\dots,n}$ a $n \times n$ symmetric positive semidefinite matrix and the quadratic function $f(x) = \langle Bx, x\rangle = \langle x, Bx\rangle = \sum\limits_{i,j=1}^{n} b_{ij}x_ix_j \geq 0$, $x = (x_1, ..., x_n)^T$, then f is convex.

Solution.

(i) Let be $y, z \in X$. Then

$$\langle By, y\rangle - \langle Bz, z\rangle - 2\langle Bz, y - z\rangle = \langle B(y - z), y - z\rangle \geq 0 \qquad (9.\ 3)$$

because

$$\begin{aligned}
\langle B(y-z), y-z\rangle &= \langle By, y-z\rangle - \langle Bz, y-z\rangle \\
&= \langle By, y\rangle - \langle By, z\rangle - \langle Bz, y\rangle + \langle Bz, z\rangle \\
&= \langle By, y\rangle - 2\langle Bz, y\rangle + \langle Bz, z\rangle \\
&= \langle By, y\rangle - 2\langle Bz, y\rangle + 2\langle Bz, z\rangle - 2\langle Bz, z\rangle + \\
&\quad +\langle Bz, z\rangle \\
&= \langle By, y\rangle - \langle Bz, z\rangle - 2\langle Bz, y-z\rangle.
\end{aligned}$$

We have used here $\langle By, z\rangle = \langle y, B^* z\rangle = \langle y, Bz\rangle = \langle Bz, y\rangle$. Further, by (9.3) follows $f(y) \geq f(z) + 2\langle Bz, y - z\rangle$.

We substitute y by $x_1 \in X$, then by $x_2 \in X$ and we get

$$f(x_1) \geq f(z) + 2\langle Bz, x_1 - z\rangle$$

and

$$f(x_2) \geq f(z) + 2\langle Bz, x_2 - z\rangle.$$

After multiplying the first inequality by $\lambda \in (0, 1)$, the second one by $1 - \lambda$, summing them and setting $z := \lambda x_1 + (1 - \lambda)x_2$, we obtain

$$\begin{aligned}
\lambda f(x_1) + (1-\lambda) f(x_2) &\geq f(z) + \langle Bz, \lambda x_1 + (1-\lambda)x_2\rangle \\
&= f(z) = f(\lambda x_1 + (1-\lambda)x_2),
\end{aligned}$$

which implies that f is convex.

(ii) Consider x and y in $\mathbb{R}^n$ and $\lambda \in (0, 1)$. We have

$$\begin{aligned}
f(\lambda x &+ (1-\lambda)y) = [\lambda x + (1-\lambda)y]^T C[\lambda x + (1-\lambda)y] \\
&= \lambda^2 x^T Cx + \lambda(1-\lambda)x^T Cy + (1-\lambda)\lambda y^T Cx + (1-\lambda)^2 y^T Cy \\
&= \lambda f(x) + (1-\lambda) f(y) + \lambda(1-\lambda)[y^T Cx + x^T Cy - x^T Cx - y^T Cy] \\
&= \lambda f(x) + (1-\lambda) f(y) + \lambda(1-\lambda)[y^T(Cx - Cy) - x^T(Cx - Cy)] \\
&= \lambda f(x) + (1-\lambda) f(y) + \lambda(1-\lambda)(y - x)^T C(x - y).
\end{aligned}$$

Because $(y-x)^T C(y-x) < 0$ the last term in the right-hand side above is less than or equal to 0, so the convexity of f follows, i.e.

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

Exercise 1.2. Show that the infimal convolution of the proper convex functions $f_i : X \to \overline{\mathbb{R}}$, $i = 1, ..., m$, is convex.

Solution. Take $x, y \in X$ and $\lambda \in (0,1)$. It follows

$$\begin{aligned} f(\lambda x + (1-\lambda)y) &= \inf\Big\{ f_1(\lambda x_1 + (1-\lambda)y_1) + ... + f_m(x_m + (1-\lambda)y_m) : \\ &\qquad x = \sum_{i=1}^m x_i, y = \sum_{i=1}^m y_i, x_i \in X, y_i \in X, i = 1, ..., m \Big\} \\ &\leq \sum_{i=1}^m f_i(\lambda x_i + (1-\lambda)y_i) \leq \lambda \sum_{i=1}^m f_i(x_i) + (1-\lambda) \sum_{i=1}^m f_i(y_i) \\ &\leq \lambda \inf_{\sum_{i=1}^m x_i = x} \left(\sum_{i=1}^m f_i(x_i) \right) + (1-\lambda) \inf_{\sum_{i=1}^m y_i = y} \left(\sum_{i=1}^m f_i(y_i) \right) \\ &= \lambda f(x) + (1-\lambda) f(y).\square \end{aligned}$$

Exercise 1.3. If C is convex, then its gauge functional $\gamma_C(x)$ is convex, too.

Solution. Let $x, y \in X$ and $\alpha \in (0,1)$. We have

$$\gamma_C(\alpha x + (1-\alpha)y) = \inf\{\lambda \geq 0 : (\alpha x + (1-\alpha)y) \in \lambda C\}.$$

If $\gamma_C(\alpha x + (1-\alpha)y) = +\infty$, then if one of $\gamma_C(x)$ and $\gamma_C(y)$ is also equal to $+\infty$ Jensen's inequality is fulfilled. Otherwise, there would exist some $\lambda_1 = \gamma_C(x) \in \mathbb{R}$ and $\lambda_2 = \gamma_C(y) \in \mathbb{R}$, such that

$$\alpha x + (1-\alpha)y \in \alpha\lambda_1 C + (1-\alpha)\lambda_2 C = (\alpha\lambda_1 + (1-\alpha)\lambda_2)C,$$

so the assumption above is denied.

More interesting is the second case, i.e. when $\gamma_C(\alpha x + (1-\alpha)y) = \lambda \in \mathbb{R}$. Provided that $\gamma_C(x) = +\infty$ or $\gamma_C(y) = +\infty$, Jensen's inequality is fulfilled. Otherwise denote $\lambda_1 := \gamma_C(x) \in \mathbb{R}$ and $\lambda_2 := \gamma_C(y) \in \mathbb{R}$. If $\lambda_1 = 0$, then $(\alpha x + (1-\alpha)y) \in 0C + \lambda_2 C = \lambda_2 C$, so $\lambda = \lambda_2$. Analogously $\lambda_1 = 0$ yields $\lambda = \lambda_1$.

In both cases Jensen's inequality is fulfilled as equality. Now we may consider $\lambda_1 > 0$ and $\lambda_2 > 0$. We have $\alpha x \in \alpha\lambda_1 C$ and $(1-\alpha)y \in (1-\alpha)\lambda_2 C$, so

$$(\alpha x + (1-\alpha)y) \in (\alpha\lambda_1 + (1-\alpha)\lambda_2)C.$$

From the definition of λ follows $\lambda \leq \alpha\lambda_1 + (1-\alpha)\lambda_2$, i.e.

$$\gamma_C(\alpha x + (1-\alpha)y) \leq \alpha\gamma_C(x) + (1-\alpha)\gamma_C(y).\square$$

Remark: According to a remark in the lecture any sublinear function is convex and because the next exercise guarantees the sublinearity of the gauge of a convex set, one may conclude easier that the gauge of a convex set is convex.

Exercise 1.4. For a convex set C, prove that its gauge functional γ_C is sublinear.

Solution. Proof. Let be given $x_1, x_2 \in X$ and let the infimum be attained (in case that infimum is not attained one has to modify the considerations a little bit: consider "infimum sequences")

$$\gamma_C(x_i) = \min\{\lambda : \lambda \geq 0, x_i \in \lambda C\} = \lambda_i < \infty, i = 1, 2.$$

From here we deduce that there exist $c_1, c_2 \in C$ such that $x_1 = \lambda_1 c_1$, $x_2 = \lambda_2 c_2$ and there exist $c \in C$ and $\lambda \geq 0$ such that $x_1 + x_2 = \lambda c$. Namely $\lambda = \lambda_1 + \lambda_2$ and $c = \frac{\lambda_1}{\lambda_1+\lambda_2}c_1 + \frac{\lambda_2}{\lambda_1+\lambda_2}c_2 \in C$ because $\frac{\lambda_1}{\lambda_1+\lambda_2} + \frac{\lambda_2}{\lambda_1+\lambda_2} = 1$ (convex linear combination).

Indeed, $x_1 + x_2 = \lambda_1 c_1 + \lambda_2 c_2$ and, on the other side $\lambda c = (\lambda_1 + \lambda_2)\frac{\lambda_1}{\lambda_1+\lambda_2}c_1 + \frac{\lambda_2}{\lambda_1+\lambda_2}c_2 = \lambda_1 c_1 + \lambda_2 c_2$. Therefore $x_1 + x_2 = \lambda c$.

By definition

$$\gamma_C(x_1 + x_2) = \inf\{\lambda : \lambda \geq 0, x_1 + x_2 \in \lambda C\}.$$

Because $x_1 + x_2 = \lambda c$ where $c \in C$ and $\lambda = \lambda_1 + \lambda_2$ it follows that

$$\gamma_C(x_1 + x_2) \leq \lambda = \lambda_1 + \lambda_2 = \gamma_C(x_1) + \gamma_C(x_2)$$

and hence the subadditivity is fulfilled.

Furthermore for $\mu > 0$

$$\begin{aligned}\gamma_C(\mu x) &= \inf\{\lambda : \lambda \geq 0, \mu x \in \lambda C\}\\ &= \inf\{\mu\frac{\lambda}{\mu} : \lambda \geq 0, x \in \frac{\lambda}{\mu}C\}\\ &= \mu \inf\{\tilde{\lambda} : \tilde{\lambda} \geq 0, x \in \tilde{\lambda}C\}\\ &= \mu\gamma_C(x),\end{aligned}$$

where we have denoted $\tilde{\lambda} := \frac{\lambda}{\mu}$.
For $\mu = 0$ we have $\gamma_C(0 \cdot x) = \gamma_C(0) = \inf\{\lambda : \lambda \geq 0, 0 \in \lambda C\} = 0 = 0 \cdot \gamma_C(x)$.
If $\gamma_C(x_1)$ or $\gamma_C(x_2)$ is equal $+\infty$ then sublinearity is trivially fulfilled. $\square$

Exercise 1.5. If f is a positively homogeneous proper convex function, then the following statements are true.

(a) $f(\lambda_1 x_1 + ... + \lambda_m x_m) \leq \lambda_1 f(x_1) + ... + \lambda_m f(x_m)$, whenever $\lambda_1 > 0, ..., \lambda_m > 0$ and for all $x_i \in X$, $i = 1, ..., m$.

(b) $f(-x) \geq -f(x)$ for every $x \in X$.

Solution.

(a) Let $\lambda_i > 0$, $i = 1, ..., m$. Consider also some alternative parameters λ_i by $\alpha_i := \frac{\lambda_i}{\sum_{i=1}^{m} \lambda_i}$, $i = 1, ..., m$. It is easy to notice that $\alpha_i > 0$, $i = 1, ..., m$, and $\sum_{i=1}^{m} \alpha_i = 1$.

As f is convex, we may write

$$f\left((1-\alpha_m)\frac{\sum_{i=1}^{m-1}\alpha_i x_i}{1-\alpha_m} + \alpha_m x_m\right) \leq (1-\alpha_m) f\left(\frac{\sum_{i=1}^{m-1}\alpha_i x_i}{1-\alpha_m}\right) + \alpha_m f(x_m),$$

that becomes using that f is positively homogeneous

$$f\left(\sum_{i=1}^{m}\alpha_i x_i\right) \leq (1-\alpha_m)\sum_{i=1}^{m-1}\frac{\alpha_i}{1-\alpha_m} f(x_i) + \alpha_m f(x_m) = \sum_{i=1}^{m}\alpha_i f(x_i).$$

Multiplying the inequality above by $\sum_{i=1}^{m} \lambda_i$ we obtain exactly the requirement.

(b) We have $f(x)+f(-x) \geq f(x-x) = f(0) = 0$, so $f(-x) \geq -f(x)$ for every $x \in X$. $\square$

Exercise 1.6. Any positively homogeneous, subadditive (i.e. sublinear), non-negative and continuous function p on X is a gauge of a convex function.

Solution. Consider $C := \{x \in X : p(x) \leq 1\}$. We will show that $p = \gamma_C$. First we need to prove that C is convex. Let $x, y \in C$ and $\alpha \in (0,1)$. Because of its properties, p is convex due to a remark in the lecture. We have

$$p(\alpha x + (1-\alpha)y) \leq \alpha p(x) + (1-\alpha)p(y) \leq \alpha + (1-\alpha) = 1,$$

so C is a convex set.

Further, using the definition of the gauge corresponding to C we can conclude that $\forall x \in X$ either $\gamma_C(x) = +\infty$ or $\gamma_C(x) = \lambda \in [0, +\infty)$. Let $x \in X$. If the first case applies, there is no $\lambda \geq 0$ such that $x \in \lambda C$, so $\frac{x}{\lambda} \notin C \ \forall \lambda > 0$, i.e. $p(\frac{x}{\lambda}) > 1$, followed by $p(x) > \lambda \ \forall \lambda > 0$. Consequently, $p(x) = +\infty$, so $\gamma_C(x) = p(x)$.

If there is a $\lambda \in (0, +\infty)$ such that $\gamma_C(x) = \lambda$, then $x \in \lambda C$. If $\lambda = 0$, then $x = 0$ and because $p(0) = 0$, we have $\gamma_C(x) = p(x)$. Otherwise, $\frac{x}{\lambda} \in C$, so $p(\frac{x}{\lambda}) \leq 1$, followed by $p(x) \leq \lambda$. Supposing that $p(x) < \lambda$ we obtain that $x \in p(x)C$, that contradicts the definition of λ. Consequently, $p \equiv \gamma_C$. $\square$

Exercise 1.7. Prove that a sublinear functional p is linear if and only if for all $x \in X$ it holds $p(-x) = -p(x)$.

Solution. *Necessity.* p linear implies $p(-x) = p(-1 \cdot x) = (-1)p(x) = -p(x) \ \forall x \in X$.

Sufficiency. For all $\lambda \geq 0$ and all $x \in X$ we have $p(\lambda x) = \lambda p(x)$ due to the sublinearity of p. Consider now a $\lambda < 0$. It is clear that $p(-\lambda x) = -p(-\lambda x) = -(-\lambda)p(x) = \lambda p(x)$. So for any real λ it is true that $p(\lambda x) = \lambda p(x)$.

Further, consider x and $y \in X$. We have

$$-p(x+y) = p(-x-y) \leq p(-x) + p(-y) = -p(x) - p(y) = -[p(x) + p(y)],$$

so $p(x+y) \geq p(x) + p(y)$. As p is sublinear it follows that it is also linear. $\square$

Exercise 3.1. Let X be a Banach space and $x^0 \in X$. The function $f : X \to \overline{\mathbb{R}}$ is lower semicontinuous at x^0 if and only if

$$\underline{\lim}_{x \to x^0} f(x) \geq f(x^0).$$

Solution. *Necessity.* Assume that $\underline{\lim}_{x \to x^0} f(x) < f(x^0)$. Then there is an $\varepsilon > 0$ less than $f(x^0) - \underline{\lim}_{x \to x^0} f(x)$. As f is lower semicontinuous at x^0, there is a neighborhood U of x^0 where any point x satisfies $-\varepsilon < f(x) - f(x^0)$. As U is a neighborhood of x^0, there is a ball $B(x^0, \delta)$ included in U or equal to it. For each $x \in B(x^0, \delta)$ it is sure that $f(x) > f(x^0) - \varepsilon$, so

$$\inf_{x \in B(x^0,\delta)} \geq f(x^0) - \varepsilon > \underline{\lim}_{x \to x^0} f(x).$$

Further follows

$$\inf_{x \in B(x^0,\delta)} > \sup_{\delta > 0} \inf_{x \in B(x^0,\delta)},$$

that is false, so the initial assumption fails.
Sufficiency. Let be $\varepsilon > 0$. We have $\underline{\lim}_{x \to x^0} f(x) \geq f(x^0) > f(x^0) - \varepsilon$, so $\sup_{\delta > 0} \inf_{x \in B(x^0,\delta)} > f(x^0) - \varepsilon$. Consequently, there is a $\delta > 0$ such that $\inf_{x \in B(x^0,\delta)} f(x) > f(x^0) - \varepsilon$, so for all $x \in B(x^0, \delta)$ we have $f(x) \geq \inf_{x \in B(x^0,\delta)} f(x) > f(x^0) - \varepsilon$, followed by $-\varepsilon < f(x) - f(x^0)$, so f is lower semicontinuous at x^0. $\square$

Exercise 3.2. Let X be a reflexive Banach space, $B : X \to X^*$ a linear, bounded, non-negative operator (i.e. $\langle Bx, x\rangle \geq 0 \; \forall x \in X$), $B^* : X \to X^*$ its adjoint operator and the function $f : X \to \mathbb{R}$, $f(x) = \langle Bx, x\rangle$. Then $f'(x) = Bx + B^*x \; \forall x \in X$.

Solution. We have for all $x \in X$ and all $h \in X$

$$\lim_{t \to 0} \frac{f(x+th) - f(x)}{t} = \lim_{t \to 0} \frac{\langle B(x+th), x+th\rangle - \langle Bx, x\rangle}{t}$$

$$= \lim_{t \to 0} \frac{\langle Bx, x\rangle + \langle Bx, th\rangle + \langle Bth, x\rangle + \langle Bth, th\rangle - \langle Bx, x\rangle}{t}$$

$$
\begin{aligned}
&= \lim_{t\to 0} \frac{t\langle Bx, h\rangle + t\langle Bh, x\rangle + t^2\langle Bh, h\rangle}{t} \\
&= \lim_{t\to 0} \frac{t\langle Bx, h\rangle + t\langle B^*x, h\rangle + t^2\langle Bh, h\rangle}{t} \\
&= \langle Bx + B^*x, h\rangle = \langle (B + B^*)x, h\rangle,
\end{aligned}
$$

i.e. $f' \equiv Bx + B^*x$. □

Exercise 3.3. Let be X a Hilbert space and consider the function $f : X \to \mathbb{R}$, $f(x) = \|x\| \; = \langle x, x\rangle^{\frac{1}{2}}$. Then

$$f'(x) = \frac{x}{\|x\|}, \quad x \neq 0.$$

Solution. For some $x, h \in X$ we have

$$
\begin{aligned}
\lim_{t\to 0} \frac{f(x_0 + th) - f(x_0)}{t} &= \lim_{t\to 0} \frac{\langle x + th, x + th\rangle^{1/2} - \langle x, x\rangle^{1/2}}{t} \\
&= \lim_{t\to 0} \frac{(\langle x, x\rangle + 2t\langle x, h\rangle + t^2\|h\|^2)^{1/2} - \langle x, x\rangle^{1/2}}{t} \\
&= \lim_{t\to 0} \frac{\|x\|^2 + 2t\langle x, h\rangle + t^2\|h\|^2 - \|x\|^2}{t[(\langle x, x\rangle + 2t\langle x, h\rangle + t^2\|h\|^2)^{1/2} + \langle x, x\rangle^{1/2}]} \\
&= \lim_{t\to 0} \frac{2t\langle x, h\rangle}{t[(\langle x, x\rangle + 2t\langle x, h\rangle + t^2\|h\|^2)^{1/2} + \langle x, x\rangle^{1/2}]} \\
&+ \lim_{t\to 0} \frac{t^2\|h\|^2}{t[(\langle x, x\rangle + 2t\langle x, h\rangle + t^2\|h\|^2)^{1/2} + \langle x, x\rangle^{1/2}]} \\
&= \lim_{t\to 0} \frac{2\langle x, h\rangle}{(\langle x, x\rangle + 2t\langle x, h\rangle + t^2\|h\|^2)^{1/2} + \langle x, x\rangle^{1/2}} \\
&= \frac{2\langle x, h\rangle}{(\langle x, x\rangle)^{1/2} + \langle x, x\rangle^{1/2}} \\
&= \left\langle \frac{x}{\|x\|}, h \right\rangle,
\end{aligned}
$$

so the assertion holds. □

Exercise 8.1. Consider the problem

$$(\mathcal{P}) \qquad \inf_{x\in X} \left\{ f(x) + \sum_{i=1}^{m} g_i(A_i x) \right\},$$

where $A \in L(X, Y_i)$, X and Y_i, $i = 1, ..., m$, are normed spaces and $f : X \to \overline{\mathbb{R}}$, $g_i : Y_i \to \mathbb{R}$, $i = 1, ..., m$, functions. Determine its dual problem and calculate the corresponding optimality conditions.

Solution. Consider the following perturbation function $\Phi : X \times Y_1 \times ... \times Y_m \to \overline{\mathbb{R}}$,

$$\Phi(x, y_1, ..., y_m) = f(x) + \sum_{i=1}^{m} g_i(A_i x - y_i).$$

To deduce the dual problem to $(\mathcal{P})$ we calculate the conjugate of the perturbation function, $\Phi : X^* \times Y_1^* \times ... \times Y_m^* \to \overline{\mathbb{R}}$,

$$\Phi^*(x^*, y_1^*, ..., y_m^*) = \sup_{\substack{x \in X, y_i \in Y_i, \\ i=1,...,m}} \Big\{ \langle x^*, x \rangle + \sum_{i=1}^{m} \langle y_i^*, y_i \rangle - f(x) - \sum_{i=1}^{m} g_i(A_i x - y_i) \Big\},$$

where by "$*$" we denote the duals of the corresponding spaces.

Introducing the new variables $p_i := A_i x - y_i$, $i = 1, ..., m$, and considering $x^* = 0$, we get after separating the terms in the expression above,

$$\begin{aligned} \Phi^*(0, y_1^*, ..., y_m^*) &= \sup_{x \in X} \Big\{ \sum_{i=1}^{m} \langle y_i^*, A_i x \rangle - f(x) \Big\} + \sum_{i=1}^{m} \sup_{p_i \in Y_i} \{ -\langle y_i^*, p_i \rangle - g_i(p_i) \} \\ &= f^* \left(\sum_{i=1}^{m} A_i^T y_i^* \right) + \sum_{i=1}^{m} g_i^*(-y_i^*). \end{aligned}$$

The dual problem to $(\mathcal{P})$ is

$$(\mathcal{D}) \qquad \sup_{\substack{y_i \in Y_i^*, \\ i=1,...,m}} \Big\{ - f^* \left(\sum_{i=1}^{m} A_i^T y_i^* \right) - \sum_{i=1}^{m} g_i^*(-y_i^*) \Big\}.$$

The optimality conditions arise easily. Considering $\bar{x}$ a solution to the primal problem and $(\bar{y}_1^*, ..., \bar{y}_m^*)$ one to the dual, we have strong duality, i.e.

$$f(\bar{x}) + \sum_{i=1}^{m} g_i(A_i \bar{x}) = -f^* \left(\sum_{i=1}^{m} A_i^T \bar{y}_i^* \right) - \sum_{i=1}^{m} g_i^*(-\bar{y}_i^*). \tag{9. 4}$$

Young's inequality yields

$$f(\bar{x}) + f^* \left(\sum_{i=1}^{m} A_i^T \bar{y}_i^* \right) \geq \langle \sum_{i=1}^{m} A_i^T \bar{y}_i^*, \bar{x} \rangle$$

and

$$g_i(A_i \bar{x}) + g_i^*(-\bar{y}_i^*) \geq \langle -\bar{y}_i^*, A_i \bar{x} \rangle, i = 1, ..., m.$$

Summing them we obtain

$$f(\bar{x}) + \sum_{i=1}^{m} g_i(A_i\bar{x}) + f^*\left(\sum_{i=1}^{m} A_i^T \bar{y}_i^*\right) + \sum_{i=1}^{m} g_i^*(-\bar{y}_i^*) \geq 0,$$

where (9.4) yields equality. So the inequalities obtained from Young's one must be also fulfilled as equalities, i.e. the optimality conditions are

$$f(\bar{x}) + f^*\left(\sum_{i=1}^{m} A_i^T \bar{y}_i^*\right) = \left\langle \sum_{i=1}^{m} A_i^T \bar{y}_i^*, \bar{x} \right\rangle$$

and

$$g_i(A_i\bar{x}) + g_i^*(-\bar{y}_i^*) = \langle -\bar{y}_i^*, A_i\bar{x} \rangle, i = 1, ..., m.\square$$

References

[1] A. Göpfert, *Mathematische Optimierung in allgemeinen Vektorräumen.* Teubner Verlag, Leipzig, 1973

[2] H. Heuser, *Funktionalanalysis*, Teubner Verlag, Stuttgart 1975

[3] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, New Jersey, 1970

[4] E. Zeidler, *Nonlinear Functional Analysis and its Applications III*, Variational Methods and Optimization, Springer Verlag, New York, Berlin, Heidelberg, Tokyo, 1985