Chapter 3

Characterization of best approximations

In this chapter we study properties which characterite solutions of the approximation problem. There is a big difference in the treatment of this question in Hilbert and Banach spaces. A harmonization is possible by considering semi-inner products.

The material presented is standard and can be found in textbooks on approximation theory and convex analysis; see for instance Cheney [6], Deutsch [8], Holmes [11] and Singer [17]. For the material concerning the semi-inner products we refer to Dragomir [9], Giles [10], Penot and Ratsimahalo [15] and Rosca [16].

3.1 Characterization of best approximations in Hilbert spaces

Theorem 3.1 (Kolmogorov's criterion). *Let $\mathcal{H}$ be a Hilbert space and let C be a nonempty closed convex subset of $\mathcal{H}$. Then for $x \in \mathcal{H}$ and $w \in C$ the following conditions are equivalent:*

(a) $w = P_C(x)$.

(b) $\langle x - w | y - w \rangle \le 0$ *for all* $y \in C$.

(c) $\|x - w\|^2 + \|y - w\|^2 \le \|x - y\|^2$ *for all* $y \in C$.

Proof:
$(a) \implies (b)$ Let $y \in C$ and $t \in (0, 1]$. Then $ty + (1 - t)w \in C$ and we have

$$0 \le \|x - (ty + (1 - t)w)\|^2 - \|x - w\|^2 = -2t\langle x - w|y - w\rangle + t^2\|y - w\|^2 .$$

Dividing by t and using $t \downarrow 0$ we obtain (b).
$(b) \implies (c)$ Using (b) we obtain

$$\|x-y\|^2 = \|x-w+w-y\|^2 = \|x-w\|^2+\|w-y\|^2+2\langle x-w|w-y\rangle \ge \|x-w\|^2+\|y-w\|^2 .$$

$(c) \implies (a)$ Obviously. ■

Kolmogorov's criterion in Hilbert spaces[1], i.e. the equivalence $(a) \iff (b)$ is saying that the best approximation w to x is such that $x - w$ forms an obtuse angle with all direction $y - w$ from w into the set C. Another way of looking at this criterion is to define, for $x \in \mathcal{H}\backslash C$ the hyperplane

$$H_x := \{y \in \mathcal{H} : \langle x - P_C(x) | y\rangle = 0\}.$$

One then notes that the translate $H_x + P_C(x)$ supports the convex set C at $P_C(x)$. Lateron, we will present Kolmogorov's criterion also for Banach spaces. Here the formulation and the proof is much more delicate.

The equivalence $(a) \iff (c)$ leads to the so called **strong uniqueness** of $P_C(x)$:

$$\|x - P_C(x)\|^2 \le \|x - y\|^2 - \|y - P_C(x)\|^2 \text{ for all } y \in C.$$

Theorem 3.2. *Let $\mathcal{H}$ be a Hilbert space and let C be a nonempty closed convex subset of $\mathcal{H}$. Then the metric projection P_C has the following properties:*

(1) $P_C(x)$ *is a singleton for each* $x \in \mathcal{H}$.

(2) $\|P_C(x) - P_C(x')\|^2 + \|(I - P_C)(x) - (I - P_C)(x')\|^2 \le \|x - x'\|^2$ *for all* $x, x' \in \mathcal{X}$.

(3) $P_C : X \longrightarrow C$ *is nonexpansive.*

(4) $P_C : X \longrightarrow C$ *is uniformly continuous.*

(5) P_C *is idempotent, i.e.* $P_C \circ P_C = P_C$.

(6) P_C *is positively homogeneous, i.e.* $P_C(tx) = tP_C(x)$ *for all* $t > 0, x \in \mathcal{H}$, *if* C *is a cone.*

Proof:
Ad (1) C is a Chebyshev set; see Theorem 2.17.
Ad (2) Let $w = P_C(x), w' = P_C(x')$. Then

$$\begin{aligned} \|x - x'\|^2 &= \|P_C(x) - P_C(x') + (I - P_C)(x) - (I - P_C)(x')\|^2 \\ &= \|P_C(x) - P_C(x')\|^2 + \|(I - P_C)(x) - (I - P_C)(x')\|^2 \\ &\quad +2\langle P_C(x) - P_C(x') | (I - P_C)(x) - (I - P_C)(x')\rangle \\ &\ge \|P_C(x) - P_C(x')\|^2 + \|(I - P_C)(x) - (I - P_C)(x')\|^2 \end{aligned}$$

where we have used (b) in Theorem 3.1.
Ad (3) Follows from (2).
Ad (4) Follows from the nonexpansivity; see (3).
Ad (5) Follows from the fact that P_C is the identity on C.
Ad (6) Let $w = P_C(x)$. Choose $t > 0$ and $y \in C$. Then $tw, t^{-1}y \in C$ and we obtain

$$\|tx - tw\| = t\|x - w\| \le t\|x - t^{-1}y\| = \|tx - y\|.$$

This shows $tP_C(x) = tw = P_C(tx)$. ■

[1]Historically, it was developed mainly for the best approximation in the space of continuous functions in an interval. The underlying Banach space $C[a, b]$ is not reflexive and not strictly convex.

Theorem 3.3. *Let $\mathcal{H}$ be a Hilbert space and let U be a closed subspace of $\mathcal{H}$. Then the metric projection P_U das the following properties:*

(1) $P_U(x)$ *is a singleton for each* $x \in \mathcal{H}$.

(2) $P_U(x) = x$ *for each* $x \in U$.

(3) P_U *is idempotent, i.e.* $P_U = P_U \circ P_U$.

(4) For each $x \in \mathcal{H}$ *we have:* $w = P_U(x)$ *if and only if* $\langle x - w | u \rangle = 0$ *for all* $u \in U$.

(5) $x - P_U(x) \in U^\perp$ *for each* $x \in X$.

(6) P_U *is selfadjoint, i.e.* $\langle P_U(x)|y\rangle = \langle x|P_U(y)\rangle$ *for all* $x, y \in \mathcal{H}$.

(7) $\|P_U(x)\| \le \|x\|$ *for each* $x \in \mathcal{H}$.

(8) P_U *is a linear mapping.*

(9) P_U *is a linear nonexpansive.*

(10) $\langle P_U(x)|x\rangle \ge 0$ *for every* $x \in \mathcal{H}$.

(11) If C *is a closed affine subset of* $\mathcal{H}$ *then* P_C *is affine and weakly continuous.*

Proof:
Ad $(1), (2), (3)$ are clear because U is Chebyshev set; see Theorem 2.17.
Ad (4) From (b) in Theorem 3.1 we conclude $\langle x - P_U(x), u\rangle$ for all $u \in U$.
Ad (5) Follows from (4)
Ad (6) Using (5) we obtain

$$\langle P_U(x)|y\rangle = \langle P_U(x)|y - P_U(y) + P_U(y)\rangle = \langle P_U(x)|P_U(y)\rangle .$$

By symmetry reasons we obtain $\langle x|P_U(y)\rangle = \langle P_U(x)|P_U(y)$.
Ad (7) We have for $x \in \mathcal{H}$ due to (5)

$$\|x\|^2 = \|P_U(x) + (x - P_U(x))\|^2 = \|P_U(x)\|^2 + \|x - P_U(x)\|^2 \ge \|P_U(x)\|^2 .$$

Hence $\|x\| \ge \|P_U(x)\|$.
Ad (8) Let $x, y \in \mathcal{H}, a, b \in \mathbb{R}$. Then by (5) $x - P_U(x), y - P_U(y) \in U^\perp$. Hence

$$ax + by - (aP_U(x) + bP_U(y)) = a(x - P_U(x)) + b(y - P_U(y)) \in U^\perp .$$

Therefore $w := aP_U(x) + bP_U(y) = P_U(ax + by) \in U$ by (4).
Ad (9) Follows from (7) and (8).
Ad (10) Follows with (6) as follows:

$$\langle P_U(x)|x\rangle = \langle P_U(P_U(x))|x\rangle = \langle P_U(x)|P_U(x)\rangle = \|P_U(x)\| \ge 0 .$$

Ad (11) Let $C = x^0 + U$ with a closed linear subspace U. Let $x \in \mathcal{H}$. Then

$$\|(x - x^0) - u\| = \|x - (x^0 + u)\| \text{ for all } u \in U .$$

This implies $x^0 + P_U(x - x^0) = P_{x^0+U}(x)$. Let

$$L : \mathcal{H} \ni x \longmapsto P_{x^0+U}(x) - P_{x^0+U}(\theta) = x^0 + P_U(x - x^0) - P_{x^0+U}(\theta) \in \mathcal{H}.$$

Since P_U is a linear continous mapping (see (8),(9)), L is linear and continuous too.
Let $x \in \mathcal{H}$ and let $x = w - \lim_n x^n$. Then

$$\lim_n \langle x^n | L^* y\rangle = \langle x | L^* y\rangle , \lim_n \langle Lx^n | y\rangle = \langle Lx | y\rangle \text{ for all } y \in \mathcal{H}.$$

This shows $w - \lim_n Lx^n = Lx$ and therefore $w - \lim_n P_{x^0+U}(x^n) = P_{x^0+U}(x)$. ■

Theorem 3.4 (Reduction principle). *Let $\mathcal{H}$ be Hilbert space, let C be a convex subset and let $U \subset \mathcal{H}$ be a linear subspace which is a Chebyshev set. Suppose that $C \subset U$. Then we have:*

(a) $P_C \circ P_U = P_C = P_U \circ P_C$.

(b) *dist*$(x, C)^2 =$ *dist*$(x, U)^2 +$ *dist*$(P_U(x), C)^2$ *for all* $x \in \mathcal{H}$.

Proof:
Notice that $P_C(x)$ is empty or a singleton due to the fact that C is a convex subset of a Hilbert space. Moreover $P_U(x)$ is a singleton for each $x \in \mathcal{H}$ since U is a Chebyshev set.
Ad (a) $P_U \circ P_C = P_C$ since $C \subset U$. Let $x \in \mathcal{H}, y \in C$. Then $y \in U$ and $x - P_U(x) \in U^\perp$ due to (5) in Theorem 3.3. Therefore

$$\|x - y\|^2 = \|x - P_U(x)\|^2 + \|P_U(x) - y\|^2 . \tag{3.1}$$

We conclude that y minimizes $u \longmapsto \|x - u\|^2$ iff y minimizes $v \longmapsto \|P_U(x) - v\|^2$. This shows that $P_C(x)$ exists iff $P_C(P_U(x))$ exists and $P_C(x) = P_C(P_U(x))$ holds true. We obtain that $P_C(x) \neq \emptyset$ iff $P_C(P_U(x)) \neq \emptyset$. This implies (a).
Ad (b) Follows from (3.1). ■

The interpretation of the reduction principle is the following: To find the best approximation to x from C we first project x onto U and then project $P_U(x)$ onto C; the result is $P_C(x)$.

Corollary 3.5. *Let $\mathcal{H}$ be a Hilbert space, let $C \subset \mathcal{X}$ be a nonempty closed convex subset of $\mathcal{H}$. Then $P_C : X \longrightarrow C$ satifies*

$$\|P_C(x) - P_C(y)\|^2 \le \langle P_C(x) - P_C(y) | x - y\rangle , \; x, y \in \mathcal{H}. \tag{3.2}$$

Proof:
Let $x, y \in \mathcal{H}$. From Kolmogorov's criterion we have:

$$\langle x - P_C(x) | P_C(y) - P_C(x)\rangle \le 0 , \; \langle y - P_C(y) | P_C(x) - P_C(y)\rangle \le 0 .$$

Adding these two inequalites implies (3.2). ■

The property (3.2) of metric projections in Hilbert spaces is called **firmly nonexpansivity**. We shall come back to this property in a general context in Chapter 7. Notice that we obtain from (3.2) the nonexpansivity of P_C.

3.2 Duality mapping

Let $\mathcal{X}$ be a Banach space. We set:

$$J(x) := \{\lambda \in \mathcal{X}^* : \langle \lambda, x \rangle = \|\lambda\|\|x\|, \|\lambda\| = \|x\|\},\ x \in \mathcal{X}\,.$$

In general, J is a set-valued mapping; we write therefore $J : \mathcal{X} \rightrightarrows \mathcal{X}^*$. This mapping is called the **(normalized) duality mapping**[2]. The duality map plays a crucial role in solving (nonlinear) problems in the setting of Banach spaces, due to the lack of an inner product in a Banach space. We will see that this map is also helpful in describing properties of the metric projections and the geometry of Banach spaces. Let us collect some facts concerning the duality map.

Lemma 3.6. *Let $\mathcal{X}$ be a Banach space. Then we have:*

(1) $J(x)$ is nonempty for each $x \in \mathcal{X}$ and $J(\theta) = \{\theta\}$.

(2) J is a homogenous mapping, i.e. $J(ax) = aJ(x)$ for all $x \in \mathcal{X}, a \geq 0$.

(3) $J(x)$ is a convex set for all $x \in \mathcal{X}$.

(4) $J(x)$ is a weak compact set.*

(5) The duality map is **monotone** *in the following sense:*

$$\langle \lambda - \mu, x - y \rangle \geq 0 \text{ for } x, y \in \mathcal{X}, \lambda \in J(x), \mu \in J(y)\,. \tag{3.3}$$

(6) The duality map is surjective if and only if every functional $\lambda \in \mathcal{X}^*$ attains its maximum on $\overline{S}_1$.

(7) The range of J is dense in $\mathcal{X}^*$.

(8) J is surjective if and only if $\mathcal{X}$ is reflexive.

(9) If $\mathcal{X}$ is a Hilbert space then the duality map is the identity.

Proof:
Ad (1) Evidentely, $J(\theta) = \{\theta\}$. Let $x \in \overline{S}_1$. Then there exists $\lambda \in \mathcal{X}^*$ with $\|\lambda\| = 1$ such hat $\langle \lambda, x \rangle = \|x\| = \|x\|\|\lambda\|$ (continuation theorem of Hahn-Banach/see the preliminaries in the preface). This shows that $\lambda \in J(x)$. For an arbitrary $x \in \mathcal{X}\backslash\{\theta\}$ apply the reasoning above to $x\|x\|^{-1}$.
Ad (2) This is easily verified by distinguishing the case $a = 0, a \neq 0$. . Notice $a\emptyset = \emptyset, a \in \mathbb{R}$.
Ad (3) For $x = \theta$ nothing has to be proved. Let $x \neq \theta$, $\lambda, \mu \in J(x)$ and $\sigma := a\lambda + (1-a)\mu$ with $a \in [0,1]$. Then $\|x\|^{-1}\|\sigma\| \leq 1$ and therefore

$$\|x\| \geq \langle \|x\|^{-1}\sigma, x \rangle = a\|x\| + (1-a)\|x\| = \|x\|\,.$$

This shows $\|x\|^2 = \langle \sigma, x \rangle$ and $\|\sigma\| = \|x\|$.
Ad (4) $J(x)$ is a closed subset of the weak* compact of the ball with radius $\|x\|$ and

[2] The concept of a duality mapping in a Banach space $\mathcal{X}$ was introduced by Beurling and Livingston [3].

therefore weak* compact (Theorem of Banach-Alaoglu).
Ad (5) Let $x, y \in \mathcal{X}$ and $\lambda \in J(x), \mu \in J(y)$. Then

$$\begin{aligned}\langle\lambda-\mu, x-y\rangle \geq &= \langle\lambda, x\rangle - \langle\lambda, y\rangle - \langle\mu, x\rangle + \langle\mu, y\rangle \\ &\geq \|x\|^2 + \|y\|^2 - (\|\lambda\|\|y\| - \|\mu\|\|x\|) = (\|x\| - \|y\|)^2 \quad (3.4)\end{aligned}$$

Now, we read off the monotonicity.
Ad (6) Let $\lambda \in \mathcal{X}^*$. Then there exists $x \in \mathcal{X}$ with $\langle\lambda, x\rangle = \|\lambda\|$: see the preliminaries in the preface. Let $y := \|\lambda\|x$. Then $\lambda \in J(y)$ since $\|\lambda\| = \|y\|$ and

$$\langle\lambda, y\rangle = \langle\lambda, \|\lambda\|x\rangle = \|\lambda\|\langle\lambda, x\rangle = \|\lambda\|^2 = \|y\|^2 .$$

Let $\lambda \in \mathcal{X}^*$ and let $y \in \mathcal{X}^*$ with $J(y) = \lambda$. If $y = \theta$ then $\lambda = \theta$ and nothing has to be proved. Let $y \neq \theta$. Then

$$\langle\lambda, y\|y\|^{-1}\rangle = \|y\|^{-1}\langle\lambda, y\rangle = \|y\| = \|\lambda\| ,$$

which implies that λ attains its norm in $\|y\|^{-1}y \in \overline{S}_1$.
Ad (7) This is the theorem of Bishop-Phelps; see [7].
Ad (8) This follows from (5) by using the Theorem of James which says that a Banach space is reflexive iff every functional $\lambda \in \mathcal{X}^*$ attains its maximum on $\overline{S}_1$.
Ad (9) Let $\lambda \in J(x)$. By the Riesz-representation theorem we have $y \in \mathcal{X}$ with $\langle y|x\rangle = \langle\lambda, x\rangle = \|x\|^2, \|\lambda\| = \|y\| = \|x\|$. By the parallelogram identity it is easily seen that $y = x$. ■

Example 3.7. *Let for $1 < p < \infty$ and let $\mathcal{X} := L_p(\Omega)$ be the Banach space of p-integrable functions on the measurable set $\Omega \subset \mathbb{R}^n$. In this case, $L_p(\Omega)^* = L_q(\Omega)$ with $1/p+1/q = 1$ and one has*

$$J(x) = \{\|x\|^{p-2}|x|^{p-1}\mathit{sign}(x)\}, \ x \neq \theta .$$

Later on, we find a way for computing the result above. □

Lemma 3.8. *Let $\mathcal{X}$ be a Banach spoace with duality mapping J_X. Let $x \in \mathcal{X}$ and $(x^n)_{n\in\mathbb{N}}$ such that*

$$\lim_n \langle J_{\mathcal{X}}(x^n) - J_{\mathcal{X}}(x), x^n - x\rangle = 0 .$$

Then $x = \lim_n x^n$.

Proof:
See [5]. ■

Lemma 3.9. *Let $\mathcal{X}$ be a Banach spoace with duality mapping J_X. Let $(s_n)_{n\in\mathbb{N}}$ be a non-increasing sequence of positive numbers with $\lim s_n = \infty$ and let $(x^n)_{n\in\mathbb{N}}$. Assume*

$$\langle s_n J_{\mathcal{X}}(x^n) - s_m J_{\mathcal{X}}(x), x^n - x^m\rangle =\leq 0 , \ m, n \in \mathbb{N} .$$

Then $(\|x^n\|)_{n\in\mathbb{N}}$ is a nondecreasing sequence. If it is bounded then $(x^n)_{n\in\mathbb{N}}$ is convergent.

Proof:
See [4]. ■

As we know, the duality mapping J on a Banach space is set-valued in general. This is more or less a consequence of the continuation theorem of Hahn-Banach. To illustrate this consider the following example.

Example 3.10. *Consider* $\mathcal{X} := \mathbb{R}^2$ *endowed with the* l_1*-norm and let* U *be the one-dimensional subspace* $\mathbb{R}^1 \times \{\theta\}$. *Then* $\mu : U \ni (x,y) \longmapsto x \in \mathbb{R}$ *is a linear continuous functional on* U *with* $\|\mu\| = 1 = \langle \mu, (1,0)\rangle$. *Each continuation of* μ *to a continuous functional* λ *on* $\mathbb{R}^2$ *has the representation* $\lambda : \mathbb{R}^2 \ni (x,y) \longmapsto \alpha x + \beta y \in \mathbb{R}$. *It is easy to verify that such a functional* λ *has the property* $\lambda_{|U} = \mu$ *and* $\|\lambda\| = 1$ *if* $|\beta| \leq 1$; *notice that the dual norm on* $\mathbb{R}^2$ *is the* l_∞*-norm. Thus, for each such* λ *we have*

$$\lambda \in X^*, \langle \lambda, (1,0)\rangle = 1 = \|(1,0)\|^2, \|\lambda\| = 1 = \|(1,0)\|,$$

so that $\lambda \in J((1,0))$. □

Example 3.11. *Let* K *be a compact topological space and let* $C(K)$ *be the space of continuous functions* f *from* K *into* $\mathbb{R}$. $C(K)$ *is a Banach space under the supremum norm* $\|\cdot\|_\infty : \|f\|_\infty := \sup_{\xi\in K}|f(\xi)|$. *Its dual space* $C(K)^*$ *can be represented in different ways. In the case* $K = [0,1]$ *the representation as the space* $BV[0,1]$ *of functions with bounded variations is approriate. Each function* g *of bounded variation leads to a Stieltjes-integral on* $C(K)$ *the elements of* $C(K)^*$ *have the following presentation:*

$$\forall \lambda \in C(K)^* \, \exists g \in BV[0,1] \, (\langle \lambda, f\rangle = \int_K f dg).$$

In the general case, we have to consider the set of **signed Borel measure** μ *on* K. *Each such measure* μ *has a decomposition* $\mu = \mu_+ - \mu_-$ *where* μ_+, μ_- *are nonnegative measures;* $|\mu| := \mu_+ + \mu_-$ *is called the* **total variation** *of* μ. *Now, the dual space* $C(K)^*$ *can be described as follows: For each* $\lambda \in C(K)^*$ *there exists a uniquely determined finite and regular signed Borel measure with*

$$\langle \lambda, f\rangle = \int_K f d\mu, \; f \in C(K).$$

„Finite" means that the range of μ *is contained in* $\mathbb{R}$ *and „regular" means that the measure* $\mu(A)$ *of each Borel subset* A *of* K *can be approximated from interior by the measure of open subsets. The norm in* $C(K)^*$ *under this representation is given by the total variation. Examples of functionals in* $C(K)^*$ *are Dirac's point measures* $\delta_\xi, \xi \in K : \langle \delta_\xi, f\rangle = f(\xi)$. *Now the duality map* J *on* $C(K)$ *is given as follows:*

$$J(f) = \{\mu : \mu \textit{ finite, regular, signed measure on } K, \int_K f d\mu = |\mu|\|f\|_\infty, |\mu| = \|f\|_\infty\}.$$

If we choose the function $f \equiv 1$ *then* $J(f)$ *is the set of probability measures on the Borel sigma algebra of* K. □

The property (4) in Lemma 3.6 has a nice consequence concerning the continuity of J on the unit sphere.

Lemma 3.12. *Let* $\mathcal{X}$ *be a Banach space with the duality mapping* J *and let* $x \in \overline{S}_1$ *with* $\#J(x) = 1$. *Suppose we have sequences* $(x^n)_{n\in\mathbb{N}}$, $(\lambda^n)_{n\in\mathbb{N}}$ *with* $\|x^n\| = 1, n \in \mathbb{N}, \lim x^n = x, \lambda^n \in J(x^n), n \in \mathbb{N}$. *Then* $w^* - \lim_n \lambda^n = J(x)$.

Proof:
We have $\|\lambda^n\| = \|x^n\| = \langle\lambda^n, x^n\rangle = 1, n \in \mathbb{N}$. Since the unit ball in $\mathcal{X}^*$ is $\sigma(X^*, X)$ compact (Theorem of Alaoglu-Banach) the sequence $(\lambda^n)_{n\in\mathbb{N}}$ posesses $\sigma(\mathcal{X}^*, \mathcal{X})$ convergent subsequences. Let $(\lambda^{n_k})_{k\in\mathbb{N}}$ be such a subsequence of $(\lambda^n)_{n\in\mathbb{N}}$; $\lambda := w^* - \lim_k \lambda^{n_l}$. Then $\|\lambda\| \leq 1$ and

$$\begin{aligned} |\langle\lambda, x\rangle - 1| &= |\langle\lambda, x\rangle - \langle\lambda^{n_k}, x^{n_k}\rangle| \\ &\leq |\langle\lambda, x\rangle - \langle\lambda^{n_k}, x\rangle| + |\langle\lambda^{n_k}, x\rangle - \langle\lambda^{n_k}, x^{n_k}\rangle| \\ &\leq |\langle\lambda, x\rangle - \langle\lambda^{n_k}, x\rangle| + \|x^{n_k} - x\| \end{aligned}$$

By taking $k \to \infty$ this implies $\langle\lambda, x\rangle = 1$, $\|\lambda\| = 1$ and therefore $\lambda \in J(x)$. Since $\#J(x) = 1$ we conclude that $\lambda = w^* - \lim_n \lambda^n = J(x)$.

■

In several cases, when one works in specific Banach spaces like $l_p, L_p(\Omega), W^{1,p}(\Omega), 1 < p < \infty$, a generalization of the normalized duality map is appropriate. In a more general consideration, one defines a duality mapping as

$$J_h : \mathcal{X} \ni x \longmapsto \{\lambda \in X^* : \langle\lambda, x\rangle = \|x\|\|h(x)\|, \|\lambda\| = \|h(x)\|\} \subset X^*.$$

Here, h is called a **gauge-function**. It is a non-decreasing map $h : [0, \infty) \longrightarrow [0, \infty)$ and has the properties $h(0) = 0, h(s) > 0$ for $s > 0$ and $\lim_{s\to\infty} h(s) = \infty$. We set

$$G := \{h : [0, \infty) \longrightarrow [0, \infty) : h \text{ a gauge-function}\}.$$

The relation between the normalized duality map and the generalized duality map J_h is easily seen to be

$$J_h(x) = \frac{h(\|x\|)}{\|x\|} J(x), \ x \in \mathcal{X}.$$

Duality mappings will be used as a main tool for studying the properties of the metric projection. We restrict ourselves to a consideration using a normalized duality mapping only.

Lemma 3.13. *Let $\mathcal{X}$ be a Banach space such that the dual space $\mathcal{X}^*$ is a strictly convex Banach space. Then the duality map J is a single-valued mapping.*

Proof:
Let $x \in \mathcal{X}$. If $x = \theta$ then $J(x) = \{\theta\}$. Suppose that $x \neq \theta$. Let $\lambda, \mu \in J(x)$. Then $\|\lambda\| = \|\mu\| = \|x\|$ and

$$2\|\lambda\|\|x\| = 2\|x\|^2 = \langle\lambda + \mu, x\rangle \leq \|\lambda + \mu\|\|x\|.$$

From this follows $2\|\lambda\| \leq \|\lambda + \mu\|$ and therefore by symmetry $\|\lambda\| + \|\mu\| \leq \|\lambda + \mu\|$. This implies $\|\lambda\| + \|\mu\| = \|\lambda + \mu\|$. By the strict convexity of $\mathcal{X}^*$ we get $\lambda = \mu$; see Lemma 2.14. ■

Lemma 3.14. *Let $\mathcal{X}$ be a Banach space with duality mapping J. Then the following conditions are equivalent:*

(a) $\mathcal{X}$ is strictly convex.

(b) *For all* $\lambda \in \mathcal{X}^*\backslash\{\theta\}$ *there exists at most one* $x \in \overline{S}_1$ *with* $\|\lambda\| = \langle\lambda, x\rangle$.

(c) J *is* **strictly monotone**, *i.e.*

$$\langle\lambda - \mu, x - y\rangle > 0 \textit{ for all } x, y \in \mathcal{X}, x \neq y, \lambda \in J(x), \mu \in J(y) .$$

Proof:
(a) $\Longrightarrow$ (b) Let $\lambda \in \mathcal{X}^*\backslash\{\theta\}$. We may assume $\|\lambda\| = 1$. Let $x, y \in \mathcal{X}$ with

$$1 = \|x\| = \|y\|, x \neq y, \langle\lambda, x\rangle = \langle\lambda, y\rangle = 1 .$$

Then $\|x + y\| \geq \langle\lambda, x + y\rangle = 2$ and hence $\mathcal{X}$ cannot be strictly convex.
(b) $\Longrightarrow$ (a) Let $x, y \in \overline{S}_1, x \neq y, \|x + y\| = 2$. Then there exists $\lambda \in \mathcal{X}^*$ with

$$\|\lambda\| = 1 \, , \, 2 = \|x + y\| = \langle\lambda, x + y\rangle = \langle\lambda, x\rangle + \langle\lambda, y\rangle .$$

This implies $\langle\lambda, x\rangle = \langle\lambda, y\rangle = 1$ and condition (b) is violated.
(a) $\Longrightarrow$ (c) Assume that J is not monotone. Then there exist $x, y \in \mathcal{X}, x \neq y$, $\lambda \in J(x), \mu \in J(y)$ with $\langle\lambda - \mu, x - y\rangle = 0$. Then from (3.4) $\|y\| = \|x\| = \|\lambda\| = \|\mu\|$. Since $x \neq y$ this implies $\|x\| = \|y\| \neq 0$. From (3.4) (first line) we obtain

$$l := \|x\|^2 - \|x\|\langle\lambda, y\|y\|^{-1}\rangle = \langle\mu, x\|x\|^{-1}\rangle - \|y\|^2 =: r .$$

Clearly, $l \geq 0$, $r \leq 0$. If $l \neq 0$ then we see that $r > 0$, a case which cannot occur. If $l = 0$ then λ attains its norm on two different points, namely x, y, and hence $\mathcal{X}$ is not strictly convex (see Lemma (3.14)).
(c) $\Longrightarrow$ (a) Assume that $\mathcal{X}$ is not strictly convex. Then there exist $\lambda \in \mathcal{X}^*\backslash\{\theta\}$ and $x, y \in \mathcal{X}, x \neq y$, with

$$\langle\lambda, x\rangle = \|\lambda\| = \langle\lambda, y\rangle \, , \, \|x\| = \|y\| = 1 .$$

We may assume $\|\lambda\| = 1$. Then $\lambda \in J(x), \lambda \in J(y)$ and $\langle\lambda - \lambda, x - y\rangle = 0$. Therefore J is not strictly monotone. ■

Lemma 3.15. *Let* $\mathcal{X}$ *be reflexive and suppose that the spaces* $\mathcal{X}$ *and* $\mathcal{X}^*$ *are strictly convex. Let* J *be the duality map of* $\mathcal{X}$ *and let* J^* *be the duality map of* $\mathcal{X}^*$ *. Then*

$$J : \overline{S}_{1,\mathcal{X}} \longrightarrow \overline{S}_{1,\mathcal{X}^*} \textit{ is bijective with } J^* \circ J = I, J \circ J^* = I .$$

Proof:
Let $\lambda \in \mathcal{X}^*$. If $\lambda = \theta$ then λ is in the range of J since $J(\theta) = \theta$. Assume now $\lambda \neq \theta$ and consider $\mu := \lambda\|\lambda\|^{-1}$. Since $\mathcal{X}$ is reflexive there exists $z \in \mathcal{X}$ with $\|z\| = 1, \langle\lambda, z\rangle = 1$. Then $\|\mu\| = \|z\|, \langle\mu, z\rangle = \|\mu\|\|z\|$ and we conclude that $\mu \in J(z)$. Since $\mathcal{X}^*$ is strictly convex J is single-valued and we have $J(z) = \mu$. With $x := \|\lambda\|z$ we have $J(x) = J(\|\lambda\|z) = \|\lambda\|J(z) = \|\lambda\|\mu = \lambda$. Now, the proof of surjectivity is complete.
Assume: J is not injective. Then there exist $u, v \in \mathcal{X}, u \neq v$, with $\lambda := J(u) = J(v)$. We have $\|\lambda\| = \|u\|, \langle\lambda, u\rangle = \|\lambda\|\|u\|$, $\|\lambda\| = \|v\|, \langle\lambda, v\rangle = \|\lambda\|\|v\|$. We set with $t \in (0, 1)$ $x := tu + (1 - t)v$. Then $\langle\lambda, x\rangle = \|\lambda\|(t\|u\| + (1 - t)\|v\|)$ and therefore $\|\lambda\|(t\|u\| + (1 - t)\|v\|) = |\langle\lambda, x\rangle| \leq \|\lambda\|\|x\|$, $t\|u\| + (1 - t)\|v\| \leq \|x\|$. Obviously, $\|x\| \leq t\|u\| + (1 - t)\|v\|$. Thus, we have

$$\|tu + (1 - t)v\| = t\|u\| + (1 - t)\|v\| .$$

We set $w := u\|u\|^{-1}, z : v\|v\|^{-1}$. Since $\|u\| = \|v\| = \lambda\|$ we have

$$\|w\| = \|z\| = 1, \|tw + (1-t)z\| = t\|w\| + (1-t)\|z\|, w \neq z.$$

This shows that the segment $[w, z]$ is contained in $\overline{S}_{1,\mathcal{X}^*}$. This is contradicting the assumption that $\mathcal{X}^*$ is strictly convex.
Now we have that J is bijective. From the definition of J and J^* we conclude $J^* \circ J = I, J \circ J^* = I$. ■

3.3 Smoothness

Let us begin with a helpful result concerning convex real functions.

Lemma 3.16. *Let* X *be a vector space and let* $g : X \longrightarrow \mathbb{R}$ *be a convex function which satisfies* $g(\theta) = 0$. *Let* $y \in X$. *Then the mapping* $(0, \infty) \ni t \longmapsto g(ty)/t \in \mathbb{R}$ *is monotone non-decreasing.*

Proof:
Let $0 < s \leq t$ and set $a := s/t \in [0, 1]$. Then

$$g(sy)/s = g(aty + (1-a)\theta)/s \leq (ag(ty) + (1-a)g(\theta))/s = g(ty)/t.$$

■

Consider in a normed space $\mathcal{X}$ the **norm-mapping**

$$\nu : \mathcal{X} \ni x \longmapsto \|x\| \in \mathbb{R}, x \in \mathcal{X},$$

and the associated **right-derivative** and **left-derivative** of ν (under the assumption that they exist):

$$\nu'_+(x, y) := \lim_{t \downarrow 0} \frac{\|x + ty\| - \|x\|}{t}, \nu'_-(x, y) := \lim_{t \uparrow 0} \frac{\|x + ty\| - \|x\|}{t}, x, y \in \mathcal{X}.$$

Lemma 3.17. *Let* $\mathcal{X}$ *be a Banach space with norm-mapping* ν. *Let* $x \in \mathcal{X} \backslash \{\theta\}$. *Then we have:*

(1) $\nu'_+(x, y), \nu'_-(x, y)$ *exist for all* $y \in \mathcal{X}$.

(2) $\nu'_-(x, y) = \lim_{t \uparrow 0} \frac{\|x + ty\| - \|x\|}{t} = -\nu'_+(x, -y)$ *for all* $y \in \mathcal{X}$.

(3) $\nu'_+(x, ay) = a\nu'_+(x, y), \nu'_-(x, ay) = a\nu'_-(x, y)$ *for all* $y \in \mathcal{X}, a \geq 0$.

(4) $\nu'_+(x, u + w) \leq \nu'_+(x, u) + \nu'_+(x, w), \nu'_-(x, u + w) \geq \nu'_-(x, u) + \nu'_-(x, w)$, $u, w \in \mathcal{X}$.

(5) $|\nu'_+(x, y)| \leq \|y\|, |\nu'_-(x, y)| \leq \|y\|$ *for all* $y \in \mathcal{X}$.

(6) $\nu'_+(x, x) = \nu'_-(x, x) = 1$.

Proof:
Ad (1) Let $x, y \in \mathcal{X}$. We apply Lemma 3.16 to $g(y) := \|x+y\| - \|x\|$ and obtain that $d_{x,y} : \mathbb{R}\backslash\{0\} \ni t \longmapsto \dfrac{\|x+ty\| - \|x\|}{t}$ is monotone non-decreasing in $(0, \infty)$. In the same way, $d_{x,-y}$ is monotone non-decreasing in $(0, \infty)$. This implies by a simple argumentation that $d_{x,y}$ is monotone non-decreasing in $\mathbb{R}\backslash\{0\}$. We conclude that

$$\nu'_+(x, y) = \lim_{t \downarrow 0} \frac{\|x+ty\| - \|x\|}{t}$$

exists. A similar argumentation shows that $\nu'_-(x, y)$ exists.
Ad (2) See the proof of (1).
Ad $(3), (4)$ This is easy to show.
Ad (5) Follows from

$$-\|y\| \le \nu'_-(x, y) \le \nu'_+(x, y) \le \|y\|, y \in \mathcal{X}.$$

Ad (6) Obviously. ■

Definition 3.18. *Let $\mathcal{X}$ be a Banach space with norm mapping ν. Then the norm in $\mathcal{X}$ is called* **Gateaux differentiable** *in x if $\nu'_+(x, y) = \nu'_-(x, y)$ for all $y \in \mathcal{X}$. If the norm is Gateaux differentiable in x we set $\nu'(x, y) := \nu'_+(x, y), y \in \mathcal{X}$, and we call $\nu'(x, y)$ the* **Gateaux derivative** *of the norm in x in direction y.* □

Clearly, a norm cannot be Gateaux differentiable in θ.

Lemma 3.19. *Let $\mathcal{X}$ be a Banach space with norm-mapping ν and let the norm be Gateaux differentiable in $x \in \mathcal{X}\backslash\{\theta\}$. Then we have:*

(1) $\nu'(x, \cdot)$ *is linear.*

(2) $\|\nu'(x, y)\| \le \|y\|$ *for all* $y \in \mathcal{X}$.

(3) $\nu'(x, \cdot) \in X^*$, $\nu'(x, x) = \|x\|$ *and* $\|\nu'(x, \cdot)\| = 1$.

Proof:
Ad (1) Follows from $(2), (3), (4)$ in Lemma 3.17.
Ad (2) See (5) in Lemma 3.17
Ad (3) We know from $(1), (2)$ that $\nu'(x, \cdot) \in \mathcal{X}^*$ and $\|\nu'(x, \cdot)\| \le 1$. Since $\nu'(x, x) = \|x\|$ we have $\|\nu'(x, \cdot)\| = 1$. ■

Lemma 3.20. *Let $\mathcal{X}$ be a Banach space with norm-mapping ν and let the norm be Gateaux differentiable in $x \in \mathcal{X}\backslash\{\theta\}$. Then we have $\|x\|\nu'(x, \cdot) \in J(x)$.*

Proof:
Due to Lemma 3.19 we have for $x \in \mathcal{X}\backslash\{\theta\}$

$$\nu'(x, \cdot) \in \{\lambda \in \mathcal{X}^* : \|\lambda\| = 1, \langle \lambda, x \rangle = \|x\|\}.$$

■

We set

$$\Sigma(x) := \{\lambda \in \mathcal{X}^* : \langle \lambda, x \rangle = \|x\|, \|\lambda\| = 1\}, \; x \in \mathcal{X}\backslash\{\theta\}.$$

Since for $x \ne \theta$ $\|x\|^{-1}J(x) = \Sigma(x)$ we know that $\Sigma(x)$ is nonempty for all $x \in \mathcal{X}\backslash\{\theta\}$. Above we have seen $\nu'(x, \cdot) \in \Sigma(x)$ if ν is Gateaux differentiable in x.

Lemma 3.21. *Let $\mathcal{X}$ be a Banach space with norm-mapping ν. Let $x \in \overline{S}_1$ and let $\lambda \in \mathcal{X}^*$. Then the following statements are equivalent:*

(a) $\lambda \in \Sigma(x)$.

(b) $\nu'_-(x,y) \le \langle \lambda, y\rangle \le \nu'_+(x,y)$ *for all* $y \in \mathcal{X}$.

Proof:
$(a) \implies (b)$ Let $y \in \mathcal{X}, t > 0$. Then

$$\langle \lambda, y\rangle = \frac{\langle \lambda, x+ty\rangle - \langle \lambda, x\rangle}{t} = \frac{\langle \lambda, x+ty\rangle - \|x\|}{t} \le \frac{\|x+ty\| - \|x\|}{t}$$

and

$$\frac{\|x-ty\| - \|x\|}{-t} = -\frac{\|x+t(-y)\| - \|x\|}{t} \le -\langle \lambda, -y\rangle = \langle \lambda, y\rangle .$$

This implies $\nu'_-(x,y) \le \langle \lambda, y\rangle \le \nu'_+(x,y)$ for all $y \in \mathcal{X}$.
$(b) \implies (a)$ From $(5),(6)$ in Lemma 3.17 we conclude $\lambda \in \Sigma(x)$. ■

Lemma 3.22. *Let $\mathcal{X}$ be a Banach space with norm-mapping ν. Let $x \in \mathcal{X}\backslash\{\theta\}$, $z \in \mathcal{X}$ and $s \in [\nu'_-(x,z), \nu'_+(x,z)]$. Then there exists $\lambda \in \mathcal{X}^*$ with*

$$\nu'_-(x,y) \le \langle \lambda, y\rangle \le \nu'_+(x,y) \text{ for all } y \in \mathcal{X},\ \langle \lambda, z\rangle = s,\ \|\lambda\| = 1,\ \langle \lambda, x\rangle = 1. \tag{3.5}$$

Proof:
We shall construct λ by the Hahn-Banach theorem. Set $\mathcal{Z} := \text{span}(\{z\})$ and define μ on $\mathcal{Z}$ by

$$\langle \mu, az\rangle := as,\ a \in \mathbb{R}.$$

Using the properties $(2),(3)$ in Lemma 3.17 it is easy to prove

$$\nu'_-(x,u) \le \langle \mu, u\rangle \le \nu'_+(x,u) \text{ for all } u \in \mathcal{Z}.$$

We know from (3) in Lemma 3.17 that $\nu'_+(x,\cdot)$ is subadditive, i.e. $\nu'_+(x,u+w) \le \nu'_+(x,u) + \nu'_+(x,w)$ for all $u, w \in \mathcal{X}$. Hence, by the Hahn-Banach Theorem we may extend μ from $\mathcal{Z}$ to the space $\mathcal{X}$ and obtain a linear functional $\lambda \in \mathcal{X}'$ such that

$$\langle \lambda, u\rangle = \langle \mu, u\rangle \text{ for all } u \in \mathcal{Z} \text{ and } \langle \lambda, y\rangle \le \nu'_+(x,y) \text{ for all } y \in \mathcal{X}.$$

Using (2) in Lemma 3.17 we obtain

$$\nu'_-(x,y) \le \langle \lambda, u\rangle \le \nu'_+(x,y) \text{ for all } y \in \mathcal{X}$$

and this implies due to (5) in 3.17

$$|\langle \lambda, y\rangle| \le \|y\| \text{ for all } y \in \mathcal{X}.$$

Therefore $\lambda \in \mathcal{X}^*$ and $\|\lambda\| \le 1$. Since $\nu'_-(x,x) = \nu'_+(x,x) = 1$ we have $\|\lambda\| = 1$. ■

Definition 3.23. *A Banach space $\mathcal{X}$ is called* **Gateaux differentiable** *if the norm is Gateaux-differentiable in each $x \in \mathcal{X}\backslash\{\theta\}$.* □

Clearly, a Hilbert space $\mathcal{H}$ is Gateaux differentiable. The Gateaux derivative $\nu'(x,\cdot)$ of the norm in $x \in \mathcal{H}\backslash\{\theta\}$ is given as follows:

$$\nu'(x,y) = \langle \frac{x}{\|x\|} | y\rangle \, , \, y \in \mathcal{H} \, .$$

Gateaux differentiability of a space is an analytic property of the norm-mapping. Now, we want to show that this property is related to the observation that in an euclidean space a boundary point of the unit sphere allows to draw a unique line through this point which is tangent to the sphere. We know $\nu'(x,\cdot) \in \Sigma(x)$. Each $\lambda \in \Sigma(x)$ defines a family of hyperplanes

$$H_{\lambda,a} := \{z \in \mathcal{X} : \langle \lambda, z\rangle = a\} \, , \, a \in \mathbb{R} \, .$$

The hyperplane $H_{\lambda,a}$ for $a := \|x\|$ touches the ball $\overline{B}_{\|x\|}$ in $\mathcal{X}$ and we say that it supports $\overline{B}_{\|x\|}$ in x.

Definition 3.24. *Let $\mathcal{X}$ be a Banach space.*

(a) $\mathcal{X}$ is **smooth** *at $x \in \overline{S}_1$ (or $x \in \overline{S}_1$ is point of smoothness of $\mathcal{X}$) if $\#\Sigma(x) = 1$.*

(b) $\mathcal{X}$ is **smooth** *if $\mathcal{X}$ is smooth at every $x \in \overline{S}_1$.*

□

Clearly, a Hilbert space $\mathcal{H}$ is smooth since we already know that $\#\Sigma(x) = 1$ for all $x \in \mathcal{H}\backslash\{\theta\}$.

Theorem 3.25. *Let $\mathcal{X}$ be a Banach space. Then we have:*

(a) If $\mathcal{X}^*$ is strictly convex then $\mathcal{X}$ is smooth.

(b) If $\mathcal{X}^*$ is smooth then $\mathcal{X}$ is strictly convex.

Proof:
Ad (a) Assume that $\mathcal{X}$ is not smooth. Then there exists $x^0 \in \overline{S}_1$ and $\lambda, \mu \in \Sigma(x^0)$ with $\lambda \neq \mu$. We obtain

$$\|\lambda\| + \|\mu\| \geq \|\lambda + \mu\| \geq \langle \lambda + \mu, x^0\rangle = 2 = \|\lambda\| + \|\mu\|$$

which implies that $\mathcal{X}^*$ is not strictly convex.
Ad (b) Assume that $\mathcal{X}$ is not strictly convex. Then there exist $x, y \in \overline{S}_1$ so that $\|\frac{1}{2}(x + y)\| = 1$. Choose $\lambda \in \Sigma(\frac{1}{2}(x+y))$. Then

$$1 = \langle \lambda, \frac{1}{2}(x+y)\rangle = \frac{1}{2}\langle \lambda, x\rangle + \frac{1}{2}\langle \lambda, y\rangle \leq \frac{1}{2}(\|x\| + \|y\|) = 1 \, .$$

Therefore we have $\langle \lambda, x\rangle = \langle \lambda, y\rangle = 1$. Consider now x, y as elements in $\mathcal{X}^{**}$. Then we conclude that $\mathcal{X}^*$ is not smooth at λ. ■

Theorem 3.26. *Let $\mathcal{X}$ be a Banach space. Let $x \in \mathcal{X}\backslash\{\theta\}$. Then the following statements are equivalent:*

(a) The duality map is single-valued.

(b) $\mathcal{X}$ is smooth at x.

(c) The norm is Gateaux differentiable in x.

Proof:
$(a) \iff (b)$ This follows from the fact that $\mathcal{X}$ is smooth at $x \in \overline{S_1}$ iff $\#SSigma(x) = 1$.
$(b) \implies (c)$ If the norm is not Gateaux differentiable at x then there exists a $z \in \mathcal{X}$ with $\nu'_-(x,z) < \nu'_+(x,z)$. Then by Lemma 3.22 there exist at least two different functionals $\lambda_1, \lambda_2 \in \Sigma(x)$. Therefore x is not a point of smoothness.
$(c) \implies (b)$ Since $\nu'_-(x,y) = \nu'_+(x,y)$ for all $y \in \mathcal{X}$ the only functional $\lambda \in \Sigma(x)$ is the functional λ with $\langle \lambda, y\rangle = \nu'(x,y) = \nu'_+(x,y)$ for all $y \in \mathcal{X}$. ■

Theorem 3.27. *Let $\mathcal{X}$ be a reflexive Banach space with the duality map J. Then the following conditions are equivalent:*

(a) J is single-valued.

(b) $\mathcal{X}^*$ is strictly convex.

(c) $\mathcal{X}$ is Gateaux differentiable.

(d) $\mathcal{X}$ is smooth.

Proof:
$(a) \implies (b)$ We know $\#\Sigma(x) = 1$ for all $x \in \overline{S}_1$. Let $\lambda, \mu \in \mathcal{X}^*$ with

$$\|\lambda\| = \|\mu\| = 1 \text{ and } \|\lambda + \mu\| = 2\,.$$

Since $\mathcal{X}^*$ is reflexive there exists $x^0 \in \mathcal{X}$ with

$$\|x^0\| = 1 \text{ and } 1 = \|\frac{1}{2}(\lambda+\mu)\| = \langle \frac{1}{2}(\lambda+\mu), x^0\rangle\,.$$

Since $|\langle \lambda, x^0\rangle| \leq 1, |\langle \mu, x^0\rangle| \leq 1$, we obtain $\langle \lambda, x^0\rangle = \langle \mu, x^0\rangle = 1 = \|x^0\|$. This shows $\lambda, \mu \in \Sigma(x^0)$ and therefore $\lambda = \mu$.
$(b) \implies (c)$ Let $x \in \overline{S}_1$. If the norm is not Gateaux differentiable in x then there exists $z \in \mathcal{X}$ with $\nu'_-(x,z) < \nu'_+(x,z)$. Choose s_1, s_2 with

$$\nu'_-(x,z) < s_1 < s_2 < \nu'_+(x,z)\,.$$

By Lemma 3.22 there exist $\lambda_1, \lambda_2 \in \mathcal{X}^*$ with $\langle \lambda_1, z\rangle \neq \langle \lambda_2, z\rangle$, $\|\lambda_1\| = \|\lambda_2\| = 1$, and

$$\nu'_-(x,y) \leq \langle \lambda_1, y\rangle \leq \nu'_+(x,y)\,,\ \nu'_-(x,y) \leq \langle \lambda_2, y\rangle \leq \nu'_+(x,y)\,,\ \text{ for all } y \in \mathcal{X}\,.$$

Then

$$\|\lambda_1 + \lambda_2\| \geq \langle \lambda_1 + \lambda_2, x\rangle = \langle \lambda_1, x\rangle + \langle \lambda_2, x\rangle \geq \nu'_-(x,x) + \nu'_-(x,x) = 2$$

and we conclude that $\mathcal{X}^*$ is not strictly convex by Lemma 3.14.
$(c) \iff (d)$ See Theorem 3.26.
$(d) \implies (a)$ If $\#J(x) = \#\Sigma(x) \geq 2$ for $x \in \overline{S}_1$ then $\mathcal{X}$ cannot be smooth at x.

■

Without reflexivity the implication $(d) \implies (b)$ in Theorem 3.27 is not true. One obtains a counterexample by remorming l_1 such that the resulting space l_1^r is strictly convex and $l_1^* = l_\infty = (l_1^r)^*$; see for instance [14].

3.4 Characterization of best approximations in Banach spaces

Now, we want to derive similar results as in Section 3.1 for the case that the approximation problem is formulated in a Banach space. First of all, we introduce the tangential and normal cone of a nonempty closed convex set.

Definition 3.28. *Let $\mathcal{X}$ be a Banach space, let C be a nonempty closed convex subset of $\mathcal{X}$ and let $z \in C$.*

(a) $T(z, C) := \overline{\cup\{s(u - z) : u \in C, s \geq 0\}}$ *is called the* **tangential cone** *of C at x.*

(b) *If $\mathcal{X}$ is a Hilbert space $\mathcal{H}$ we set $N(z, C) := \{y \in \mathcal{H} : \langle y|u - z\rangle \leq 0$ for all $u \in C\}$ and $N(z, C)$ is called the* **normal cone** *of C at z.*

□

Clearly, $T(z, C), N(z, C)$ are nonempty closed convex cones. One can easily see that $T(z, C)$ is the smallest closed convex cone which contains alle vectors $u - z, u \in C$. The normal cone is the polar cone of the tangential cone. For interior points z of C both cones are not very interesting: $T(z, C) = \mathcal{H}, N(z, C) = \{\theta\}$.

As we know from Kolmogorov's criterion, if $x \in \mathcal{X}$ and $w \in C$ then

$$w = P_C(x) \iff x - w \in N(w, C) \tag{3.6}$$

This form of Kolmogorov's criterion will be now transfered to the Banach space case. The way we do that is to use optimality conditions for convex programs.

Let $\mathcal{X}$ be a Banach space, let C be a nonempty closed convex subset of $\mathcal{X}$ and let $x \in \mathcal{X}$. The approximation problem may be reformulated as follows:

$$\text{Minimize } (f + \delta_C)(u) \text{ subject to } u \in \mathcal{X}. \tag{3.7}$$

Here $f : \mathcal{X} \ni y \longmapsto \|x - y\| \in \mathbb{R}$ and

$$\delta_C(y) := \begin{cases} 0 & \text{if } y \in C \\ \infty & \text{if } y \notin C \end{cases}.$$

Of course, $f : \mathcal{X} \longrightarrow \mathbb{R}$ is a convex function and $\delta_C : \mathcal{X} \longrightarrow \overline{\mathbb{R}} := \mathbb{R} \cup \{\infty\}$ is convex function too, since both functions have an epigraph which is a convex subset on $\mathcal{X} \times \mathbb{R}$. Set $G := f + \delta_C$. G is a convex function too. A supporting hyperplane of G in a point $(y^0, G(y^0)) \in \text{epi}(G)$ with $G(y^0) < \infty$ is the graph of a mapping

$$\mathcal{X} \ni y \longmapsto \langle \lambda, y - y^0\rangle + G(y^0) \in \mathbb{R}\ (\lambda \neq \theta)$$

such that

$$\text{epi}(G) \subset H_+ := \{(y, s) : \langle \lambda, y - y^0\rangle + G(x^0) \leq s\}.$$

This leads us to the subdifferential of a convex function.

Definition 3.29. *Let $\mathcal{X}$ be a Banach space and let $F : \mathcal{X} \longrightarrow \overline{\mathbb{R}}$ convex. If $y^0 \in$ dom$(F) := \{y : F(y) < \infty\}$ then the set*

$$\partial F(y^0) := \{\lambda \in X^* : \langle \lambda, y - y^0 \rangle + F(y^0) \le F(y) \text{ for all } y \in \mathcal{X} \tag{3.8}$$

is called the **subdifferential of** F *in* y^0. □

The subdifferential induces a set-valued mapping

$$\partial F : \mathcal{X} \rightrightarrows \mathcal{X}^* .$$

If F is continuous in a point $y^0 \in \text{dom}(G)$ then $\partial G(y^0)$ is nonempty; see Theorem 10.44 in Chapter 10.

Let us come back to the minimization problem (3.7). Suppose that $w \in C$ belongs to dom(g), $g := f + \delta_C$. Then we have the obvious observation that

$$w \in P_C(x) \text{ holds true iff } \theta \in \partial g(w) . \tag{3.9}$$

Now the application of the subdifferential formula

$$\partial(f + \delta_C)(w) = \partial f(w) + \partial \delta_C(w) \tag{3.10}$$

is allowed since $w \in \text{dom}(f) \cap \text{dom}(\delta_C)$ and f is continuous in w; see Theorem 10.49 in Chapter 10. Consequentely, if $x \in \mathcal{X}$ and $w \in C$ then

$$w \in P_C(x) \text{ holds true iff } -\partial f(w) \cap \partial \delta_C(w) \ne \emptyset . \tag{3.11}$$

Next we analyze $\partial f(w), \partial \delta_C(w)$.

Lemma 3.30. *Let $\mathcal{X}$ be a Banach space and let C be a nonempty closed convex subset of $\mathcal{X}$. Let $w \in C$. Then*

$$\partial \delta_C(w) = N(w, C) \tag{3.12}$$

where $N(w, C) := \{\mu \in X^* : \langle \mu, z \rangle \le 0 \text{ for all } z \in T(w, C)\}$ *is the so called* **normal cone** *of* C *at* w.

Proof:
Suppose $\lambda \in \partial \delta_C(w)$. Then

$$\langle \lambda, y - w \rangle \le \delta_C(y) - \delta_C(w) \text{ for all } y \in \mathcal{X} .$$

This implies $\langle \lambda, u - w \rangle \le 0$ for all $u \in C$ and hence, $\lambda \in N(w, C)$. ■

Lemma 3.31. *Let $\mathcal{X}$ be a Banach space, let $x \in \mathcal{X}$ and let $f : \mathcal{X} \ni y \longmapsto \|x - y\| \in \mathbb{R}$. If $x - w \ne \theta$ then*

$$\partial f(w) = \Sigma(x - w) . \tag{3.13}$$

Proof:
Clearly, $\partial f(w)$ is nonempty since f is convex and f is continuous in w; see Theorem 10.47 in Chapter 10. Suppose $\mu \in \partial f(w)$. Then

$$\langle \mu, h \rangle \le \|x - (w + h)\| - \|x - w\| \text{ for all } h \in \mathcal{X} . \tag{3.14}$$

Clearly,

$$\langle \mu, h \rangle \le \|x - w\| + \|h\| - \|x - w\| = \|h\| \text{ for all } h \in \mathcal{X}$$

and

$$\langle \mu, x - w \rangle \le \|x - x\| - \|x - w\|$$

which implies $\|\mu\| = 1$ since $x - w \neq \theta$. Setting $h := w - x$ in (10.27) we obtain

$$\langle \mu, w - x \rangle \le \|x - (2w - x)\| - \|x - w\| = \|x - w\|$$

and we obtain for $\rho := -\mu$

$$\rho \in X^*, \|\rho\| = 1, \langle \rho, x - w \rangle = \|x - w\|, \text{i.e. } \rho \in \Sigma(x)\,. \tag{3.15}$$

Conversely, if $\lambda \in \Sigma(x - w)$, then

$$\|x-(w+h)\|-\|x-w\| = \|x-(w+h)\|-\langle \lambda, x-w \rangle \ge \langle \lambda, x-(w+h) \rangle - \langle \lambda, x-w \rangle = \langle \lambda, h \rangle$$

for all $h \in \mathcal{X}$. This shows $\lambda \in \partial f(w)$. ■

Theorem 3.32. *Let $\mathcal{X}$ be a Banach space and let C be a nonempty closed convex subset of $\mathcal{X}$. Let $x \in \mathcal{X}, w \in C$. Then the following conditions are equivalent:*

(a) $w \in P_C(x)$.

(b) $J(x - w) \cap N(w, C) \neq \emptyset$.

Proof:
If $x - w = \theta$ then $x \in P_C(x)$ and $\lambda := \theta \in J(x - w) \cap N(w, C)$. If $x - w \neq \theta$ then we have shown above $w \in P_C(x)$ iff $\Sigma(x-w) \cap N(w, C) \neq \emptyset$. But as we know, $J(x-w) = \{\|x-w\|\mu : \mu \in \Sigma(x - w)\}$ and therefore $J(x - w) \cap N(w, C) \neq \emptyset$ iff $\Sigma(x - w) \cap N(x - w, C) \neq \emptyset$. ■

In Lemma 3.20 we have shown that in a Banach space $\mathcal{X}$ the Gateaux dervative $\nu'(x, \cdot)$ at the point $x \in \mathcal{X}\backslash\{\theta\}$ belongs to $\Sigma(x)$ (when it exists). Hence, $\|x\|\nu'(x, \cdot) \in J(x)$.

Theorem 3.33 (Asplund, 1966)**.** *Let $\mathcal{X}$ be a Banach space and let $j : \mathcal{X} \ni x \longmapsto \frac{1}{2}\|x\|^2 \in \mathbb{R}$. Then*

$$\partial j(x) = J(x) \textit{ for all } x \in \mathcal{X}\,. \tag{3.16}$$

Proof:
We follow [1]. Notice that $\partial j(x)$ is nonempty for all $x \in \mathcal{X}$ since j is continuous. Moreover, for $x = \theta$ nothing has to be proved. Let $x \in \mathcal{X}\backslash\{\theta\}$.
Let $\lambda_x \in \partial j(x)$. Choose any $y \in \overline{S}_{\|x\|}$. Then $\langle \lambda_x, y - x \rangle \le 0$, i.e. $\langle \lambda_x, y \rangle \le \langle \lambda_x, x \rangle$. This implies

$$\|\lambda_x\|\|x\| = \sup_{y \in \overline{S}_{\|x\|}} \langle \lambda_x, y \rangle \le \langle \lambda_x, x \rangle\,.$$

Hence, we have shown

$$\|\lambda_x\|\|x\| = \langle \lambda_x, x \rangle \text{ for all } \lambda_x \in \partial j(x), \text{ for all } x \in \mathcal{X}\backslash\{\theta\}. \tag{3.17}$$

Let $z \in \overline{S}_1$ and let $t > 0, s > 0$. Let $\lambda_{tz} \in \partial j(tz)$. Then using (3.17) we obtain

$$\begin{aligned}
\frac{1}{2}t^2 - \frac{1}{2}s^2 &\geq \langle \lambda_{sz}, tz\rangle - \|\lambda_{sz}\|\|sz\| \\
&= \langle \lambda_{sz}, tz\rangle - \|\lambda_{sz}\|\|sz\| + \|\lambda_{sz}\|\|tz\| - \|\lambda_{sz}\|\|tz\| \\
&= \|\lambda_{sz}\|(t-s) + (\langle \lambda_{sz}, tz\rangle - \|\lambda_{sz}\|\|tz\|) \\
&= \|\lambda_{sz}\|(t-s) + t(\langle \lambda_{sz}, sz\rangle - \|\lambda_{sz}\|\|sz\|)/s \\
&\geq \|\lambda_{sz}\|(t-s)
\end{aligned}$$

This shows the property $\lambda_w \in J(w)$ on the ray from θ through z. From this we conclude that $\lambda_x \in J(x)$ for all $x \in J(x)$.
Conversely, suppose that $\lambda_x \in J(x)$. Then by the definition of $J(x)$

$$\frac{1}{2}\|x\|^2 + \langle \lambda_x, y - x\rangle \leq \frac{1}{2}\|x\|^2 + \|\lambda_x\|(\|y\| - \|x\|)$$

and

$$\frac{1}{2}\|x\|^2 + \|\lambda_x\|(\|y\| - \|x\|) \leq \frac{1}{2}\|x\|^2 + \|\lambda_x\|\|y\| - \|x\|^2 \leq \frac{1}{2}\|x\|^2 + \frac{1}{2}\|\lambda_x\|^2 + \frac{1}{2}\|y\|^2 - \|x\|^2 = \frac{1}{2}\|y\|^2 .$$

This shows $\lambda_x \in \partial j(x)$. ■

Now, we know the subdifferential of the norm mapping $\nu : \mathcal{X} \ni x \longmapsto \|x\| \in \mathbb{R}$ and $j : \mathcal{X} \ni x \longmapsto \frac{1}{2}\|x\|^2 \in \mathbb{R}$:

$$\partial\nu(\theta) = \overline{B}_1 \tag{3.18}$$
$$\partial\nu(x) = \Sigma(x) \text{ if } x \neq \theta \tag{3.19}$$
$$\partial j(x) = J(x) \tag{3.20}$$

The assertion (3.18) follows from the observation

$$\langle \lambda, y - \theta\rangle \leq \nu(y - \theta) - \nu(\theta) = \|y\|, y \in \mathcal{X}$$

for all $\lambda \in \partial\nu(\theta)$.

3.5 Semi-inner products and characterization of best approximations

The characterization results in Hilbert spaces and Banach spaces look very different. In this section we want to bring together these results by using semi-inner products in Banach spaces.[3] This can be done by introducing the concept of semi-inner products defined by Lumer [13], Giles [10] and used in approximation problems by Penot [15]. Here we follow mainly Dragomir [9]; see also Rosca [16].

[3] Again, remember that we discuss our results strictly under the completeness assumption. Since we discuss the question of characterization of best approximations it would be sufficient to consider normed spaces only.

Definition 3.34. *Let* X *be a real vector space*[4]. *A mapping* $[\cdot|\cdot] : X \times X \longrightarrow \mathbb{R}$ *is called a* **semi-inner product** *if the following properties are satisfied:*

(1) $[x + y|z] = [x|z] + [y|z]$ *for all* $x, y, z \in X$;

(2) $[ax|y] = a[x|y], [x|ay] = a[x|y]$ *for all* $x, y \in X, a \in \mathbb{R}$;

(3) $[x|x] \geq 0$ *for all* $x \in X$ *and* $[x|x] = 0$ *implies* $x = \theta$;

(4) $|[x|y]|^2 \leq [x|x]\,[y|y]$ *for all* $x, y \in X$.

$\square$

Corollary 3.35. *Let* X *be a real vector space and* $[\cdot|\cdot] : X \times X \longrightarrow \mathbb{R}$ *be a semi-inner product on* X. *Then the mapping* $\|\cdot\| : X \longrightarrow \mathbb{R}$ *defined by*

$$\|x\| := [x|x]^{\frac{1}{2}}, \; x \in X,$$

is a norm on X.

Proof:
This can easily be shown. For convenience we verify the triangle inequality. Let $x, y \in X$ with $\|x + y\| \neq 0$; the case $\|x + y\| = 0$ is trivial. Then

$$\|x + y\|^2 = [x + y|x + y] = [x|x + y] + [y|x + y] \leq \|x\|\|x + y\| + \|y\|\|x + y\|$$

and dividing by $\|x + y\|$ the triangle inequality follows. $\blacksquare$

Corollary 3.36. *Let* X *be a real vector space and* $[\cdot|\cdot] : X \times X \longrightarrow \mathbb{R}$ *be a semi-inner product on* X. *Let* $x \in \mathcal{X}$. *Then the mapping* $\lambda : \mathcal{X} \longrightarrow \mathbb{R}$, *defined by*

$$\lambda(y) := [y|x] \in \mathbb{R}$$

is a functional in $\mathcal{X}^*$ *with* $\|\lambda\| = \|x\|$ *when* X *is endowed with the norm* $[\cdot|\cdot|^{\frac{1}{2}}$ *(see Corollary 3.35.*

Proof:
The linearity follows from the linearity of $[\cdot|\cdot]$ in the first argument. From $|\lambda(y)| = |[y|x]| \leq [x|x]^{\frac{1}{2}}[y|y]^{\frac{1}{2}}$ we conclude $\|\lambda\| \leq [x|x]^{\frac{1}{2}} = \|x\|$ and hence, λ is continuous. On the other hand, we have $[x|x] = |\lambda(x)| \leq \|\lambda\|[x|x]^{\frac{1}{2}}$, i.e. $\|\lambda\| \geq [x|x]^{\frac{1}{2}} = \|x\|$. $\blacksquare$

Definition 3.37. *Let* X *be a Banach space with norm* $\|\cdot\|$ *and endowed with a semi-inner product* $[\cdot|\cdot]$. *Then we say that* $[\cdot|\cdot]$ *generates the norm if* $\|x\| = [x|x]^{\frac{1}{2}}$ *for all* $x \in \mathcal{X}$. $\square$

Definition 3.38. *Let* $\mathcal{X}$ *be a Banach space and let* $J : X \rightrightarrows X^*$ *be the duality map. Then a mapping* $\tilde{J} : X \longrightarrow X^*$ *with* $\tilde{J}(x) \in J(x), x \in \mathcal{X}$, *is called a* **section of** J. *Such a section* $\tilde{J}$ *is called a* **homogeneous section** *if the mapping* $x \longmapsto \tilde{J}(x)$ *is homogeneous, i.e.* $\tilde{J}(ax) = a\tilde{J}(x), x \in \mathcal{X}, a \in \mathbb{R}$. $\square$

[4]Semi-inner products can be defined for vector spaces with complex scalars. The homogenity property has to be adapted.

Lemma 3.39. *Let* X *be a Banach space with duality map* J *. Then the following statements hold:*

(a) *The duality map allows homogeneous sections.*

(b) *The duality map has a uniquely determined homogeneous section if and only if* J *is single-valued.*

(c) *Every semi-inner product* $[\cdot|\cdot]$ *on* $\mathcal{X}$ *which generates the norm in* $\mathcal{X}$ *is of the form*

$$[x|y] = \langle\tilde{J}(y), x\rangle \text{ for all } x, y \in \mathcal{X}, \tag{3.21}$$

where $\tilde{J}$ *is a homogenous section of the duality map* J *on* $\mathcal{X}$ *.*

Proof:
Ad (a) Let $z \in \overline{S}_1$. Choose (exactly one) $\lambda_z \in X^*$ with $\|\lambda_z\| = 1, \langle\lambda_z, z\rangle = 1$. Then $\lambda_z \in \Sigma(z)$. Now, if $x = az, a \in \mathbb{R}, z \in \overline{S}_1$, set $\lambda_x := a\lambda_z$.

Let $x, y \in \mathcal{X}, x \neq \theta, r \in \mathbb{R}, r \neq 0$. Then for $r > 0$

$$\langle r\lambda_x, y\rangle = \langle r\|x\|\lambda_{x\|x\|^{-1}}, y\rangle = \langle\|x\|\lambda_{rx\|rx\|^{-1}}, y\rangle = \langle\lambda_{rx}, y\rangle$$

and for $r > 0$

$$\begin{aligned}\langle r\lambda_x, y\rangle &= \langle(-r)\lambda_x, -y\rangle = \langle\lambda_{-rx}, -y\rangle \\ &= \langle\|r(-x)\|\lambda_{r(-x)\|r(-x)\|^{-1}}, -y\rangle = \langle r\|x\|\lambda_{x\|x\|^{-1}}, y\rangle = \langle r\lambda_x, y\rangle .\end{aligned}$$

Moreover,

$$\langle\lambda_x, y\rangle = \langle\|x\|\lambda_{x\|x\|^{-1}}, y\rangle \leq \|x\|\|y\| , \langle\lambda_x, x\rangle = \langle\|x\|\lambda_{x\|x\|^{-1}}, x\rangle = \|x\|^2 .$$

Therfore, a homogeneous section $\tilde{J}$ of J is defined by setting

$$\tilde{J}(x) = \lambda_x , \ x \in \mathcal{X} .$$

Ad (b) From the proof in (a) it is clear that homogeneous sections are not uniquely determined if $\#J(x) \geq 2$ for some $x \in \mathcal{X}\backslash\{\theta\}$.
Ad (c) Let $\tilde{J}$ be a homogeneous section of J . Define the mapping

$$[\cdot|\cdot] : \mathcal{X} \times \mathcal{X} \longrightarrow \mathbb{R} , \ [x|y] := \langle\tilde{J}(y), x\rangle .$$

Then $|\cdot|\cdot]$ is linear in the first argument and homogeneous in each argument. We have

$$[x|x] = \langle\tilde{J}(x), x\rangle = \|x\|^2 = \|\tilde{J}(x)\|\|x\| \geq 0, x \in \mathcal{X},$$

and therefore $[x, x] = 0$ iff $x = \theta$. Since

$$|[x|y]|^2 = |\langle\tilde{J}(y), x\rangle|^2 \leq \|\tilde{J}(y)\|^2\|x\|^2 = \|y\|^2\|x\|^2 = [x|x]|y|y]$$

$[\cdot|\cdot]$ is a semi-inner product which generates the norm.
Conversely, let $[\cdot|\cdot]$ be a semi-inner product on $\mathcal{X}$ which generates the norm of $\mathcal{X}$. Define $\tilde{J} : \mathcal{X} \longrightarrow \mathcal{X}^*$ by

$$\langle\tilde{J}(y), x\rangle := [x|y] , \ x \in \mathcal{X} .$$

Then

$$\langle \tilde{J}(x), x\rangle = [x|x] = \|x\|^2\,,\ \|\tilde{J}(y)\| = \|x\|, x, y \in \mathcal{X},$$

Consequentely, we have

$$\langle \tilde{J}(x), x\rangle = \|x\|^2 \text{ and } \|\tilde{J}(x)\| = \|x\|\,.$$

Moreover $y \longmapsto \tilde{J}(y)$ is homogeneous section. ∎

Lemma 3.39 says that every Banach space can be endowed with a semi-inner product.

Corollary 3.40. *Let $\mathcal{X}$ be a smooth Banach space. Then the duality map is single-valued and the uniquely determined semi-inner product which generates the norm is given by*

$$[x, y] := \langle J(y), x\rangle\,,\ x, y \in \mathcal{X}\,. \tag{3.22}$$

Proof:
We know that in a smooth Banach space the duality map is single-valued; see Theorem 3.27. Therefore by Lemma 3.39 the semi-inner product is uniquely determined. ∎

Example 3.41. *Consider the Banach space $\mathcal{X} := \mathbb{R}^3$ endowed with the l_1-norm. It is easy to check that by setting*

$$[x|y] := \|y\|_1 \sum_{k=1, y_k \neq 0}^{3} \frac{x_k y_k}{|y_k|}\,,\ x = (x_1, x_2, x_3), y = (y_1, y_2, y_3)\,.$$

a semi-inner product in $\mathcal{X}$ is defined which generates the l_1 norm in $\mathbb{R}^3$. □

With the help of a semi-inner product which generates the norm in a Banach space we can reformulate the property of strictly convexity.

Theorem 3.42. *Let $\mathcal{X}$ be a Banach space and let $[\cdot|\cdot]$ be a semi-inner product which generates the norm in $\mathcal{X}$. Then the following statements are equivalent:*

(a) *$\mathcal{X}$ is strictly convex.*

(b) *If $x, y \in \mathcal{X}\backslash\{\theta\}$ with $[x|y] = \|x\|\|y\|$ then there exists $a > 0$ with $x = ay$.*

Proof:
$(a) \Longrightarrow (b)$ Let $\tilde{J}$ be a homogeneous section of J which generates the semi-inner product. Let $x, y \in \mathcal{X}\backslash\{\theta\}$ with $[x|y] = \|x\|\|y\|$. Then $[x|y] = \langle \tilde{J}(y), x\rangle = \|x\|\|y\|$. This implies

$$\langle \tilde{J}(y), \frac{x}{\|x\|}\rangle = \|y\| = \|\tilde{J}(y)\|\,,\ \langle \tilde{J}(y), \frac{y}{\|y\|}\rangle = \|y\| = \|\tilde{J}(y)\|\,.$$

Since $\mathcal{X}$ is strictly convex, every continuous linear functional attains its norm on at most one point; see Lemma 3.14. This implies

$$\frac{x}{\|x\|} = \frac{y}{\|y\|}$$

and we have $x = ay$ with $a = \|x\|\|y\|^{-1}$.
(b) $\Longrightarrow$ (a) Let $x, y \in \mathcal{X}\backslash\{\theta\}$ with $\|x+y\| = \|x\| + \|y\|$. Then

$$[x|x+y] \leq \|x\|\|x+y\|\,,\ [y|x+y] \leq \|y\|\|x+y\|\,. \tag{3.23}$$

Suppose that in (3.23) one of inequalities above is strict. Then, by addition, we obtain

$$\|x+y\|^2 = [x+y|x+y] = [x|x+y] + [y|x+y] < (\|x\| + \|y\|)\|x+y\| = \|x+y\|^2$$

which is a contradiction. Therefore, $[x|x+y] = \|x\|\|x+y\|$ and by (b) we obtain $x = a(x+y)$ with $a > 0$. Since obviously $a \neq 0$ we have $x = sy$ with $s = (1-a)^{-1}$. ∎

Lemma 3.43. *Let $\mathcal{X}$ be a Banach space and let $[\cdot|\cdot]$ be a semi-inner product which generates the norm in $\mathcal{X}$. Then for all $x, y \in \mathcal{X}, t > 0, x \neq \theta, x + ty \neq \theta$,*

$$\frac{[y|x]}{\|x\|} \leq \frac{\|x+ty\| - \|x\|}{t} \leq \frac{[y|x+ty]}{\|x+ty\|} \tag{3.24}$$

Proof:
Let $\tilde{J}$ be a homogeneous section which generates $[\cdot|\cdot]$:

$$[x|y] = \langle \tilde{J}(y), x\rangle, x, y \in \mathcal{X}\,.$$

Let $x, y \in \mathcal{X}, t > 0, x \neq \theta, x + ty \neq \theta$.

$$\begin{aligned}
\frac{\|x+ty\| - \|x\|}{t} &= \frac{\|x+ty\|\|x\| - \|x\|^2}{t\|x\|} \geq \frac{\langle \tilde{J}(x), x+ty\rangle - \|x\|^2}{t\|x\|} \\
&= \frac{\langle \tilde{J}(x), y\rangle}{\|x\|} = \frac{[y|x]}{\|x\|}
\end{aligned}$$

and

$$\begin{aligned}
\frac{\|x+ty\| - \|x\|}{t} &= \frac{\|x+ty\|^2 - \|x\|\|x+ty\|}{t\|x+ty\|} \\
&\leq \frac{\|x+ty\|^2 - \langle \tilde{J}(x+ty), x\rangle}{t\|x+ty\|} \\
&= \frac{\|x+ty\|^2 - \langle \tilde{J}(x+ty), x+ty\rangle + \langle \tilde{J}(x+ty), ty\rangle}{t\|x+ty\|} \\
&= \frac{\langle \tilde{J}(x+ty), y\rangle}{\|x+ty\|} = \frac{[y|x+ty]}{\|x+ty\|}
\end{aligned}$$

∎

Theorem 3.44. *Let $\mathcal{X}$ be a Banach space and let $[\cdot|\cdot]$ be a semi-inner product which generates the norm in $\mathcal{X}$. Then the following statements are equivalent:*

(a) $\mathcal{X}$ is smooth.

(b) $\mathcal{X}$ is Gateaux-differentiable.

(c) The semi-inner product $[\cdot|\cdot]$ which generates the norm in $\mathcal{X}$ is uniquely determined.

(d) *The duality map of* $\mathcal{X}$ *is single-valued.*

(e) $\lim_{t\to 0}[y|x+ty]$ *exists for all* $x \in \mathcal{X}\backslash\{\theta\}, y \in \mathcal{X}$.

Additionally, if one of the above statements is true, the limit in (e) *is given as* $[y,x]$.

Proof:
Let J be the duality map of $\mathcal{X}$ and let Let $\tilde{J}$ be a homogeneous section which generates $[\cdot|\cdot]$.
$(a) \iff (b)$ Theorem 3.26.
$(a) \implies (c)$ If $\mathcal{X}$ is smooth, $\#J(x) = 1$ for all $x \in \mathcal{X}$; see Theorem 3.26. Therefore there exists exactly one homogeneous section of J which generates the semi-inner product.
$(c) \implies (d)$ Obviously, there exists only one homogeneous section of J and by Lemma 3.39, $\#J(x) = 1$ for all $x \in \mathcal{X}$.
$(d) \implies (e)$ Let $x \in \mathcal{X}\backslash\{\theta\}, y \in \mathcal{X}$.
Let $(t_n)_{n\in\mathbb{N}}$ be a sequence in $(0,\infty)$ with $\lim_n t_n = 0$. Then $(\tilde{J}((x+t_n y)\|x+t_n y\|^{-1}))_{n\in\mathbb{N}}$ is a sequence in the unit ball of $\mathcal{X}^*$. By Lemma 3.12 $\lim_{t\downarrow 0}\tilde{J}((x+t_n y)\|x+t_n y\|^{-1})) = \tilde{J}(x\|x\|^{-1})$. This shows

$$\lim_{t\downarrow 0}\tilde{J}((x+ty\|x+ty\|^{-1})) = \tilde{J}(x\|x\|^{-1})$$

which implies $\lim_{t\to 0}[y|x+ty] = [y|x]$.
$(e) \implies (b)$ We conclude from Lemma 3.43 that $\nu'_-(x,y) = \nu'_+(x,y)$ which shows that ν is Gateaux differentiable. ■

Definition 3.45. *Let* $\mathcal{X}$ *be a Banach space and let* $[\cdot|\cdot] : X \times X \longrightarrow \mathbb{R}$ *be a semi-inner product on* X . *The semi-inner product is called* **continuous** *iff*

$$\lim_{t\to 0}[x|y+tx] = [x|y] \text{ for all } x, y \in \overline{S}_1 .$$

□

Theorem 3.46. *Let* $\mathcal{X}$ *be a Banach space and let* $[\cdot|\cdot] : X \times X \longrightarrow \mathbb{R}$ *be a semi-inner product on* X *which generates the norm. Let* C *be a nonempty closed convex set and* $x \in \mathcal{X}, w \in C$. *Then the following statements are equivalent:*

(a) $w \in P_C(x)$.

(b) $[(u-w|x-w-t(u-w)] \le 0$ *for all* $u \in C$ *and* $t \in [0,1]$.

(c) $[u-w|x-w] \le 0$ *for all* $u \in C$.

Proof:
$(a) \implies (b)$ Assume by contradiction, $[(u-w|x-w-t(u-w)] > 0$ for some $u \in C$ and $t \in [0,1]$. Clearly, $u \ne w$. Set $z := tu + (1-t)w$. Then $z \in C$ and we have

$$\begin{aligned}
\|x-z\| &= \|x-w-t(u-w)\|^{-1}[x-w-t(u-w)|x-w-t(u-w)] \\
&= \|x-w-t(u-w)\|^{-1}[x-w-t(u-w)|x-w-t(u-w)] \\
&\quad -t[u-w|x-w-t(u-w)] \\
&< \|x-w-t(u-w)\|^{-1}[x-w|x-w-t(u-w)] \\
&\le \|x-w\|
\end{aligned}$$

This is a contradiction to the best approximation property of w.
$(b) \implies (c)$ Take $t := 0$ in (b).
$(c) \implies (a)$ We have for all $u \in C$:

$$\begin{aligned} \|x-w\|^2 &= [x-w|x-w] \\ &= [x-u|x-w] + [u-w|x-w] \\ &\le [x-u|x-w] \\ &\le \|x-u\|\|x-w\| \end{aligned}$$

We conclude $\|x-w\| \le \|x-u\|$ for all $u \in C$. Hence, $w \in P_C(x)$ and since C is a Chebyshev set $w = P_C(x)$.

■

Definition 3.47. *Let $\mathcal{X}$ be a Banach space and let $[\cdot|\cdot] : X \times X \longrightarrow \mathbb{R}$ be a semi-inner product on X. Let K be a nonempty subset of $\mathcal{X}$. Then we set*

$$K^\circ := \{x \in \mathcal{X} : [y|x] \le 0 \text{ for all } y \in K\}$$

and we call K° the **dual cone** *(relative to $[\cdot|\cdot]$) of K.* □

Theorem 3.48. *Let $\mathcal{X}$ be a Banach space and let $[\cdot|\cdot] : X \times X \longrightarrow \mathbb{R}$ be a semi-inner product on X which generates the norm. Let C be a nonempty closed convex set and $x \in \mathcal{X}, w \in C$. Then the following statements are equivalent:*

(a) $w \in P_C(x)$.

(b) $x - w - t(y-w) \in (C-w)^\circ$ *for all* $y \in C$ *and* $t \in [0,1]$.

(c) $x - w \in (C_w)^\circ$.

Proof:
Obviously with Theorem 3.46. ■

Let $\mathcal{X}$ be a Banach space with norm $\|\cdot\|$. Since $j : \mathcal{X} \ni x \longmapsto \frac{1}{2}\|x\| \in \mathbb{R}$ is a convex function the following limits exist:

$$[x,y]_+ := \lim_{t\downarrow 0} \frac{\|y+tx\|^2 - \|y\|^2}{2t}, \; [x,y]_- := \lim_{t\uparrow 0} \frac{\|y+tx\|^2 - \|y\|^2}{2t}; \tag{3.25}$$

apply Lemma 3.16. The binary functions $[\cdot|\cdot]_+, [\cdot|\cdot]_-$ are called the **superior semi-inner product** and the **inferior semi-inner product** associated to the norm in $\mathcal{X}$ respectively. It is not difficult to see that for all $x, y \in \mathcal{X}, t > 0, s < 0$ we have

$$\lim_{t\downarrow 0} \frac{\|y+tx\|^2 - \|y\|^2}{2t} \ge [x,y]_+ \ge [x,y]_- \ge \lim_{s\uparrow 0} \frac{\|y+sx\|^2 - \|y\|^2}{2s}. \tag{3.26}$$

For details see [9].

Theorem 3.49. *Let $\mathcal{X}$ be a Banach space with norm $\|\cdot\|$. Then we have for all $x, y \in \mathcal{X}$:*

(a) $[x,y]_+ = \sup\{\langle\lambda, y\rangle : \lambda \in J(y)\}$.

(b) $[x,y]_- = \inf\{\langle\lambda, y\rangle : \lambda \in J(y)\}$.

Proof:
See Dragomir [9]. ■

$[\cdot,\cdot]_+, [\cdot,\cdot]_-$ may be use to develop Kolmogorov-like results for best approximations in Banach spaces. This is done in [15].

A semi-inner product $[\cdot|\cdot]$ in a real vector space X max be used to introduce an orthogonality relation via „x orthogonal to y“ if and only if $[y|x] = 0$. This is not a symmetric relation and transitivity of the orthogonality of vectors is not guaranteed in general.

Example 3.50. *Consider the space $\mathbb{R}^3$ endowed with the l_1-norm; see Example 3.41. Then the semi-inner product*

$$[x|y] := \|y\| \sum_{i=1, y_i \neq 0}^{3} |y_i|^{-1} x_i y_i$$

generates the norm. Consider $x := (-2, 1, 0), y := (1, 1, 0)$. Then we have $[y|x] = 0$, $[x|y] = -2$ and we see that x is orthogonal to y but y is not orthogonal to x. □

There are other various concepts of orthogonality in normed spaces. In general, they are modeled along a property which holds in inner product spaces. Let $(X, \langle\cdot|\cdot\rangle)$ be an inner product space. Here are three properties in an inner product space which may serve as a model for orthogonality in normed spaces:

(1) $x \perp y \iff \langle x|y\rangle = 0$.

(2) $x \perp y \iff \|x+y\| = \|x-y\|$

(3) $x \perp y \iff \|x\| \leq \|x+ty\|$ for all $t \in \mathbb{R}$.

The orthogonality via (1) can be generalized to Banach spaces by saying

$$x \text{ is orthogonal to } y \text{ iff } \langle J(y), x\rangle = 0 .$$

where we suppose that J is single-valued. This is the concept which corresponds to semi-inner products; see above. We denote this orthogonality as **J-orthogonality**.

Orthogonality in the **sense of Birkhoff** in normed spaces is derived from (3):

$$x \text{ is orthogonal to } y \text{ iff } \|x\| \leq \|x+ty\| \text{ for all } t \in \mathbb{R} .$$

In general, Birkhoff-orthogonality is not a symmetric property too, i.e. $x \perp y$ does not imply $y \perp x$. Nevertheless, Birkhoff-orthogonality may be used to generalize orthogonal projections from Hilbert spaces to general Banach spaces; see for instance [9, 12].

3.6 Appendix: Convexity II

We want to consider the following optimization problem:

$$\text{Minimize } (f+g)(x)\,,\ x \in X\,. \tag{3.27}$$

Here X is a Banach space and $f, g : X \longrightarrow \hat{\mathbb{R}}$ are proper closed convex functions. In order to associate to this problem a dual problem we consider the process of conjugation for convex functions.

Definition 3.51. *Let* X *be a Banach space and let* $f : X \longrightarrow \hat{\mathbb{R}}$ *be a convex function. Then we set*

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

and call f^* *the* **conjugate of** f. □

Conjugation defines a convex function $f^* : X^* \longrightarrow \bar{\mathbb{R}}$. It is obvious that $f^* : X \longrightarrow \hat{\mathbb{R}}$ if f is proper. An immediate consequence of the definition is the **Fenchel-Young inequality**

$$f(x) + f^*(\lambda) \geq \langle \lambda, x\rangle\,,\ x \in X, \lambda \in X^*\,. \tag{3.28}$$

This is an extension to non-quadratic convex functions f of the inequality

$$\frac{1}{2}\|x\|^2 + \frac{1}{2}\|y\|^2 \geq \langle x|y\rangle$$

in the euclidean space $\mathbb{R}^n$.
We also note that f^* is convex and weak*-lower semicontinuous because it is the supremum of the family of affine continuous functions $(\langle \cdot, x\rangle - f(x))_{x\in X}$. When X is a Hilbert space then the conjugate may be considered as a function on the Hilbert space itself and the only function f which is selfdual ($f = f^*$) is the function $x \longmapsto \frac{1}{2}\|x\|^2$.

We set

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

and call f^{**} the **double conjugate** of f. One has the remarkable duality (see [2])

$$f = f^{**} \iff f \text{ convex and lower semicontinuous}$$

The most important theorem concerns the **Fenchel-Rockafellar duality** which is the content of the following theorem. To formulate this duality we associate to the primal problem (10.33) a **dual problem** using the conjugates of f and g:

$$\text{Maximimize } (-f^*(-\lambda) - g^*(-\lambda))\,,\ \lambda \in X^* \tag{3.29}$$

and we set

$$p := \inf_{x\in X}(f(x) + g(x))\,,\ d := \sup_{\lambda\in X^*}(-(f^* + g^*)(-\lambda))\,,$$

$p \in \bar{\mathbb{R}}, d \in \bar{\mathbb{R}}$ are the values of the **primal** and **dual problem**, respectively.

Corollary 3.52 (Weak duality). *Let* X *be a Banach space and let* $f, g : X \longrightarrow (-\infty, \infty]$ *be convex. Then* $p \geq d$.

Remark 3.53. *In case of* $d < p$ *we say that there exists a* **duality gap**. *Without further assumptions we cannot guarantee that there is no duality gap. If there is no duality gap the identity* $p = d$ *can be used to find a* **lower bound** *for* p *since we have*

$$-(f^* + g^*)(-\lambda) \leq d = p \text{ for all } \lambda \in X^* .$$

An **upper bound** *for* p *is easily found:*

$$p \leq f(x) + g(x) \text{ for all } x \in X .$$

□

Theorem 3.54 (Fenchel, 1949, Rockafellar, 1966). *Let* X *be a Banach space and let* $f, g : X \longrightarrow \hat{\mathbb{R}}$ *be convex. We assume that one of the functions* f, g *is continuous in some point* $x^0 \in dom(f) \cap dom(g)$ *. Then*

$$p = \inf_{x \in X}(f(x) - g(x)) = \max_{\lambda \in X^*}(-(f^* + g^*)(-\lambda)) = d . \tag{3.30}$$

Remark 3.55. *The duality asserts that in the presence of a so called* **primal constraint qualification** *such as „*f *(or* g*) is continuous at a point* $x^0 \in dom(f) \cap dom(g)$*" one has*

$$p := \inf_{x \in X}(f(x) + g(x)) = d := \sup_{\lambda (\in X^*}(-f^* + g^*)(-\lambda))$$

and the supremum for d *is a maximum. If we add a bf dual constraint qualification such as „*f^* *(or* g^**) is continuous at a point* $\lambda^0 \in dom(f^*) \cap dom(g^*)$*" one has that the infimum for* p *is a minimum and we have that the primal and the dual problem are solvable and that there exists no duality gap.* □

Let X be a Banach space. In the theory of monotone operators $T : X \rightrightarrows X^*$ a very important role is played by the so called duality mapping. It essentially does in Banach space the job done by the identity in Hilbert spaces.

Definition 3.56. *Let* X *be a Banach space. The* **duality mapping** $J = J_X : X \rightrightarrows X^*$ *is defined as follows:*

$$J(\theta) = \theta \,,\ J(x) = \{\mu \in X^* | \langle \mu, x \rangle = \|x\|^2, \|\mu\| = \|x\|\} \text{ for } x \neq \theta . \tag{3.31}$$

□

Expression (10.38) is equivalent to

$$J(x) = \{\mu \in X^* | \langle \mu, x \rangle \geq \|x\|^2, \|\mu\| \leq \|x\|\} \text{ for } x \in X . \tag{3.32}$$

One can show directly from the definition, using the Hahn-Banach, that $J(x)$ is a non-empty, weak*-closed convex subset of X for each $x \in X$.

Lemma 3.57. *Let* X *be a Banach space and let* $j : X \longrightarrow \mathbb{R}$ *be the function defined by* $x \longmapsto \frac{1}{2}\|x\|^2$ *. Then*

$$\partial j(x) = J(x) \,,\ x \in X . \tag{3.33}$$

Proof:
Let $\mu \in \partial j(x)$. It suffices to consider the case $x \neq \theta$ since $\partial j(\theta) = \{\theta\}$ and $J(\theta) = \{\theta\}$. Then

$$\frac{1}{2}\|y\|^2 \geq \frac{1}{2}\|x\|^2 + \langle \mu, y - x \rangle \,,\ x \in X\,. \tag{3.34}$$

Let $t > 0$ and $u \in X$ be arbitrary, and replace in (10.40) y by $x + tu$:

$$t\langle \mu, u \rangle \leq \frac{1}{2}\|x + tu\|^2 - \frac{1}{2}\|x\|^2 \leq t\|x\|\|u\| + \frac{1}{2}t^2\|u\|^2\,.$$

Dividing through by t and passing to the limit $t \downarrow 0$ we obtain $\langle \mu, u \rangle \leq \|x\|\|u\|$. Since u was arbitrary this implies $\|\mu\| \leq \|x\|$. On the other hand, let $y = tx$ in (10.40):

$$\frac{1}{2}(t^2 - 1)\|x\|^2 \geq (t-1)\langle \mu, x \rangle\,. \tag{3.35}$$

For $0 < t < 1$ we obtain from (10.41) $\frac{1}{2}(t+1)\|x\|^2 \leq \langle \mu, x \rangle$. Letting $t \to 1$ we have $\langle \mu, x \rangle \geq \|x\|^2$, completing the proof that $\mu \in J(x)$.
For other way around, assume now that $\mu \in J(x)$. Let us estimate the righthand side in (10.40) using the properties of μ :

$$\langle \mu, y \rangle - \langle \mu, x \rangle + \frac{1}{2}\|x\|^2 \leq \|x\|\|y\| - \|x\|^2 + \frac{1}{2}\|x\|^2 \leq \frac{1}{2}\|y\|^2\,.$$

■

Definition 3.58. *Let $\mathcal{X}$ be a Banach space and let C be asubset of $\mathcal{X}$. Then the set*

$$\mathit{core}(C) := \{x \in C : \cup_{r>0}(r(C - x)) = \overline{\mathit{span}(C - x)}$$

is called the **strong relative interior** *of* C . □

Clearly, $\mathrm{int}(C) \subset \mathrm{core}(C)$.

Definition 3.59. *Let $\mathcal{X}$ be a Banach space and let $f, g \in \Gamma_0(\mathcal{X})$. Then the* **inf-convolution** *$f \,\Box\, g$ is defined as follows:*

$$(f \,\Box\, g)(x) := \inf_{u \in \mathcal{X}} (f(u) + g(x - u))\,,\ x \in \mathcal{X}\,.$$

□

Lemma 3.60. *Let $\mathcal{X}$ be a Banach space and let $f, g \in \Gamma_0(\mathcal{X})$, $h := f \,\Box\, g$. Then*

$$\mathit{epi}(h) = \mathit{epi}(f) + \mathit{epi}(g)$$

if the infimum in $f \,\Box\, g$ is attained for all $x \in \mathcal{X}$.

Proof:

■

Theorem 3.61. *Let $\mathcal{X}$ be a Banach space and let $f, g \in \Gamma_0(\mathcal{X})$. Suppose that the constraint qualification*

$$\theta \in \mathit{core}(\mathit{dom}(f) - \mathit{dom}(g))$$

is satisfied. Then

(a) $(f+g)^* = f^* \square g^*$.

(b) *For all* $\lambda \in \mathcal{X}^*$ *exists* $\mu \in \mathcal{X}^*$ *with* $(f+g)^*(\lambda) = f^*(\mu) + g^*(\lambda - \mu)$.

(c) $\partial(f+g) = \partial f + \partial g$.

Proof:

■

3.7 Conclusions and comments

3.8 Exercises

1.) Let $\mathcal{X}$ be a Banach space and let C be a closed convex cone. Then for $w \in \mathcal{X}$ the following conditions are equivalent:

   (a) $w \in P_C(x)$.

   (b) There exists $\lambda \in J(x-w) \cap C^\circ$ such that $\langle l, w\rangle = 0$. Here $C^\circ := \langle \mu \in X^* : \langle \mu, u\rangle \leq 0$ for all $u \in C\}$ is the polar cone of C.

2.) Let $\mathcal{H}$ be a Hilbert space, let A, B be a closed subsets of $\mathcal{H}$ and let $x \in \mathcal{H}$. Show: If $P_B(x) \in A$, then $P_A(x) = P_B(x)$.

3.) Let $\mathcal{X}$ be a Banach space and let $H := H_{\lambda,\alpha} := \{x \in \mathcal{X} : \langle \lambda, x\rangle = \alpha\}$ be a hyperplane ($\lambda \in \mathcal{X}^*, \lambda \neq \theta, \alpha \in \mathbb{R}$). Show $\mathrm{dist}(x, H) = |\langle \lambda, x\rangle - \alpha| \|\lambda\|^{-1}$ for all $x \in \mathcal{X}$.

4.) Let $\mathcal{X}$ be a Banach space and let $H := H_{\lambda,\alpha} := \{x \in \mathcal{X} : \langle \lambda, x\rangle = \alpha\}$ be a hyperplane ($\lambda \in \mathcal{X}^*, \lambda \neq \theta, \alpha \in \mathbb{R}$) which is a Chebyshev set. Show that P_H is linear.

5.) Definiere $f : \mathbb{R} \ni r \longmapsto \frac{1}{2}\mathrm{dist}(r, 2\mathbb{Z}) \in \mathbb{R}$. Set $C := \mathrm{graph}(f)$ and consider $\mathbb{R}^2$ endowed with the l_1-norm. Show that C is a nonconvex Chebyshev set.

6.) Let $\mathbb{R}^2$ endowed with the maximum norm. Show that the orthogonality relation of Birkhoff is not symmetric relation.

7.) Show that in an inner product space $(X, \langle\cdot|\cdot\rangle)$ for $x, y \in X$ the following conditions are equivalent:

   (a) $\langle x|y\rangle = 0$

   (b) $\|x\| \leq \|x + ty\|$ for all $t \in \mathbb{R}$.

8.) Let $\mathcal{X}$ be a normed space and let $x, y \in \mathcal{X}$. Set $L_y := \mathrm{span}(\{y\})$. Show that the following conditions are equivalent:

   (a) $x \perp y$ in the sense of Birkhoff.

   (b) $\|x\| = \mathrm{dist}(x, L_y)$.

9.) Let $\mathcal{X}^*$ be a strictly convex Banach space with duality map J. Show that $x, y \in \mathcal{X}$ are orthogonal in the sense of Birkhoff iff that x, y are J-orthogonal.

10.) Let $\mathcal{X}^*$ be a strictly convex Banach space with duality map J. Show: If $x^1, \dots, x^m \in \mathcal{X}$ are linearly independent then $x^1, \dots, x^m$ are pairwise J-orthogonal.

11.) Let $\mathcal{X}$ be a Banach space with a single-valued duality mappings $J : \mathcal{X} \rightrightarrows \mathcal{X}^*, J^* : \mathcal{X}^* \rightrightarrows \mathcal{X}$. If J is strongly monotone, i.e.

$$\langle J(x) - J(y), x - y\rangle \geq b\|x - y\|^2 \text{ for all } x, y \in \mathcal{X}$$

with a positive constant b, then J is Lipschitz-continuous iff J^* is strongly monotone.

12.) Let $\mathcal{X}$ be a smooth Banach space. Consider

$$g : \mathcal{X} \times \mathcal{X} \longrightarrow \mathbb{R}\,,\; g(x, y) := \frac{1}{2}\|x\|(\lim_{t\downarrow 0} \frac{\|x + ty\| - \|x\|}{t} + \lim_{t\uparrow 0} \frac{\|x + ty\| - \|x\|}{t})\,.$$

Then:

(a) $g(x, \cdot) \in J(x)$ for all $x \neq \theta$.

(b) An angle $\phi \in [0, \pi]$ between x, y is defined by

$$\cos(\phi) = \frac{g(x, y) + g(y, x)}{2\|x\|\|y\|}$$

(c) If $\mathcal{X}$ is a Hilbert space $\mathcal{H}$ then the angle above is the usual in inner product spaces.

13.) Let $\mathcal{H}$ be a Hilbert space and let C be a nonempty closed convex subset of $\mathcal{H}$. Suppose that $U : \mathcal{H} \longrightarrow \mathcal{H}$ is a surjective linear isometry. Then $U(C)$ is a nonempty closed convex subset of $\mathcal{H}$ and $P_{U(C)} = U \circ P_C \circ U^*$.

14.) Let $\mathcal{H}$ be a Hilbert space and let C be a nonempty closed convex cone of $\mathcal{H}$. The set $C^\circ := \{x \in \mathcal{H} : \langle x|y\rangle \leq 0 \text{ for all } y \in C\}$ is called the polar cone of C. Verify $P_{C^\circ} = I - P_C$.

15.) Show $x = P_C(x + z)$ if $z \in N(x, C)$.

16.) Let $\mathcal{H}$ be a Hilbert space and let $S \subset \mathcal{H}$ be a nonempty set. S is called a **sun** if for all $x \in \mathcal{H}$ the following assertion hold:

$$P_S(x) = P_S(u) \text{ for all } u \in \{P_S(x) + t(x - P_S(x)) : t \geq 0\}.$$

Let $C \subset \mathcal{H}$ be a no$(f + g)^* 0 f^* \square g^*$.nempty Chebyshev set. Prove the equivalence of the following conditions:

(a) C is convex.

(b) C is a sun.

(c) P_C is a nonexpansive mapping.

17.) Let $\mathcal{X}$ be a Banach space and let $[\cdot|\cdot]$ be a semi-inner product which generates the norm in $\mathcal{X}$. Then the following statements are equivalent:

(a) $\mathcal{X}$ is smooth.

(b) $\lim_{t\to 0} \frac{[x|x + ty] - \|x\|^2}{t}$ exists for all $x \in \mathcal{X}\backslash\{\theta\}, y \in \mathcal{X}$.

18.) Give an elementary proof of the fact that each convex function $f : \mathbb{R} \longrightarrow \mathbb{R}$ is continuous.

Bibliography

[1] E. Asplund. Positivity of duality mappings. *Amer. Math. Soc.*, 73:200–203, 1966.

[2] J. Baumeister. Funktionalanalysis, 2013. Skriptum einer Vorlesung, Goethe–Universität Frankfurt/Main.

[3] A. Beuerling and A.E. Livingston. A theorem on duality mappings in Banach spaces. *Ark. Mat.*, 4:405–411, 1962.

[4] H. Brézis, M.G. Crandall, and A. Pazy. Perturbations of nonlinear maximal monotone sets in Banach spaces. *Comm. Pure Appl. Math.*, 23:123–144, 1970.

[5] F. Browder. *Nonlinear operators and nonlinear equations of evolution in Banach spaces.* Proc. Symp. Pure and Appl. Math. 18. Providence, 1976.

[6] E.W. Cheney. *Introduction to approximation theory, 2nd ed.* AMS publsihing, Providence, 1982.

[7] I. Cioranescu. *Geometry of Banach Spaces, Duality Mappings and Nonlinear Problems.* Kluwer, Dordrecht, 1990.

[8] F. Deutsch. *Best approximation in inner product spaces.* Springer, New York, 2001.

[9] S.S. Dragomir. *Semi-inner products and applications.* Nova Sciene Hauppage, 2004.

[10] J.R. Giles. Classes of semi-inner-product spaces. *Trans. of Amer. Math. Soc.*, 129:436–446, 1967.

[11] R. Holmes. *A Course on Optimization and Best Approximation.* Springer, 1971.

[12] S. Kinnunen. On projections and birkhoff-james orthogonality in Banach spaces. *Nieuw. Arch. Wisk.*, (IV), 2:235–255, 1984.

[13] G. Lumer. Semi-inner product spaces. *Trans. Amer. Math. Soc.*, 100:29–43, 1961.

[14] R.E. Megginson. *An introduction to Banach space theory.* Springer, New York, 1998.

[15] J.-P. Penot and R. Ratsimahalo. Characteritations of metric projections in Banach spaces and applications. 1998. Preprint.

[16] I. Rosca. Semi produits scalaires et réprésentaion du type riesz pour les functionélles linéaires te bornées sur les espace normés. *C. R. Acad. Sci. Paris*, 283 (19), 1976.

[17] I. Singer. *The theory of best approximation and functional analysis*, volume 13 of *Series in applied mathematics.* SIAM, Philadelphia, 1974.