Optimization and Optimal Control in Banach Spaces

Bernhard Schmitzer

October 15, 2018

1 Convex non-smooth optimization with proximal operators

Remark 1.1 (Motivation)**.** Convex optimization:

- easier to solve, global optimality,
- convexity is strong regularity property, even if functions are not differentiable, even in infinite dimensions,
- usually strong duality,
- special class of algorithms for non-smooth, convex problems; easy to implement and to parallelize. Objective function may assume value $+\infty$, i.e. well suited for implementing constraints.

So if possible: formulate convex optimization problems.
Of course: some phenomena can only be described by non-convex problems, e.g. formation of transport networks.

Definition 1.2. Throughout this section H is Hilbert space, possibly infinite dimensional.

1.1 Convex sets

Definition 1.3 (Convex set)**.** A set $A \subset H$ is convex if for any $a, b \in A$, $\lambda \in [0, 1]$ one has $\lambda \cdot a + (1 - \lambda) \cdot b \in A$.

Comment: Line segment between any two points in A is contained in A

Sketch: Positive example with ellipsoid, counterexample with 'kidney'

Comment: Study of geometry of convex sets is whole branch of mathematical research. See lecture by Prof. Wirth in previous semester for more details. In this lecture: no focus on convex sets, will repeat all relevant properties where required.

Proposition 1.4 (Intersection of convex sets)**.** If $\{C_i\}_{i \in I}$ is family of convex sets, then $C := \bigcap_{i \in I} C_i$ is convex.

Proof.
- Let $x, y \in C$ then for all $i \in I$ have $x, y \in C_i$, thus $\lambda \cdot x + (1 - \lambda) \cdot y \in C_i$ for all $\lambda \in [0, 1]$ and consequently $\lambda \cdot x + (1 - \lambda) \cdot y \in C$.

$\square$

Definition 1.5 (Convex hull)**.** The *convex hull* $\operatorname{conv} C$ of a set C is the intersection of all convex sets that contain C.

Proposition 1.6. Let $C \subset H$, let T be the set of all convex combinations of elements of C, i.e.,

$$T := \left\{ \sum_{i=1}^{k} \lambda_i \, x_i \,\middle|\, k \in \mathbb{N},\, x_1, \dots, x_k \in C,\, \lambda_1, \dots, \lambda_k > 0,\, \sum_{i=1}^{k} \lambda_i = 1 \right\}.$$

Then $T = \operatorname{conv} C$.

Proof.
- **Part I,** $\operatorname{conv} C \subset T$**:** T is convex: any $x, y \in T$ are (finite) convex combinations of points in C. Thus, so is any convex combination of x and y. Also, $C \subset T$. So $\operatorname{conv} C \subset T$.
- **Part II,** $\operatorname{conv} C \supset T$**:** Let S be convex and $S \supset C$. We will show that $S \supset T$ and thus $\operatorname{conv} C \supset T$, which with the previous step implies equality of the two sets.
- We show $S \supset T$ by induction. By definition, any element in T can be represented as follows: For some $k \in \mathbb{N}$, $x_1, \dots, x_k \in C$, $\lambda_1, \dots, \lambda_k > 0$, $\sum_{i=1}^{k} \lambda_i = 1$ let

$$s_k = \sum_{i=1}^{k} \lambda_i \, x_i \,.$$

- When $k = 1$ clearly $s_k = x_1 \in C \subset S$.
- Assume, we have shown that all linear combinations up to $k - 1$ elements in T are also contained in S.
- For $k > 1$ set $\tilde{\lambda}_i = \lambda_i / (1 - \lambda_k)$ for $i = 1, \dots, k - 1$. Then

$$s_k = \lambda_k \, x_k + (1 - \lambda_k) \cdot \underbrace{\sum_{i=1}^{k-1} \tilde{\lambda}_i \, x_i}_{:= s_{k-1}} \,.$$

- We have $x_k \in C \subset S$ and by assumption $s_{k-1} \in S$. Therefore, $s_k \in S$. □

Proposition 1.7 (Carathéodory)**.** Let $H = \mathbb{R}^n$. Every $x \in \operatorname{conv} C$ can be written as convex combination of at most $n + 1$ elements of C.

Proof. Consider arbitrary convex combination $x = \sum_{i=1}^{k} \lambda_i \, x_i$ for $k > n + 1$.
Claim: without changing x can change $(\lambda_i)_i$ such that one λ_i becomes 0.

- The vectors $\{x_2 - x_1, \dots, x_k - x_1\}$ are linearly dependent, since $k - 1 > n$.
- $\Rightarrow$ There are $(\beta_2, \dots, \beta_k) \in \mathbb{R}^{k-1} \setminus \{0\}$ such that

$$0 = \sum_{i=2}^{k} \beta_i \, (x_i - x_1) = \sum_{i=2}^{k} \beta_i \, x_i - \underbrace{\sum_{i=2}^{k} \beta_i}_{:= -\beta_1} x_1 \,.$$

- Define $\tilde{\lambda}_i = \lambda_i - t^* \, \beta_i$ for $t^* = \frac{\lambda_{i^*}}{\beta_{i^*}}$ and $i^* = \operatorname{argmin}_{i=1,\dots,k : \beta_i \neq 0} \frac{\lambda_i}{|\beta_i|}$.

- $\tilde{\lambda}_i \geq 0$: $\tilde{\lambda}_i = \lambda_i \cdot \big(1 - \underbrace{\tfrac{\lambda_{i^*}/\beta_{i^*}}{\lambda_i/\beta_i}}_{|\cdot|\leq 1}\big)$

- $\tilde{\lambda}_{i^*} = 0$

- $\displaystyle\sum_{i=1}^k \tilde{\lambda}_i = \underbrace{\sum_{i=1}^k \lambda_i}_{=1} - t^* \underbrace{\sum_{i=1}^k \beta_i}_{=0} = 1$

- $\displaystyle\sum_{i=1}^k \tilde{\lambda}_i \, x_i = \underbrace{\sum_{i=1}^k \lambda_i \, x_i}_{=x} - t^* \underbrace{\sum_{i=1}^k \beta_i \, x_i}_{=0} = x$

□

1.2 Convex functions

Definition 1.8 (Convex function)**.** A function $f : H \to \mathbb{R} \cup \{\infty\}$ is convex if for all $x, y \in H$, $\lambda \in [0, 1]$ one has $f\big(\lambda \cdot x + (1 - \lambda) \cdot y\big) \leq \lambda \cdot f(x) + (1 - \lambda) \cdot f(y)$. Set of convex functions over H is denoted by $\mathrm{Conv}(H)$.

- f is *strictly convex* if for $x \neq y$ and $\lambda \in (0, 1)$: $f\big(\lambda \cdot x + (1 - \lambda) \cdot y\big) < \lambda \cdot f(x) + (1 - \lambda) \cdot f(y)$.
- f is *concave* if $-f$ is convex.
- The *domain* of f, denoted by $\mathrm{dom}\, f$ is the set $\{x \in H | f(x) < +\infty\}$. f is called *proper* if $\mathrm{dom}\, f \neq \emptyset$.
- The *graph* of f is the set $\{(x, f(x)) | x \in \mathrm{dom}\, f\}$.
- The *epigraph* of f is the set 'above the graph', $\mathrm{epi}\, f = \{(x, r) \in H \times \mathbb{R} | r \geq f(x)\}$.
- The *sublevel set* of f with respect to $r \in \mathbb{R}$ is $S_r(f) = \{x \in H | f(x) \leq r\}$.

Sketch: Strictly convex, graph, secant, epigraph, sublevel set

---

Proposition 1.9. (i) f convex $\Rightarrow$ $\mathrm{dom}\, f$ convex.

(ii) $[f$ convex$] \Leftrightarrow [\mathrm{epi}\, f$ convex$]$.

(iii) $[(x, r) \in \mathrm{epi}\, f] \Leftrightarrow [x \in S_r(f)]$.

Example 1.10. (i) *characteristic or indicator function* of convex set $C \subset H$:

$$\iota_C(x) = \begin{cases} 0 & \text{if } x \in C \\ +\infty & \text{else.} \end{cases} \qquad \text{Do not confuse with} \qquad \chi_C(x) = \begin{cases} 1 & \text{if } x \in C \\ 0 & \text{else.} \end{cases}$$

(ii) any *norm* on H is convex: For all $x, y \in H$, $\lambda \in [0, 1]$:

$$\|\lambda \cdot x + (1 - \lambda) \cdot y\| \leq \|\lambda \cdot x\| + \|(1 - \lambda) \cdot y\| = \lambda \cdot \|x\| + (1 - \lambda) \cdot \|y\|$$

(iii) for $H = \mathbb{R}^n$ the *maximum function*

$$\mathbb{R}^n \ni x \mapsto \max\{x_i | i = 1, \ldots, n\}$$

is convex (follows from previous point, since it is also a norm).

(iv) *linear and affine functions* are convex.

Example 1.11 (Optimization with constraints)**.** Assume we want to solve an optimization problem with linear constraints, e.g.,

$$\min\{f(x) | x \in \mathbb{R}^n,\ A\,x = b\}$$

where $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$, $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$. This can be formally rewritten as unconstrained problem:

$$\min\{f(x) + g(A\,x) | x \in \mathbb{R}^n\} \qquad \text{where} \qquad g = \iota_{\{b\}}\,.$$

We will later discuss algorithms that are particularly suited for problems of this form where one only has to 'interact' with f and g separately, but not their combination.

As mentioned in the motivation: convexity is a strong regularity property. Here we give some examples of consequences of convexity.

Definition 1.12. A function $f : H \to \mathbb{R} \cup \{\infty\}$ is (sequentially) continuous in x if for every convergent sequence $(x_k)_k$ with limit x one has $\lim_{k\to\infty} f(x_k) = f(x)$. The set of points x where $f(x) \in \mathbb{R}$ and f is continuous in x is denoted by $\operatorname{cont} f$.

Remark 1.13 (Continuity in infinite dimensions)**.** If H is infinite dimensional, it is a priori not clear, whether closedness and sequential closedness coincide. But since H is a Hilbert space, it has an inner product, which induces a norm, which induces a metric. On metric spaces the notions of closedness and sequential closedness coincide and thus so do the corresponding notions of continuity.

Proposition 1.14 (On convexity and continuity I)**.** Let $f \in \operatorname{Conv}(H)$ be proper and let $x_0 \in \operatorname{dom} f$. Then the following are equivalent:

(i) f is locally Lipschitz continuous near x_0.

(ii) f is bounded on a neighbourhood of x_0.

(iii) f is bounded from above on a neighbourhood of x_0.

Proof. The implications (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii) are clear. We show (iii) $\Rightarrow$ (i).

- If f is bounded from above in an environment of x_0 then there is some $\rho \in \mathbb{R}_{++}$ such that $\sup f(\overline{B(x_0, \rho)}) = \eta < +\infty$.
- Let $x \in H$, $x \neq x_0$, such that $\alpha := \|x - x_0\|/\rho \in (0, 1]$

Sketch: Draw position of $\tilde{x}$.

- Let $\tilde{x} = x_0 + \frac{1}{\alpha}(x - x_0) \in \overline{B(x_0, \rho)}$. Then $x = (1-\alpha) \cdot x_0 + \alpha \cdot \tilde{x}$ and therefore by convexity of f

$$f(x) \leq (1-\alpha) \cdot f(x_0) + \alpha \cdot f(\tilde{x})$$
$$f(x) - f(x_0) \leq \alpha \cdot (\eta - f(x_0)) = \|x - x_0\| \cdot \tfrac{\eta - f(x_0)}{\rho}$$

Sketch: Draw position of new $\tilde{x}$.

- Now let $\tilde{x} = x_0 + \frac{1}{\alpha}(x_0 - x) \in \overline{B(x_0, \rho)}$. Then $x_0 = \frac{\alpha}{1+\alpha} \cdot \tilde{x} + \frac{1}{1+\alpha} \cdot x$. So:

$$f(x_0) \leq \tfrac{1}{1+\alpha} \cdot f(x) + \tfrac{\alpha}{1+\alpha} \cdot f(\tilde{x})$$
$$f(x_0) - f(x) \leq \tfrac{\alpha}{1+\alpha} \cdot (f(\tilde{x}) - f(x_0) + f(x_0) - f(x))$$
$$f(x_0) - f(x) \leq \alpha \cdot (\eta - f(x_0)) = \|x - x_0\| \cdot \tfrac{\eta - f(x_0)}{\rho}$$

We combine to get:

$$|f(x) - f(x_0)| \leq \|x - x_0\| \cdot \tfrac{\eta - f(x_0)}{\rho}$$

- Now need to extend to other 'base points' near x_0.
- For every $x_1 \in \overline{B(x_0, \rho/4)}$ have $\sup f(\overline{B(x_1, \rho/2)}) \leq \eta$ and $f(x_1) \geq f(x_0) - \frac{\rho}{4} \cdot \frac{\eta - f(x_0)}{\rho} \geq 2\, f(x_0) - \eta$. With arguments above get for every $x \in \overline{B(x_1, \rho/2)}$ that

$$|f(x) - f(x_1)| \leq \|x - x_1\| \cdot \tfrac{\eta - f(x_1)}{\rho/2} \leq \|x - x_1\| \cdot \tfrac{4(\eta - f(x_0))}{\rho}\,.$$

- For every $x_1, x_2 \in \overline{B(x_0, \rho/4)}$ have $\|x_1 - x_2\| \leq \rho/2$ and thus

$$|f(x_1) - f(x_2)| \leq \|x_1 - x_2\| \cdot \tfrac{4(\eta - f(x_0))}{\rho}\,.$$

□

Proposition 1.15 (On convexity and continuity II)**.** If any of the conditions of Proposition 1.14 hold for some $x_0 \in \operatorname{dom} f$, then f is locally Lipschitz continuous on $\operatorname{int}\operatorname{dom} f$.

Proof. **Sketch:** Positions of x_0, x, y and balls $B(x_0, \rho)$, $B(x, \alpha \cdot \rho)$

- By assumption there is some $x_0 \in \operatorname{dom} f$, $\rho \in \mathbb{R}_{++}$ and $\eta < \infty$ such that $\sup f(\overline{B(x_0, \rho)}) \leq \eta$.
- For any $x \in \operatorname{int}\operatorname{dom} f$ there is some $y \in \operatorname{dom} f$ such that $x = \gamma \cdot x_0 + (1-\gamma) \cdot y$ for some $\gamma \in (0, 1)$.
- Further, there is some $\alpha \in (0, \gamma)$ such that $\overline{B(x, \alpha \cdot \rho)} \subset \operatorname{dom} f$ and $y \notin \overline{B(x, \alpha \cdot \rho)}$.
- Then, $\overline{B(x, \alpha \cdot \rho)} \subset \operatorname{conv}(\overline{B(x_0, \rho)} \cup \{y\})$.
- So for any $z \in \overline{B(x, \alpha \cdot \rho)}$ there is some $w \in B(x_0, \rho)$ and some $\beta \in [0, 1]$ such that $z = \beta \cdot w + (1-\beta) \cdot y$. Therefore,

$$f(z) \leq \beta \cdot f(w) + (1-\beta) \cdot f(y) \leq \max\{\eta, f(y)\}\,.$$

- So f is bounded from above on $\overline{B(x, \alpha \cdot \rho)}$ and thus by Proposition 1.14 f is locally Lipschitz near x.

□

Remark 1.16. One can show: If $f : H \to \mathbb{R} \cup \{\infty\}$ is proper, convex and lower semicontinuous, then $\operatorname{cont} f = \operatorname{int} \operatorname{dom} f$.

Proposition 1.17 (On convexity and continuity in finite dimensions)**.** If $f \in \operatorname{Conv}(H = \mathbb{R}^n)$ then f is locally Lipschitz continuous at every point in $\operatorname{int} \operatorname{dom} f$.

Proof.
- Let $x_0 \in \operatorname{int} \operatorname{dom} f$.
- If H is finite-dimensional then there is a finite set $\{x_i\}_{i \in I} \subset \operatorname{dom} f$ such that $x_0 \in \operatorname{int} \operatorname{conv}(\{x_i\}_{i \in I}) \subset \operatorname{dom} f$.
- For example: along every axis $i = 1, \ldots, n$ pick $x_{2i-1} = x + \varepsilon \cdot e_i$, $x_{2i} = x - \varepsilon \cdot e_i$ for sufficiently small ε where e_i denotes the canonical i-th Euclidean basis vector.
- Since every point in $\operatorname{conv}(\{x_i\}_{i \in I})$ can be written as convex combination of $\{x_i\}_{i \in I}$ we find $\sup f(\operatorname{conv}(\{x_i\}_{i \in I})) \leq \max_{i \in I} f(x_i) < +\infty$.
- So f is bounded from above on an environment of x_0 and thus Lipschitz continuous in x_0 by the previous Proposition.

□

Comment: Why is interior necessary in Proposition above?

Example 1.18. The above result does not extend to infinite dimensions.

- For instance, the H^1-norm is not continuous with respect to the topology induced by the L^2-norm.
- An unbounded linear functional is convex but not continuous.

Definition 1.19 (Lower semi-continuity)**.** A function $f : H \to \mathbb{R} \cup \{\infty\}$ is called (sequentially, see Remark 1.13) *lower semicontinuous* in $x \in H$ if for every sequence $(x_n)_n$ that converges to x one has

$$\liminf_{n \to \infty} f(x_n) \geq f(x) \,.$$

f is called lower semicontinuous if it is lower semicontinuous on H.

Example 1.20. $f(x) = \begin{cases} 0 & \text{if } x \leq 0, \\ 1 & \text{if } x > 0 \end{cases}$ is lower semicontinuous, $f(x) = \begin{cases} 0 & \text{if } x < 0, \\ 1 & \text{if } x \geq 0 \end{cases}$ is not.

Sketch: Plot the two graphs.

Comment: Assuming continuity is sometimes impractically strong. Lower semi-continuity is a weaker assumption and also sufficient for well-posedness of minimization problems: If $(x_n)_n$ is a convergent minimizing sequence of a lower semicontinuous function f with limit x then x is a minimizer.

Proposition 1.21. Let $f : H \to \mathbb{R} \cup \{\infty\}$. The following are equivalent:

(i) f is lower semicontinuous.

(ii) $\operatorname{epi} f$ is closed in $H \times \mathbb{R}$.

(iii) The sublevel sets $S_r(f)$ are closed for all $r \in \mathbb{R}$.

Proof. **(i) $\Rightarrow$ (ii).** Let $(y_k, r_k)_k$ be a converging sequence in $\operatorname{epi} f$ with limit (y, r). Then

$$r = \lim_{k\to\infty} r_k \geq \liminf_{k\to\infty} f(y_k) \geq f(y) \qquad \Rightarrow \qquad (y, r) \in \operatorname{epi} f\,.$$

(ii) $\Rightarrow$ (iii). For $r \in \mathbb{R}$ let $A_r : H \to H \times \mathbb{R}$, $x \mapsto (x, r)$ and $Q_r = \operatorname{epi} f \cap (H \times \{r\})$. Q_r is closed, A_r is continuous.

$$S_r(f) = \{x \in H | f(x) \leq r\} = \{x \in H \colon (x, y) \in Q_r\} = A_r^{-1}(Q_r) \qquad \text{is closed.}$$

(iii) $\Rightarrow$ (i). Assume (i) is false. Then there is a sequence $(y_k)_k$ in H converging to $y \in H$ such that $\rho := \lim_{k\to\infty} f(y_k) < f(x)$. Let $r \in (\rho, f(y))$. For $k \geq k_0$ sufficiently large, $f(y_k) \leq r < f(y)$, i.e. $y_k \in S_r(f)$ but $y \notin S_r(f)$. Contradiction. $\square$

1.3 Subdifferential

Definition 1.22. The power set of H is the set of all subsets of H and denoted by 2^H.

Comment: Meaning of notation.

Definition 1.23 (Subdifferential)**.** Let $f : H \to \mathbb{R} \cup \{\infty\}$ be proper. The *subdifferential* of f is the set-valued operator

$$\partial f : H \to 2^H, \qquad x \mapsto \{u \in H | f(y) \geq f(x) + \langle y - x, u\rangle \text{ for all } y \in H\}$$

For $x \in H$, f is *subdifferentiable* at x if $\partial f(x) \neq \emptyset$. Elements of $\partial f(x)$ are called *subgradients* of f at x.

Sketch: Subgradients are slopes of affine functions that bound f from below and are equal to f in x.

Definition 1.24. The *domain* $\operatorname{dom} A$ of a set-valued operator A are the points where $A(x) \neq \emptyset$.

Definition 1.25. Let $f : H \to \mathbb{R} \cup \{\infty\}$ be proper. x is a *minimizer* of f if $f(x) = \inf f(H)$. The set of minimizers of f is denoted by $\operatorname{argmin} f$.

The following is an adaption of first order optimality condition for differentiable functions to convex non-smooth functions.

Proposition 1.26 (Fermat's rule)**.** Let $f : H \to \mathbb{R} \cup \{\infty\}$ be proper. Then

$$\operatorname{argmin} f = \{x \in H \,|\, 0 \in \partial f(x)\}\,.$$

Proof. Let $x \in H$. Then

$$[x \in \operatorname{argmin} f] \Leftrightarrow [f(y) \geq f(x) = f(x) + \langle y - x, 0\rangle \text{ for all } y \in H] \Leftrightarrow [0 \in \partial f(x)]\,. \qquad \square$$

Proposition 1.27 (Basic properties of subdifferential)**.** Let $f : H \to \mathbb{R} \cup \{\infty\}$.

(i) $\partial f(x)$ is closed and convex.

(ii) If $x \in \operatorname{dom} \partial f$ then f is lower semicontinuous at x.

Proof. **(i)**:

$$\partial f(x) = \bigcap_{y \in \operatorname{dom} f} \{u \in H \colon f(y) \geq f(x) + \langle y - x, u\rangle\}$$

So $\partial f(x)$ is the intersection of closed and convex sets. Therefore it is closed and convex.
(ii): Let $u \in \partial f(x)$. Then for all $y \in H$: $f(y) \geq f(x) + \langle y - x, u\rangle$. So, for any sequence $(x_k)_k$ converging to x one finds

$$\liminf_{k\to\infty} f(x_k) \geq f(x) + \liminf_{k\to\infty} \langle x_k - x, u\rangle = f(x)\,. \qquad \square$$

Definition 1.28 (Monotonicity)**.** A set-valued function $A : H \to 2^H$ is monotone if

$$\langle x - y, u - v\rangle \geq 0$$

for every tuple $(x, y, u, v) \in H^4$ such that $u \in A(x)$ and $v \in A(y)$.

Proposition 1.29. The subdifferential of a proper function is monotone.

Proof. Let $u \in \partial f(x)$, $v \in \partial f(y)$. We get:

$$\begin{aligned} f(y) &\geq f(x) + \langle y - x, u \rangle \,, \\ f(x) &\geq f(y) + \langle x - y, v \rangle \,, \end{aligned}$$

and by combining:

$$0 \geq \langle y - x, u - v \rangle$$

□

Proposition 1.30. Let I be a finite index set, let $H = \bigotimes_{i \in I} H_i$ a product of several Hilbert spaces. Let $f_i : H_i \to \mathbb{R} \cup \{\infty\}$ be proper and let $f : H \to \mathbb{R} \cup \{\infty\}$, $x = (x_i)_{i \in I} \mapsto \sum_{i \in I} f_i(x_i)$. Then $\partial f(x) = \bigotimes_{i \in I} \partial f_i(x_i)$.

Proof. $\bigotimes_{i \in I} \partial f_i(x_i) \subset \partial f(x)$**:** For $x \in H$ let $p_i \in \partial f_i(x_i)$. Then

$$f(x + y) = \sum_{i \in I} f_i(x_i + y_i) \geq \sum_{i \in I} f_i(x_i) + \langle y_i, p_i \rangle = f(x) + \langle y, p \rangle \ .$$

Therefore $p = (p_i)_{i \in I} \in \partial f(x)$.
$\partial f(x) \subset \bigotimes_{i \in I} \partial f_i(x_i)$**:** Let $p = (p_i)_{i \in I} \in \partial f(x)$. For $j \in I$ let $y_j \in H_j$ and let $y = (\tilde{y}_i)_{i \in I}$ where $\tilde{y}_i = 0$ if $i \neq j$ and $\tilde{y}_j = y_j$. We get

$$f(x + y) = \sum_{i \in I} f_i(x_i + \tilde{y}_i) = \sum_{i \in I \setminus \{j\}} f_i(x_i) + f_j(x_j + y_j) \geq f(x) + \langle y, p \rangle = \sum_{i \in I} f_i(x_i) + \langle y_j, p_j \rangle$$

This holds for all $y_j \in H_j$. Therefore, $p_j \in \partial f_j(x_j)$. □

Example 1.31.

- $f(x) = \frac{1}{2}\|x\|^2$: f is Gâteaux differentiable (see below) with $\nabla f(x) = x$. We will show that this implies $\partial f(x) = \{\nabla f(x)\} = \{x\}$.
- $f(x) = \|x\|$:
  - For $x \neq 0$ f is again Gâteaux differentiable with $\nabla f(x) = \frac{x}{\|x\|}$.
  - For $x = 0$ we get $f(y) \geq \langle y, p \rangle = f(0) + \langle y - 0, p \rangle$ for $\|p\| \leq 1$ via the Cauchy-Schwarz inequality. So $\overline{B(0,1)} \subset \partial f(0)$.
  - Assume some $p \in \partial f(0)$ has $\|p\| > 1$. Then $\frac{p}{\|p\|} \in \partial f(p)$. We test: $\left\langle p - 0, \frac{p}{\|p\|} - p \right\rangle = \|p\| - \|p\|^2 < 0$ which contradicts monotonicity of the subdifferential. Therefore $\partial f(0) = \overline{B(0,1)}$.
- $H = \mathbb{R}$, $f(x) = |x|$ is a special case of the above.

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

Sketch: Draw 'graph' of subdifferential.

- $H = \mathbb{R}^n$, $f(x) = \|x\|_1$. The ℓ_1 norm is not induced by an inner product. Therefore the above does not apply. We can use Proposition 1.30:

$$\partial f(x) = \bigotimes_{k=1}^{n} \partial \mathrm{abs}(x_k)$$

Sketch: Draw subdifferential 'graph' for 2D.

Proposition 1.32. Let $f, g : H \to \mathbb{R} \cup \{\infty\}$. For $x \in H$ one finds $\partial f(x) + \partial g(x) \subset \partial(f+g)(x)$.

Proof. Let $u \in \partial f(x)$, $v \in \partial g(x)$. Then

$$f(x+y) + g(x+y) \geq f(x) + \langle u, y\rangle + g(x) + \langle v, y\rangle = f(x) + g(x) + \langle u+v, y\rangle \ .$$

Therefore, $u + v \in \partial(f+g)(x)$. □

Remark 1.33. The converse inclusion is not true in general and much harder to proof. A simple counter-example is $f(x) = \|x\|^2$ and $g(x) = -\|x\|^2/2$. The subdifferential of g is empty but the subdifferential of $f+g$ is not.

An application of the sub-differential is a simple proof of Jensen's inequality.

Proposition 1.34 (Jensen's inequality)**.** Let $f : H = \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ be convex. Let μ be a probability measure on H such that

$$\overline{x} = \int_H x \,\mathrm{d}\mu(x) \in H$$

and $\overline{x} \in \mathrm{dom}\, \partial f$. Then

$$\int_H f(x)\,\mathrm{d}\mu(x) \geq f(\overline{x})\,.$$

Proof. Let $u \in \partial f(\overline{x})$.

$$\int_H f(x)\,\mathrm{d}\mu(x) \geq \int_H f(\overline{x}) + \langle x - \overline{x}, u\rangle \;\mathrm{d}\mu(x) = f(\overline{x})$$

□

Let us examine the subdifferential of differentiable functions.

Definition 1.35 (Gâteaux differentiability)**.** A function $f : H \to \mathbb{R} \cup \{\infty\}$ is *Gâteaux differentiable* in $x \in \mathrm{dom}\, f$ if there is a unique *Gâteaux gradient* $\nabla f(x) \in H$ such that for any $y \in H$ the directional derivative is given by

$$\lim_{\alpha \searrow 0} \tfrac{f(x+\alpha\cdot y) - f(x)}{\alpha} = \langle y, \nabla f(x)\rangle \ .$$

Proposition 1.36. Let $f : H \to \mathbb{R} \cup \{\infty\}$ be proper and convex, let $x \in \mathrm{dom}\, f$. If f is Gâteaux differentiable in x then $\partial f(x) = \{\nabla f(x)\}$.

Proof. $\nabla f(x) \in \partial f(x)$**:**

- For fixed $y \in H$ consider the function $\phi : (0,\infty) \to \mathbb{R} \cup \{\infty\}$, $\alpha \mapsto \frac{f(x+\alpha\cdot y)-f(x)}{\alpha}$.
- ϕ is increasing: let $\beta \in (0,\alpha)$. Then $x + \beta \cdot y = (1-\beta/\alpha)\cdot x + \beta/\alpha \cdot (x+\alpha\cdot y)$. So

$$\begin{aligned} f(x+\beta\cdot y) &\leq (1-\beta/\alpha)\cdot f(x) + \beta/\alpha \cdot f(x+\alpha\cdot y), \\ \phi(\beta) &\leq \frac{(1-\beta/\alpha)\cdot f(x) + \beta/\alpha\cdot f(x+\alpha\cdot y) - f(x)}{\beta} \\ &= \frac{\beta/\alpha \cdot (f(x+\alpha\cdot y)-f(x))}{\beta} = \phi(\alpha)\,. \end{aligned}$$

- Therefore,

$$\langle y, \nabla f(x)\rangle = \lim_{\alpha \searrow 0} \frac{f(x+\alpha\cdot y)-f(x)}{\alpha} = \inf_{\alpha\in\mathbb{R}_{++}} \phi(\alpha) \leq f(x+y)-f(x)\,.$$

  (We set $\alpha = 1$ to get the last inequality.)

$\partial f(x) \subset \{\nabla f(x)\}$:

- For $u \in \partial f(x)$ we find for any $y \in H$

$$\langle y, \nabla f(x)\rangle = \lim_{\alpha\searrow 0} \frac{f(x+\alpha\cdot y)-f(x)}{\alpha} \geq \lim_{\alpha\searrow 0} \frac{f(x)+\langle \alpha\cdot y, u\rangle - f(x)}{\alpha} = \langle y, u\rangle\ .$$

- This inequality holds for any y and $-y$ simultaneously. Therefore $u = \nabla f(x)$.

□

Remark 1.37. For differentiable functions in one dimension this implies monotonicity of the derivative: Let $f \in C^1(\mathbb{R})$. With Propositions 1.36 and 1.29 we get: if $x \geq y$ then $f'(x) \geq f'(y)$.

1.4 Cones and support functions

Cones are a special class of sets with many applications in convex analysis.

Definition 1.38. A set $C \subset H$ is a *cone* if for any $x \in C$, $\lambda \in \mathbb{R}_{++}$ one has $\lambda \cdot x \in C$. In short notation: $C = \mathbb{R}_{++} \cdot C$.

Remark 1.39. A cone need not contain 0, but for any $x \in C$ it must contain the open line segment $(0, x]$.

Proposition 1.40. The intersection of a family $\{C_i\}_{i \in I}$ of cones is cone. The *conical hull* of a set $C \subset H$, denoted by $\operatorname{cone} C$ is the smallest cone that contains C. It is given by $\mathbb{R}_{++} \cdot C$.

Proof.

- Let $C = \bigcap_{i \in I} C_i$. If $x \in C$ then $x \in C_i$ for all $i \in I$ and for any $\lambda \in \mathbb{R}_{++}$ one has $\lambda \cdot x \in C_i$ for all $i \in I$. Hence $\lambda \cdot x \in C$ and C is also a cone.
- Let $D = \mathbb{R}_{++} \cdot C$. Then D is a cone, $C \subset D$ and therefore $\operatorname{cone} C \subset D$. Conversely, let $y \in D$. Then there are $x \in C$ and $\lambda \in \mathbb{R}_{++}$ such that $y = \lambda \cdot x$. So $x \in \operatorname{cone} C$, therefore $y \in \operatorname{cone} C$ and thus $D \subset \operatorname{cone} C$.

$\square$

Proposition 1.41. A cone C is convex if and only if $C + C \subset C$.

Proof. C **convex** $\Rightarrow C + C \subset C$**:** Let $a, b \in C$. $\Rightarrow \frac{1}{2} \cdot a + \frac{1}{2} \cdot b \in C \Rightarrow a + b \in C \Rightarrow C + C \subset C$.
$C + C \subset C \Rightarrow C$ **convex:** Let $a, b \in C$. $\Rightarrow a + b \in C$ and $\lambda \cdot a, (1 - \lambda) \cdot b \in C$ for all $\lambda \in (0, 1)$. $\Rightarrow \lambda \cdot a + (1 - \lambda) \cdot b \in C$. $\Rightarrow [a, b] \in C \Rightarrow C$ convex. $\square$

Definition 1.42. Let $C \subset H$. The *polar cone* of C is

$$C^{\ominus} = \{y \in H \,|\, \sup \langle C, y \rangle \leq 0\} .$$

Sketch: Draw a cone in 2D with angle $< \pi/2$ and its polar cone.

Proposition 1.43. Let C be a linear subspace of H. Then $C^{\ominus} = C^{\perp}$.

Proof.

- Since C is a linear subspace, if $\langle x, y \rangle \neq 0$ for some $y \in H$, $x \in C$ then $\sup \langle C, y \rangle = \infty$.
- Therefore, $C^{\ominus} = \{y \in H\colon\ \langle x, y \rangle = 0 \text{ for all } x \in C\}$.

$\square$

Definition 1.44. Let $C \subset H$ convex, non-empty and $x \in H$. The *tangent cone* to C at x is

$$T_C x = \begin{cases} \overline{\operatorname{cone}(C - x)} & \text{if } x \in C, \\ \emptyset & \text{else.} \end{cases}$$

The *normal cone* to C at x is

$$N_C x = \begin{cases} (C - x)^{\ominus} = \{u \in H\colon\ \sup \langle C - x, u \rangle \leq 0\} & \text{if } x \in C, \\ \emptyset & \text{else.} \end{cases}$$

Example 1.45. Let $C = \overline{B(0,1)}$. Then for $x \in C$:

$$T_C x = \begin{cases} \{y \in H\colon\ \langle y, x\rangle \leq 0\} & \text{if } \|x\| = 1, \\ H & \text{if } \|x\| < 1. \end{cases}$$

Note: the $\leq$ in the $\|x\| = 1$ case comes from the closure in the definition of $T_C x$. Without closure it would merely be $<$.

$$N_C x = \begin{cases} \mathbb{R}_+ \cdot x & \text{if } \|x\| = 1, \\ \{0\} & \text{if } \|x\| < 1. \end{cases}$$

Example 1.46. What are tangent and normal cone for the L_1-norm ball in $\mathbb{R}^2$?

We start to see connections between different concepts introduced so far.

Proposition 1.47. Let $C \subset H$ be a convex set. Then $\partial \iota_C(x) = N_C x$.

Proof.

- $x \notin C$: $\partial \iota_C(x) = \emptyset = N_C x$.
- $x \in C$:

$$\begin{aligned}[u \in \partial \iota_C(x)] \quad &\Leftrightarrow \quad [\iota_C(y) \geq \iota_C(x) + \langle y - x, u\rangle\ \forall\, y \in C] \Leftrightarrow [0 \geq \langle y - x, u\rangle\ \forall\, y \in C] \\ &\Leftrightarrow [\sup \langle C - x, u\rangle \leq 0] \Leftrightarrow [u \in N_C x]\end{aligned}$$

□

Comment: This will become relevant, when doing constrained optimization, where parts of the objective are given by indicator functions.

Now we introduce the projection onto convex sets. It will play an important role in analysis and numerical methods for constrained optimization.

Proposition 1.48 (Projection)**.** Let $C \subset H$ be non-empty, closed convex. For $x \in H$ the problem

$$\inf\{\|x - p\| \,|\, p \in C\}$$

has a unique minimizer. This minimizer is called the *projection* of x onto C and is denoted by $P_C x$.

Proof.

- We will need the following inequality for any $x, y, z \in H$, which can be shown by careful expansion:

$$\|x - y\|^2 = 2\,\|x - z\|^2 + 2\,\|y - z\|^2 - 4\,\|(x + y)/2 - z\|^2$$

- C is non-empty, $y \mapsto \|x - y\|$ is bounded from below, so the infimal value is a real number, denoted by d.
- Let $(p_k)_{k \in \mathbb{N}}$ be a minimizing sequence. For $k, l \in \mathbb{N}$ one has $\frac{1}{2}(p_k + p_l) \in C$ by convexity and therefore $\|x - \frac{1}{2}(p_k + p_l)\| \geq d$.
- With the above inequality we find:

$$\|p_k - p_l\|^2 = 2\|p_k - x\|^2 + 2\|p_l - x\|^2 - 4\|\tfrac{p_k + p_l}{2} - x\|^2 \leq 2\|p_k - x\|^2 + 2\|p_l - x\|^2 - 4\,d^2$$

- So by sending $k, l \to \infty$ we find that $(p_k)_k$ is a Cauchy sequence which converges to a limit p. Since C is closed, $p \in C$. And since $y \mapsto \|x-y\|$ is continuous, p is a minimizer.
- Uniqueness of p, quick answer: the optimization problem is equivalent to minimizing $y \mapsto \|x-y\|^2$, which is strictly convex. Therefore p must be unique.
- Uniqueness of p, detailed answer: assume there is another minimizer $q \neq p$. Then $\frac{1}{2}(p+q) \in C$ and we find:

$$\|x-p\|^2 + \|x-q\|^2 - 2\|x - \tfrac{1}{2}(p+q)\|^2 = \tfrac{1}{2}\|p-q\|^2 > 0$$

  So the sum of the objectives at p and q is strictly larger than twice the objective at the midpoint. Therefore, neither p nor q can be optimal.

$\square$

Proposition 1.49 (Characterization of projection)**.** Let $C \subset H$ be non-empty, convex, closed. Then $p = P_C x$ if and only if

$$[p \in C] \wedge [\langle y-p, x-p\rangle \leq 0 \text{ for all } y \in C]\,.$$

Sketch: Illustrate inequality.

Proof.

- It is clear that $[p = P_C x] \Rightarrow [p \in C]$, and that $[p \notin C] \Rightarrow [p \neq P_C x]$.
- So, need to show that for $p \in C$ one has $[p = P_C x] \Leftrightarrow [\langle y-p, x-p\rangle \leq 0 \text{ for all } y \in C]$.
- For some $y \in C$ and some $\varepsilon \in \mathbb{R}_{++}$ consider:

$$\begin{aligned}\|x-(p+\varepsilon\cdot(y-p))\|^2 - \|x-p\|^2 &= \|p+\varepsilon\cdot(y-p)\|^2 - \|p\|^2 - 2\,\varepsilon\,\langle x, y-p\rangle \\ &= \varepsilon^2\|y-p\|^2 - 2\,\varepsilon\,\langle x-p, y-p\rangle\end{aligned}$$

  If $\langle x-p, y-p\rangle > 0$ then this is negative for sufficiently small ε and thus p cannot be the projection. Conversely, if $\langle x-p, y-p\rangle \leq 0$ for all $y \in C$, then for $\varepsilon = 1$ we see that p is indeed the minimizer of $y \mapsto \|x-y\|^2$ over C and thus the projection.

$\square$

Corollary 1.50 (Projection and normal cone)**.** Let $C \subset H$ be non-empty, closed, convex. Then $[p = P_C x] \Leftrightarrow [x \in p + N_C p]$.

Proof. $[p = P_C x] \Leftrightarrow [p \in C \wedge \sup\langle C-p, x-p\rangle \geq 0] \Leftrightarrow [x - p \in N_C p]$. $\square$

Comment: This condition is actually useful for computing projections.

Example 1.51 (Projection onto L_1-ball in $\mathbb{R}^2$)**.** Let $C = \{(x,y) \in \mathbb{R}^2 \,|\, |x|+|y| \leq 1\}$. We find:

$$N_C(x,y) = \begin{cases} \emptyset & \text{if } |x|+|y| > 1, \\ \{0\} & \text{if } |x|+|y| < 1, \\ \overline{\text{cone}}\{(1,1),(-1,1)\} & \text{if } (x,y) = (0,1), \\ \overline{\text{cone}}\{(1,1),(1,-1)\} & \text{if } (x,y) = (1,0), \\ \overline{\text{cone}}\{(1,1)\} & \text{if } x+y = 1, x \in (0,1), \\ \ldots \end{cases}$$

Sketch: Draw normal cones attached to points in C.

Now compute projection of $(a, b) \in \mathbb{R}^2$. W.l.o.g. assume $(a, b) \in \mathbb{R}^2_+ \setminus C$. Then

$$P_C(a, b) = \begin{cases} (0,1) & \text{if } [a+b \geq 1] \wedge [b-a \geq 1], \\ (1,0) & \text{if } [a+b \geq 1] \wedge [a-b \geq 1], \\ ((1+a-b)/2, (1-a+b)/2) & \text{else.} \end{cases}$$

Comment: Do computation in detail.

Comment: Result is very intuitive, but not so trivial to prove rigorously due to non-smoothness of problem. Comment: Eistüte.

We now establish a sequence of results that will later allow us to analyze the subdifferential via cones and prepare results for the study of the Fenchel–Legendre conjugate.

Proposition 1.52. Let $K \subset H$ be a non-empty, closed, convex cone. Let $x, p \in H$. Then

$$[p = P_K x] \quad \Leftrightarrow \quad [p \in K,\ x - p \perp p,\ x - p \in K^{\ominus}] .$$

Proof.

- By virtue of Corollary 1.50 (Characterization of projection with normal cone inclusion) we need to show

$$[x - p \in N_K p] \quad \Leftrightarrow [p \in K,\ x - p \perp p,\ x - p \in K^{\ominus}] .$$

- **⇒:** Let $x - p \in N_K p$. Then $p \in K$. By definition have $\sup \langle K - p, x - p\rangle \leq 0$. Since $2p, 0 \in K$ (K is closed) this implies $\langle p, x - p\rangle = 0$. Further, since K is convex, we have (Prop. 1.41) $K + K \subset K$, and in particular $K + p \subset K$. Therefore $\sup \langle K + p - p, x - p\rangle \leq \sup \langle K - p, x - p\rangle \leq 0$ and thus $x - p \in K^{\ominus}$.

  Sketch: Recall that $K + p \subset K$. Counter-example for non-convex K.

- **⇐:** Since $p \perp x - p$ have $\sup \langle K - p, x - p\rangle = \sup \langle K, x - p\rangle \leq 0$ since $x - p \in K^{\ominus}$. Then, since $p \in K$ have $x - p \in N_K p$.

□

Proposition 1.53. Let $K \subset H$ be a non-empty, closed, convex cone. Then $K^{\ominus\ominus} = K$.

Proof.

- **$K \subset K^{\ominus\ominus}$:** Recall: $K^{\ominus} = \{u \in H \mid \sup \langle K, u\rangle \leq 0\}$.
- Let $x \in K$. Then $\langle x, u\rangle \leq 0$ for all $u \in K^{\ominus}$. Therefore $\sup \langle x, K^{\ominus}\rangle \leq 0$ and so $x \in K^{\ominus\ominus}$. Therefore: $K \subset K^{\ominus\ominus}$.
- **$K^{\ominus\ominus} \subset K$:** Let $x \in K^{\ominus\ominus}$, set $p \in P_K x$. Then by Proposition 1.52 (Projection onto closed, convex cone): $x - p \perp p$, $x - p \in K^{\ominus}$.
- $[x \in K^{\ominus\ominus}] \wedge [x - p \in K^{\ominus}] \Rightarrow \langle x, x - p\rangle \leq 0$.
- $\|x - p\|^2 = \langle x, x - p\rangle - \langle p, x - p\rangle \leq 0 \Rightarrow x = p \Rightarrow x \in K$. Therefore $K^{\ominus\ominus} \subset K$.

□

For subsequent results we need the following Lemma that once more illustrates that convexity implies strong regularity.