Optimization and Optimal Control in Banach Spaces

Bernhard Schmitzer

October 13, 2017

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 $C \stackrel{\text{def.}}{=} \bigcap_{i \in I} C_i$ 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 $\lambda \cdot x + (1 - \lambda) \cdot y \in C$.

□

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.,

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

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

Proof. $\operatorname{conv} C \subset T$**.** 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 $\operatorname{conv} C \subset T$.
$\operatorname{conv} C \supset T$**.** 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 recursion. 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 $s_k \in S$.
Otherwise, set $\tilde{\lambda}_i = \lambda_i / (1 - \lambda_k)$ for $i = 1, \dots, k-1$. Then

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

We find that $s_k \in S$ if $s_{k-1} \in S$. Applying this argument recursively to s_{k-1} until we reach s_1, we have shown that $s_k \in S$. □

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

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

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

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

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

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

□

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 $\operatorname{Conv}(H)$.

- 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 $\operatorname{dom} f$ is the set $\{x \in H \colon f(x) < +\infty\}$. f is called *proper* if $\operatorname{dom} f \neq \emptyset$.
- The *graph* of f is the set $\{(x, f(x)) | x \in \operatorname{dom} f\}$.
- The *epigraph* of f is the set 'above the graph', $\operatorname{epi} f = \{(x, r) \in H \times \mathbb{R} \colon r \geq f(x)\}$.
- The *sublevel set* of f with respect to $r \in \mathbb{R}$ is $S_r(f) = \{x \in H \colon f(x) \leq r\}$.

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

---

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

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

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

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.

(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.,

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

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

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

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 $\alpha \overset{\text{def.}}{=} \|x - x_0\|/\rho \in (0, 1]$

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:

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

  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, 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 semi-continuous, then $\operatorname{cont} f = \operatorname{int}\operatorname{dom} f$.

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

Proof.

- Let $x_0 \in \mathrm{int}\,\mathrm{dom}\, f$.
- If H is finite-dimensional then there is a finite set $\{x_i\}_{i \in I} \subset \mathrm{dom}\, f$ such that $x_0 \in \mathrm{int}\,\mathrm{conv}(\{x_i\}_{i \in I}) \subset \mathrm{dom}\, f$.
- 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 $\mathrm{conv}(\{x_i\}_{i \in I})$ can be written as convex combination of $\{x_i\}_{i \in I}$ we find $\sup f(\mathrm{conv}(\{x_i\}_{i \in I})) \leq \max_{i \in I} f(x_i) < +\infty$.
- 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 semi-continuous* 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 semi-continuous if it is lower semi-continuous on H.

Example 1.20. $f(x) = \begin{cases} 0 & \text{if } x \leq 0, \\ 1 & \text{if } x > 0 \end{cases}$ is lower semi-continuous, $f(x) = \begin{cases} 0 & \text{if } x < 0, \\ 1 & \text{if } x \geq 0 \end{cases}$ 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 semi-continuous 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 semi-continuous.

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

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

Proof. **(i) ⇒ (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) ⇒ (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.

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

(iii) ⇒ (i). Assume (i) is false. Then there is a sequence $(y_k)_k$ in H converging to $y \in H$ such that $\rho \overset{\text{def.}}{=} \lim_{k\to\infty} f(y_k) < f(x)$. 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. □