Seminar on Convex Analysis Freie Universität Berlin Summer Term 2023	valentinpi 25 July, 2023

Presentation Notes on: Further Properties of Subdifferentials and The Situation in $\mathbb{R}^n$

1 Preliminaries We first recall the following notions (without the relevant properties for now) from the previous talks. Most of the necessary related, technical theorems will be recited in dedicated boxes when needed. From now on, assume that X and Y are separable locally convex topological vector spaces over $\mathbb{R}$.

Topological Vector Spaces (TVS)	[1, pp. 30-31]
Convex Sets	[2, p. 45]
Convex Hulls	[2, p. 162]
Kones K, K_U	[2, p. 45, p. 162]
Lower/Upper Semicontinuity $f(x) \geq f(x_0) - \varepsilon$ / $f(x) \leq f(x_0) + \varepsilon$	[2, p. 12]
Epigraphs $\operatorname{epi}(f)$	[2, p. 45]
Convex Functions $\operatorname{conv}(\operatorname{epi}(f)) = \operatorname{epi}(f)$	[2, p. 45]
Jensens Inequality $f((1-\lambda)x + \lambda y) \leq (1-\lambda)f(x) + \lambda f(y)$	[2, p. 167]
Proper Functions $-\infty < f \neq \infty$	[2, p. 45]
Effective Domain $\operatorname{dom}(f)$	[2, p. 45]
Affine Hull $\operatorname{aff}(A)$	[2, p. 186]
Relative Interior $\operatorname{ri}(f)$	[2, p. 187]
Subgradients $f(z) \geq f(x) + \langle x^*, z - x\rangle \,\forall\, z \in X$	[2, p. 46]
Subdifferentials $\partial f(\cdot)$	[2, p. 46]
Conjugates f^*	[2, pp. 171-172]
Directional Derivatives $f'(\cdot;\cdot)$	[2, p. 193]

We denote the space of continuous linear operators between them as $L(X, Y)$ and the dual spaces as X^*, Y^*. Recall that we equip these dual spaces with their respective weak* topologies, giving especially that any continuous functional in the bi-dual X^{**} and analogously Y^{**} is of form $\hat{x}\colon X^* \to \mathbb{R}, x^* \mapsto \hat{x}(x^*) := x^*(x)$ for an $x \in X$ [3, pp. 439-440]. Let further $n \in \mathbb{N}_{\geq 1}$ and $f\colon X \to \overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$, where the closure of $\mathbb{R}$ is taken in the canonical topology after extension by $\pm\infty$. Also note the convention of expressing the subgradient as above [4, p. 214]. To test it out one can quickly verify that $f \in \{\pm\infty\}$ is subdifferentiable everywhere. Furthermore, note that we will not cite every theorem we use directly here, but we will give references at selected positions. Especially, we want to highlight the following characterizations, which we will use more-or-less implicitly throughout.

Theorem 1.1 ([2, pp. 170-171])**.** Let f be proper and convex. Then the following statements are equivalent.
(i) f is bounded on a neighborhood of $x \in X$.
(ii) f is continuous at x.

Theorem 1.2 ([2, pp. 193-199])**.** Let f be convex. Then the following statements are equivalent.
(i) $x^* \in \partial f(x)$.
(ii) $f(z) \geq f(x) + \langle x^*, z - x\rangle \,\forall\, z \in X$.
(iii) $f(x) + f^*(x^*) = \langle x^*, x\rangle$
(iv) $f'(x, y) = \sup_{x^* \in \partial f(x)} \langle x^*, y\rangle \geq \langle x^*, y\rangle \,\forall\, y \in X$, i.e. $x^* \in \partial f'(x; 0) = \partial f(x) = \operatorname{dom}(f'(x; \cdot)^*)$, if f is additionally proper.

We may remark especially the rather surprising/fascinating statement in Theorem 1.2 (iv). It can be proven by showing that $f'(x; \cdot)$ supports precisely this set [2, p. 192, p. 196], but is not clear from the definitions. We will also use the following fact about the directional derivative.

Theorem 1.3 (Infimum Form [2, pp. 194-195])**.** Let f be proper and convex. Then $f'(x; \cdot)$ is well defined for any $x \in \operatorname{dom}(f)$, convex, proper and for any $y \in X$

$$f'(x; y) = \inf_{\lambda \in \mathbb{R}_{>0}} \frac{f(x + \lambda y) - f(x)}{\lambda} \tag{1.1}$$

2 A Chain Rule Recall the chain rule from univariate differential calculus, which is of form $(f \circ g)' = g' \cdot (f' \circ g)$. Similar statements can be made for the subdifferential of a function with a concatenated continuous linear operator, as the following theorem establishes.

Theorem 2.1 ([2, p. 201])**.** Let $\Lambda \in L(X, Y)$ and $f \colon Y \to \overline{\mathbb{R}}$ be a function.
(i) For any $x \in X$, we have

$$\partial(f\Lambda)(x) \supseteq \Lambda^* \partial f(\Lambda x) \tag{2.2}$$

(ii) If f is convex and proper on X, as well as continuous at a point in $\mathrm{Im}(\Lambda)$, then for any $x \in X$

$$\partial(f\Lambda)(x) = \Lambda^* \partial f(\Lambda x) \tag{2.3}$$

$$\begin{tikzcd} X \arrow[d, "\Lambda"'] \arrow[drr, "\partial(f\Lambda)"] & & \\ Y \arrow[r, "\partial f"'] & \mathcal{P}(Y^*) \arrow[r, "\Lambda^*"'] & \mathcal{P}(X^*) \end{tikzcd}$$

Proof. (i) Fix an $x \in X$ and let $x^* \in \Lambda^* \partial f(\Lambda x)$. Then $\exists\, y^* \in \partial f(\Lambda x)$, s.t. $x^* = \Lambda^* y^* = y^* \Lambda$. Then y^* fulfills by definition

$$f(z') \geq f(\Lambda x) + \langle y^*, z' - \Lambda x \rangle \tag{2.4}$$

for any $z' \in Y$. As $\mathrm{Im}(\Lambda) \subset Y$, we may assume $z' = \Lambda z$ for a $z \in X$ and obtain

$$(f\Lambda)(z) \geq (f\Lambda)(x) + \langle y^*, \Lambda(z - x) \rangle = (f\Lambda)(x) + \langle x^*, z - x \rangle \tag{2.5}$$

by the notation used.[1]
(ii) If $(f\Lambda)(x) = \infty$, then $\partial(f\Lambda)(x) = \emptyset$, giving with (i) the equality. Denote now the point at which f is continuous as $\Lambda\bar{x}$, $\bar{x} \in X$ and additionally fix an $x \in X$. First, we show that $f'(\Lambda x; \cdot)$ is continuous at a point in $\mathrm{Im}(\Lambda)$ (a). In the second step, we use a rule on interchanging conjugates with adjoints to directly obtain the statement (b).
(a) The argument requires the following theorem.

Theorem 2.2 ([2, pp. 195-196])**.** Let $f \colon X \to \overline{\mathbb{R}}$ be a proper, convex function, which is continuous on $\emptyset \neq U \subseteq X$ and $x \in X$ be fixed.
(i) If $|f'(x; \bar{x})| < \infty$ for $\bar{x} \in X$ with $x + \bar{x} \in U$, then $f'(x; \cdot)$ is continuous on $K_{U - \{x\}} \setminus \{0\}$[a].
(ii) If f is continuous at x, then $f'(x; \cdot)$ is finite and continuous on X.

[a] The minus sign $-$ here denotes the algebraic difference of sets.

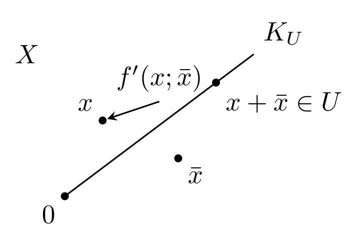

FIGURE 1. Illustration of the situation in Theorem 2.2.

[1] The full calculation reads $\langle y^*, \Lambda(z - x) \rangle = (y^*\Lambda)(z - x) = (\Lambda^* y^*)(z - x) = \langle x^*, z - x \rangle$.

Consider that f is continuous in $\{\Lambda\bar{x} = \Lambda x + \Lambda(\bar{x} - x)\}$ and use the bound

$$f'(\Lambda x; \Lambda(\bar{x} - x)) = \lim_{\lambda \downarrow 0} \frac{f(\Lambda x + \lambda\Lambda(\bar{x} - x)) - f(\Lambda x)}{\lambda} \tag{2.6}$$

$$\leq \lim_{\lambda \downarrow 0} \frac{(1-\lambda)f(\Lambda x) + \lambda f(\Lambda\bar{x}) - f(\Lambda x)}{\lambda} = f(\Lambda\bar{x}) - f(\Lambda x) \overset{(1)}{\rightsquigarrow} |f'(\Lambda x; \Lambda(\bar{x} - x))| < \infty \tag{2.7}$$

(1) Since f is convex and continuous at $\Lambda\bar{x}$, it is finite there.
By Theorem 2.2 (i), $f'(\Lambda x; \cdot)$ is continuous on $K_{\{\Lambda\bar{x}\}-\{\Lambda x\}} \setminus \{0\}$. If $\bar{x} \neq x$, then we are done. Otherwise, as f is continuous at Λx, thus by Theorem 2.2 (ii), $f'(\Lambda x; \cdot)$ is continuous everywhere, especially at a point of $\mathrm{Im}(\Lambda)$.
(b) Consider first the following adjoint rule.

Theorem 2.3 ([2, p. 179, p. 183])**.** Let $\Lambda \in L(X, Y)$ and $f\colon Y \to \overline{\mathbb{R}}$ be a convex function, continuous at a point in $\mathrm{Im}(\Lambda)$. Then $(f\Lambda)^* = \Lambda^* f^*$ and for each $x^* \in \mathrm{dom}((f\Lambda)^*)$, there is a $y^* \in Y^*$ with $x^* = \Lambda^* y^*$ and $(f\Lambda)^*(x^*) = f^*(y^*)$.

The statement is now given by the following calculation.

$$\partial(f\Lambda)(x) \overset{(1)}{=} \partial(f'(\Lambda x; \cdot)\Lambda)(0) \overset{(2)}{=} \Lambda^* \partial f'(\Lambda x; 0) = \Lambda^* \partial f(\Lambda x) \tag{2.8}$$

(1) Recall the equivalent definition $\partial(f\Lambda)(x) = \partial(f\Lambda)'(x; 0) = \{x^* \in X^* \mid (f\Lambda)'(x; y) \geq \langle x^*, y\rangle \,\forall\, y \in X\}$ for subdifferentials at a point. Then use the fact that $(f\Lambda)'(x; z) = f'(\Lambda x; \Lambda z)$ for $x, z \in X$ by

$$(f\Lambda)'(x; z) = \lim_{\lambda \downarrow 0} \lambda^{-1}((f\Lambda)(x + \lambda z) - (f\Lambda)(x)) = \lim_{\lambda \downarrow 0} \lambda^{-1}(f(\Lambda x + \lambda\Lambda z) - f(\Lambda x)) = f'(\Lambda x; \Lambda z) \tag{2.9}$$

(2) We show both directions of the set equality.
($\subseteq$) Recall that $f'(\Lambda x; \cdot)$ is convex, and that $x^* \in \partial(f'(\Lambda x; \cdot)\Lambda)(0)$ means $x^* \in \mathrm{dom}((f'(\Lambda x; \cdot)\Lambda)^*)$. Applying Theorem 2.3 then gives a $y^* \in Y^*$ with $x^* = \Lambda^* y^*$ and $(f'(\Lambda x; \cdot)\Lambda)^*(x^*) = f'(\Lambda x; \cdot)^*(y^*)$. For the direction, it suffices to prove that $y^* \in \partial f'(\Lambda x; 0) = \partial f(\Lambda x)$. By the characterization of subgradients for convex functions, it further suffices to prove $f'(\Lambda x; \Lambda x) + (f'(\Lambda x; \cdot))^*(y^*) = \langle y^*, \Lambda x\rangle$. But we have

$$\langle y^*, \Lambda x\rangle = \langle x^*, x\rangle = f'(\Lambda x; \Lambda x) + (f'(\Lambda x; \cdot)\Lambda)^*(x^*) = f'(\Lambda x; \Lambda x) + f'(\Lambda x; \cdot)^*(y^*) \tag{2.10}$$

Reversing the steps then gives the argument.
($\supseteq$) Let $\Lambda^* y^* \in \Lambda^* \partial f'(\Lambda x; 0)$ for a $y^* \in Y^*$ and let $x' \in X$ be arbitrary. Then we have

$$(f'(\Lambda x; \cdot)\Lambda)(x') = f'(\Lambda x; \Lambda x') \geq \langle y^*, \Lambda x'\rangle = \langle \Lambda^* y^*, x'\rangle \tag{2.11}$$

so $\Lambda^* y^* \in \partial(f'(\Lambda x; \cdot)\Lambda)(0)$.

∎

Remark 2.4. Note that the statement from [2, p. 201] misses the fact that f needs to at least be proper. If f is not proper, i.e. $-\infty \in \mathrm{Im}(f)$ or $f = \infty$, and assuming that we have defined the subdifferential analogously for arbitrary functions of form $X \to \overline{\mathbb{R}}$, then we can distinguish three cases.

(i) Suppose $f = -\infty$. Then $\partial(f\Lambda)(x) = X^*$, but $\Lambda^* \partial f(\Lambda x) = \mathrm{Im}(\Lambda^*)$. Unless Λ^* is an epimorphism, the equality does not hold.
(ii) Now suppose $-\infty \in \mathrm{Im}(f)$ and that there is a point $x' \in X$ with $f(x') > -\infty$. Then $\partial f(x') = \emptyset$ and $\partial f(x_{-\infty}) = X^*$ generally for any $x_{-\infty} \in f^{-1}(\{-\infty\})$. Note that we let x' and $x_{-\infty}$ be loose. We shall skip a full characterization of this case depending on $\mathrm{Im}(\Lambda)$ here.
(iii) For $f = \infty$, we apply the same argument as in (i).

3 Supremum Functions We look at descriptions of the subderivative of a supremum function over a family of convex functions.

Lemma 3.1. Let $\{f_s\colon X \to \overline{\mathbb{R}}\}_{s \in S}$, with S an index set, be a family of convex functions and $\hat{f}\colon X \to \overline{\mathbb{R}}, x \mapsto \sup_{s \in S} f(x)$. Then $\hat{f}$ is convex.

Proof. Consider

$$\operatorname{epi}(\hat{f}) = \left\{(\alpha, x) \in \mathbb{R} \times X \mid \alpha \geq \hat{f}(x) = \sup_{s \in S} f_s(x)\right\} \tag{3.12}$$

$$= \bigcap_{s \in S} \{(\alpha, x) \in \mathbb{R} \times X \mid \alpha \geq f_s(x)\} = \bigcap_{s \in S} \operatorname{epi}(f_s) \tag{3.13}$$

Arbitrary cuts of convex sets are convex due to the fact that if two points are contained in all convex sets, the line between them is also contained, so $\hat{f}$ is convex. ■

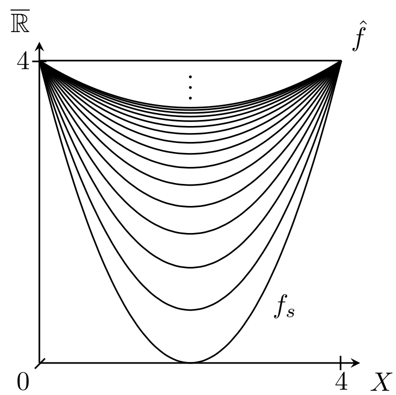

FIGURE 2. Illustration of the proof argument for Lemma 3.1 by a sequence of parabolas. Consider for this case $S = [0, 4] = X$ and for any $s \in S$ the function f_s to be the parabola obtained by Lagrange Interpolation on the points $\{(0, 4), (2, s), (4, 4)\}$.

Theorem 3.2 ([2, pp. 201-204])**.** Let S be a compact topological space and $f \colon S \times X \to \overline{\mathbb{R}}$, s.t. for any $(s, x) \in S \times X$, $f|_{\{s\} \times X}$ is convex and proper, and that $f|_{S \times \{x\}}$ is upper semicontinuous, where we each omit the respective fixed argument. Let $f_s \coloneqq f|_{\{s\} \times X}$ and set

$$\hat{f} \colon X \to \overline{\mathbb{R}}, x \mapsto \sup_{s \in S} f_s(x) \text{ and } S_0 \colon X \to \mathcal{P}(S), x \mapsto \{s \in S \mid f_s(x) = \hat{f}(x)\}. \tag{3.14}$$

In both of the following statements, the closures are taken wrt. the weak*-topology of X^*.

(i) For any $x \in X$

$$\overline{\operatorname{conv}}\left(\bigcup_{s \in S_0(x)} \partial f_s(x)\right) \subseteq \partial \hat{f}(x) \tag{3.15}$$

(ii) If for all $s \in S$, f_s is continuous at a point $x_0 \in X$, then

$$\overline{\operatorname{conv}}\left(\bigcup_{s \in S_0(x_0)} \partial f_s(x_0)\right) = \partial \hat{f}(x_0) \tag{3.16}$$

Proof. (i) Apply Lemma 3.1 to $\hat{f}$ to conclude its convexity. For $x \in X$ and $s \in S_0(x)$, we have $f_s(x) = \hat{f}(x)$ and $\partial f_s(x) \subseteq \partial \hat{f}(x)$, as for any $x^* \in \partial f_s(x)$ and $z \in X$

$$\hat{f}(x) + \langle x^*, z - x\rangle = f_s(x) + \langle x^*, z - x\rangle \leq f_s(z) \leq \hat{f}(z) \tag{3.17}$$

Since s was arbitrary, $\bigcup_{s \in S_0(x)} \partial f_s(x) \subseteq \partial \hat{f}(x)$. Since $\partial \hat{f}(x)$ is convex and weak*-closed [2, p. 198], we have

$$\overline{\operatorname{conv}}\left(\bigcup_{s \in S_0(x)} \partial f_s(x)\right) \subseteq \partial \hat{f}(x) \tag{3.18}$$

which was the claim.

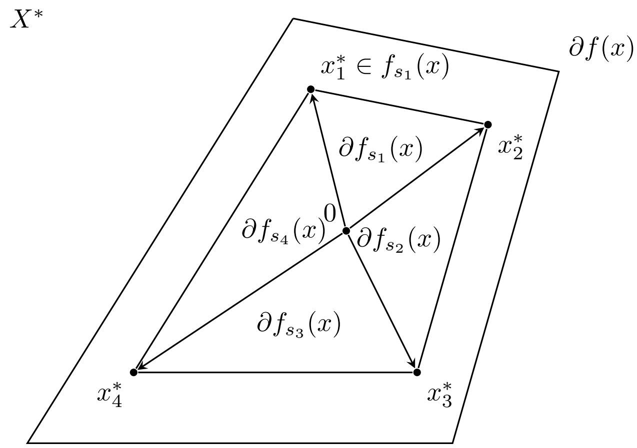

FIGURE 3. Illustration of the statement in Theorem 3.2 on an example of a 4-point space S and triangle-formed subdifferentials.

(ii) Our overall strategy is a proof of equality by contradiction. We first prove the existence of a set-separating functional (a), then we use a separation theorem to obtain a point $x \in X$ with a useful property (b) and then we argue that we can wlog. assume $\hat{f}(x_0 + x) < \infty$ (c). Lastly, we prove, that the first assumption violates the upper semicontinuity of $f_{S\times\{x_0\}}$ (d).

Theorem 3.3 ([2, pp. 164-165])**.** Let $A \subseteq X$ be closed, convex and let $x \notin A$. Then there exists a functional $x^* \in X^*$, s.t. $\langle x^*, y\rangle \leq \langle x^*, x\rangle - \varepsilon$ for a fixed $\varepsilon \in \mathbb{R}_{>0}$ and any $y \in A$, strongly separating A and $\{x\}$.

(a) Consider, that
- for all $s \in S$, $\partial f_s(x_0) \neq \emptyset$ by [2, p. 199], as the functions are continuous there, and
- $S_0(x_0) \neq \emptyset$, as by Weierstrass [2, p. 13], $f|_{S\times\{x_0\}}$ attains a maximum $f_s(x_0)$ for some $s \in S$.

The latter may not be true for any $x \in X$, it is only true since every f_s, $s \in S$, is finite at x_0. Set $Q \coloneqq \overline{\operatorname{conv}}(\bigcup_{s\in S_0(x_0)} \partial f_s(x_0))$. By the preceding argument we have $Q \neq \emptyset$. Thus, we can suppose $Q \neq \partial\hat{f}(x_0)$ and let $x^* \in \partial\hat{f}(x_0) \setminus Q$.
(b) Q is convex and closed, and $x^* \notin Q$. Theorem 3.3 gives the existence of some $\hat{x} \in X^{**}$ and $\varepsilon \in \mathbb{R}_{>0}$, s.t. as discussed in the preliminaries in Section 1, an $x \in X$ fulfills

$$\hat{x}(x^*) = x^*(x) = \langle x^*, x\rangle \geq \sup_{z^*\in Q} \langle z^*, x\rangle + \varepsilon \tag{3.19}$$

(c) Note that scaling by a constant does not change the separation criterion for x. So we may find such a constant. Since the functions $f|_{\{s\}\times X}$ for a $s \in S$ are continuous in x_0, as well as proper and convex, they are especially finite there, and thus $x_0 \in \operatorname{dom}(\hat{f})$. The goal is first to prove that there is some $\lambda \in \mathbb{R}_{>0}$, s.t. $\hat{f}(x_0 + \lambda x) < \infty$, s.t. replacing x by λx gives $\hat{f}(x_0), \hat{f}(x_0 + x) < \infty$.
We proceed in a couple of steps. Let $\varepsilon' \in \mathbb{R}_{>0}$.
- For every $s \in S$, there is a $\lambda_s \in \mathbb{R}_{>0}$ with $f_s(x_0+\lambda_s x) \leq f_s(x_0)+\varepsilon'$ by continuity of f_s and by the seminorm structure of locally convex TVS [3, p. 426], more exactly by the fact that there is a null environment basis element that is circular.
- Fix an $s \in S$. $f|_{S\times\{x_0+\lambda_s x\}}$ is upper semicontinuous, so there is an open neighborhood $s \in U_s \subseteq S$ with $f_{s'}(x_0 + \lambda_s x) \leq f_s(x_0 + \lambda_s x) + \varepsilon' \leq f_s(x_0) + 2\varepsilon'$ for any $s' \in U_s$.
- $\{U_s\}_{s\in S}$ is thus an open cover of S. By compactness, there are $s_1, ..., s_m \in S$ for $m \in \mathbb{N}_{\geq 1}$, s.t. $\{U_{s_1}, ..., U_{s_m}\}$ covers S. Choosing $\lambda \coloneqq \min\{\lambda_{s_1}, ..., \lambda_{s_m}\} > 0$ gives $f_s(x_0 + \lambda x) \leq \hat{f}(x_0) + 2\varepsilon'$ for any $s \in S$, which is the desired finiteness result.

Replace now for everything following, as announced, x with λx. As $\hat{f}$ is convex, we can further directly conclude $\hat{f}(x_0+tx) \leq (1-t)\hat{f}(x_0)+t\hat{f}(x_0+x) < \infty$ for any $t \in [0,1]$ by Jensens Inequality. In other words, $x_0 + [0,1]x \subseteq \mathrm{dom}(\hat{f})$.

(d) At last, we proceed with the main contradiction, which is again divided up into multiple steps.

1. We claim $\lim_{t\to 0} \hat{f}(x_0+tx) = \hat{f}(x_0)$. ($\leq$) Let $s_0 \in S_0(x_0)$. We then have for any $t \in [0,1]$, that $f_{s_0}(x_0+tx) \leq \hat{f}(x_0+tx)$. Taking the limit, as $f|_{\{s_0\}\times X}$ is continuous in x_0, gives $\hat{f}(x_0) \leq \lim_{t\to 0} \hat{f}(x_0+tx)$. ($\geq$) Use for a fixed $t \in [0,1]$ Jensens Inequality to obtain $\hat{f}(x_0+tx) \leq (1-t)\hat{f}(x_0)+t\hat{f}(x_0+x)$. Taking the limit gives $\lim_{t\to 0} \hat{f}(x_0+tx) \leq \hat{f}(x_0)$. The sandwich rule of limit calculus now gives the statement.

2. Let $t \in (0,1)$ be fixed and choose $s_t \in S$ with $f_{s_t}(x_0+tx) = \hat{f}(x_0+tx)$. By Jensens inequality

$$\hat{f}(x_0+tx) = f_{s_t}(x_0+tx) \leq (1-t)f_{s_t}(x_0) + tf_{s_t}(x_0+x) \tag{3.20}$$

Since $f_{s_t}(x_0+x) \leq \hat{f}(x_0+x) < \infty$, we thus get

$$\hat{f}(x_0+tx) - t\hat{f}(x_0+x) \leq (1-t)f_{s_t}(x_0) \tag{3.21}$$

Taking the limit $t \to 0$ and using $\lim_{t\to 0} \hat{f}(x_0+tx) = \hat{f}(x_0)$ from the argument in 1., as well as the product rule from limit calculus on the right side, we obtain $\hat{f}(x_0) \leq \lim_{t\to 0} f_{s_t}(x_0)$. $\lim_{t\to 0} f_{s_t}(x_0) \leq \hat{f}(x_0)$ follows directly by the monotonicity of limits. So $\lim_{t\to 0} f_{s_t}(x_0) = \hat{f}(x_0)$ by the sandwhich rule of limit calculus.

3. Let $s_0 \in S$ be a cluster point of $\{s_t\}_{t\in(0,1)}$, which we recall exists, due to S being compact and $\{s_t\}_{t\in(0,1)}$ being infinite [2, p. 12]. $f|_{S\times\{x_0\}}$ is upper semicontinuous, so consider for any ε' an open neighborhood $U_{\varepsilon'} \subseteq S$ of S_0 with $f_s(x_0) \leq f_{s_0}(x_0)+\varepsilon'$ for any $s \in U_{\varepsilon'}$. As s_0 is a cluster point, there is some s_t with $t \in [0,1]$ with $s_t \in U_{\varepsilon'}$. Using this fact now with any monotonically decreasing null sequence $(\varepsilon'_n \in \mathbb{R}_{\geq 0})_{n\in\mathbb{N}}$, we obtain $\lim_{t\to 0} f_{s_t}(x_0) = \hat{f}(x_0) \leq f_{s_0}(x_0) \leq \hat{f}(x_0)$, where the latter inequality follows from the definitions. This gives the claimed equality by transitivity. Especially, $s_0 \in S_0(x_0)$, so $\partial f_{s_0}(x_0) \subseteq Q$.

4. We have the inequalities

$$\frac{f_{s_t}(x_0+tx) - f_{s_t}(x_0)}{t} \geq \frac{\hat{f}(x_0+tx) - \hat{f}(x_0)}{t} \tag{3.22}$$

$$\geq \hat{f}'(x_0;x) \geq \langle x^*, x\rangle \geq \sup_{z^*\in\partial f_{s_0}(x_0)\subseteq Q} \langle z^*, x\rangle + \varepsilon = f'_{s_0}(x_0;x) + \varepsilon \tag{3.23}$$

$$\rightsquigarrow f_{s_t}(x_0+tx) \geq f_{s_t}(x_0) + t(f'_{s_0}(x_0;x)+\varepsilon) \tag{3.24}$$

and

$$\frac{f_{s_0}(x_0+t_1x) - f_{s_0}(x_0)}{t_1} \leq f'_{s_0}(x_0;x) + \frac{\varepsilon}{2} \tag{3.25}$$

with a sufficient choice of t_1, using the limit in the definition of the directional derivative. For any $t \in (0,t_1)$, we now have with Jensens Inequality

$$\left(1-\frac{t}{t_1}\right) f_{s_t}(x_0) + \frac{t}{t_1} f_{s_t}(x_0+t_1x) \geq f_{s_t}\left(\left(1-\frac{t}{t_1}\right)x_0 + \frac{t}{t_1}(x_0+t_1x)\right) \tag{3.26}$$

$$= f_{s_t}(x_0+tx) \geq f_{s_t}(x_0) + t(f'_{s_0}(x_0;x) + 2\cdot\varepsilon/2) \tag{3.27}$$

$$\geq f_{s_t}(x_0) + t\left(\frac{f_{s_0}(x_0+t_1x) - f_{s_0}(x_0)}{t_1} + \frac{\varepsilon}{2}\right) \tag{3.28}$$

Dividing both sides by t/t_1 and reordering after $f_{s_t}(x_0)$, we have

$$f_{s_t}(x_0+t_1x) \geq f_{s_t}(x_0+t_1x) + \varepsilon t_1/2 \tag{3.29}$$

Letting $t \to 0$, we thus have with 3., where we observe that the argument can be repeated with x_0+t_1x analogously, that

$$\lim_{t\to 0} f_{s_t}(x_0+t_1x) \geq f_{s_0}(x_0+t_1x) + \varepsilon t_1/2 \tag{3.30}$$

contradicting the upper semicontinuity of $f|_{S\times\{x_0+t_1x\}}$ ↯.

■

Remark 3.4. Some remarks about this version of the proof, especially in comparison with the book.

(i) We have additionally assumed that f_s is proper for any $s \in S$, as the use of [2, p. 199] requires that.
(ii) In part (c) of the proof, we defined U_s differently by taking neighborhoods directly from the definition of upper semicontinuity. In [2, pp. 201-204], the sets $\{s' \in S \mid f(s', x_0 + \lambda_s x) < \hat{f}(x_0) + 2\}$, claimed as being open, were used, leading to the same conclusion however.
(iii) Further, the use of the partial subdifferential notation has been omitted here, instead f_s was used directly [2, p. 47].
(iv) As a personal note, the TVS X cannot have discrete topology, as the field is $\mathbb{R}$ [3, p. 426].

4 Situation in $\mathbb{R}^n$ Let $n \in \mathbb{N}_{\geq 1}$ be fixed for the following.

Subdifferentiability in Finite Dimensions We first consider the existence of subdifferentials in finite dimensions.

Lemma 4.1 ([2, p. 188]). Let $f\colon \mathbb{R}^n \to \overline{\mathbb{R}}$ be convex and proper.
(i) f is continuous with respect to $\operatorname{aff}(\operatorname{dom}(f))$ on $\operatorname{ri}(\operatorname{dom}(f))$.
(ii) f^* is proper.

Remark 4.2. Continuity with respect to a set is to be understood as treating the neighborhoods for continuity as subsets of the respected set. Recall especially, that we defined the relative interior as [4, p. 44]

$$\operatorname{ri}(A) = \{a \in \operatorname{aff}(A) \mid \exists\, \varepsilon \in \mathbb{R}_{>0}\colon B(a, \varepsilon) \cap \operatorname{aff}(A) \subseteq A\} \tag{4.31}$$

for some set $A \subseteq \mathbb{R}^n$ and $B(a, \varepsilon) = \{a' \in \mathbb{R}^n \mid \|a - a'\|_{\ell_2} \leq \varepsilon\}$.

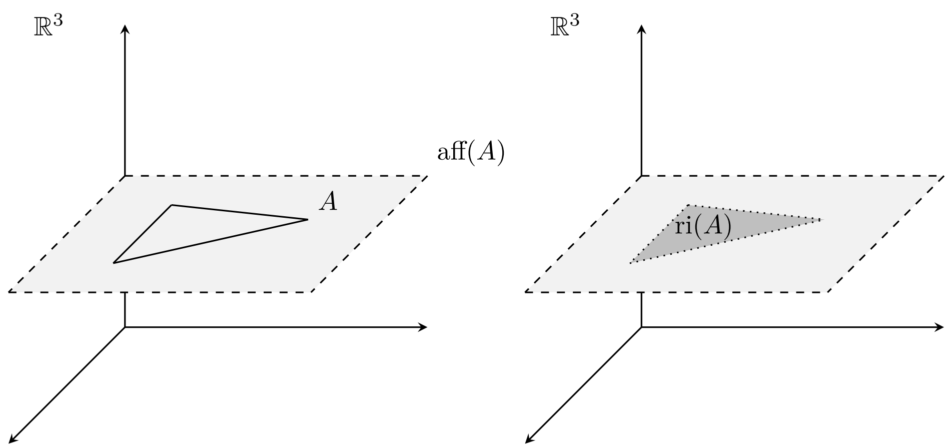

Figure 4. Illustration of the concept of the relative interior.

Theorem 4.3 ([2, p. 204])**.** Let $f\colon \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper and convex. Then f is subdifferentiable in $\operatorname{ri}(\operatorname{dom}(f))$.

Proof. Take $x \in \operatorname{ri}(\operatorname{dom}(f))$. Consider the following chain of applications of proven theorems.
(1) Since f is convex and proper, Lemma 4.1 (i) gives the continuity at x with respect to $\operatorname{aff}(\operatorname{dom}(f))$.
(2) Theorem 2.2 (ii) gives the finiteness of $f'(x; \cdot)$ with respect to $\operatorname{aff}(\operatorname{dom}(f))$.
(3) $f'(x; \cdot)$ is convex and proper by that, so applying Lemma 4.1 (i) again gives, that $f'(x; \cdot)^*$ is proper.
(4) The effective domain of $f'(x; \cdot)^*$ is the subdifferential $\partial f'(x; 0) = \partial f(x)$ [2, p. 196]. Now, since $f'(x; \cdot)^*$ is proper, $\emptyset \neq \operatorname{dom}(f'(x; \cdot)^*) = \partial f(x)$, concluding the proof.

■

Remark 4.4. In [2, p. 172], it is stated that the supremum of the Young-Fenchel transform can only be taken for $\operatorname{dom}(f'(x; \cdot))$, which is false. Generally, taking it over $f'(x; \cdot)^{-1}([0, \infty])$ suffices, which is $\mathbb{R}^n$ for a proper $f'(x; \cdot)$.

Representations of Subgradients in Finite Dimensions The next main theorem will study representations of the functionals in subdifferentials.

Lemma 4.5 ([2, pp. 185-186]). Let $A \subseteq \mathbb{R}^n$ be bounded and closed. Then $\mathrm{conv}(A) = \overline{\mathrm{conv}}(A)$.

Lemma 4.6 ([2, p. 199]). For a proper convex function $f\colon X \to \overline{\mathbb{R}}$, which is continuous at a point x_0, $\partial f(x_0)$ is non-empty, weakly* bounded and weakly* compact.

Remark 4.7. Weakly* boundedness means that for a set $A \subseteq X^*$ and any weak* neighborhood of zero U in X^*, there is some $\varepsilon \in \mathbb{R}_{>0}$ with $\varepsilon A \subseteq U$.

Lemma 4.8. Let $X, S, f, f_s, \hat{f}, S_0, x_0$ be defined as in the setting of Theorem 3.2 (ii). Then $S_0(x_0)$ is compact.

Proof. Since S is compact, it suffices to show that $S_0(x_0)$ is closed [5, p. 165]. Let $s_* \in S$ be a cluster point of $S_0(x_0)$. By upper semicontinuity, there is for any $\varepsilon \in \mathbb{R}_{>0}$ an open neighborhood $s_* \in U_\varepsilon \subseteq S$, s.t. $f_s(x_0) \leq f_{s_*}(x_0) + \varepsilon \,\forall\, s \in U_\varepsilon$. Since s_* is a cluster point of $S_0(x_0)$, U_ε contains a point $s' \in (U_\varepsilon \setminus \{s_*\}) \cap S_0(x_0)$. So $\hat{f}(x_0) = f_{s'}(x_0) \leq f_{s_*}(x_0) + \varepsilon$. Since ε was chosen arbitrarily, $\hat{f}(x_0) = f_{s_*}(x_0)$ and thus $s_* \in S_0(x_0)$. ∎

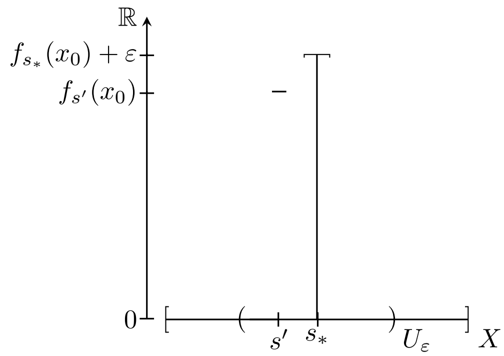

FIGURE 5. Illustration of the argument of the proof of Lemma 4.8 using for X a closed interval. Tightening both U_ε and the gap between $f_{s_*}(x_0) + \varepsilon$ and $f_{s'}(x_0)$, of which the latter is constant as $\hat{f}(x_0)$, we obtain the statement.

Remark 4.9. The proof of Lemma 4.8 is due to Prof. Dr. Marita Thomas.

Theorem 4.10 ([2, pp. 204-205])**.** Let $X, S, f, f_s, \hat{f}, S_0, x_0$ for $X = \mathbb{R}^n$ and any $s \in S$ be defined as in the setting of Theorem 3.2 (ii). Then every $y \in \partial \hat{f}(x_0)$ can be represented as a convex combination of form

$$y = \sum_{i=1}^{r} \alpha_i y_i \tag{4.32}$$

with $r \in \mathbb{N}$, $1 \leq r \leq n+1$, $\sum_{i=1}^{r} \alpha_i = 1$ and $s_i \in S_0(x_0)$, $(\alpha_i, y_i) \in \mathbb{R}_{>0} \times \partial f_{s_i}(x_0)$ for any $i \in \mathbb{N}$, $1 \leq i \leq r$.

Proof. The statement corresponds to the statement, that for $P := \bigcup_{s \in S_0(x_0)} \partial f_s(x_0)$, P is bounded and closed, because in that case $\partial \hat{f}(x) = \overline{\mathrm{conv}(P)} = \mathrm{conv}(P)$ by Theorem 3.2 (ii) and Lemma 4.5, since $(\mathbb{R}^n)^* \cong \mathbb{R}^n$. Because in that case, the convex combinations $y = \sum_{i=1}^{r} \alpha_i y_i$ correspond exactly to all elements of $\mathrm{conv}(P)$. It only remains to show, that P is (a) bounded and (b) closed.

(a) Using part (c) of the proof of Theorem 3.2, we find for any $x \in X$ a $\lambda \in \mathbb{R}_{>0}$ with $\hat{f}(x_0 + \lambda x) < \infty$. So $\hat{f}'(x_0; \cdot)$ is finite, since $\hat{f}'(x_0; x) = \inf_{\lambda' \in \mathbb{R}_{>0}} (\hat{f}(x_0 + \lambda' x) - \hat{f}(x_0))/\lambda' \leq (\hat{f}(x_0 + \lambda x) - \hat{f}(x_0))/\lambda < \infty$, where we also used, that $\hat{f}$ is convex and proper. $\hat{f}'(x_0; \cdot)$ is thus convex and proper and applying Lemma 4.1 (i) gives, that it is continuous, as $\mathrm{dom}(f'(x_0; \cdot)) = \mathbb{R}^n$. By that, $\hat{f}$ must be continuous in x_0, as it would otherwise be unbounded in a neighborhood there and $f'(x_0; \cdot)$ would thus not be continuous. Lemma 4.6 gives the weakly* boundedness of $\partial \hat{f}(x_0) \supseteq P$, thus also of P.

(b) Let $(z_n \in P)_{n\in\mathbb{N}_{\geq 1}}$, s.t. $z_n \to z \in \mathbb{R}^n$. Denote accordingly $z_k \in \partial f_{s_k}(x_0)$ with $s_k \in S_0(x_0)$ for every $k \in \mathbb{N}_{\geq 1}$. As $S_0(x_0)$ is compact by Lemma 4.8, $(s_k)_{k\in\mathbb{N}_{\geq 1}}$ has a cluster point $s_0 \in S_0(x_0)$ by sequential compactness [5, pp. 179-180]. With the upper semicontinuity of $\hat{f}$, we thus have for any $x \in \mathbb{R}^n$

$$f_{s_0}(x) - f_{s_0}(x_0) \overset{(1)}{\geq} f_{s_0}(x) - \hat{f}(x_0) \overset{(2)}{\geq} \limsup_{k\to\infty} f_{s_k}(x) - \hat{f}(x_0) \tag{4.33}$$

$$\overset{(1)}{=} \limsup_{k\to\infty} f_{s_k}(x) - f_{s_k}(x_0) \geq \lim_{k\to\infty} \langle z_k, x - x_0\rangle = \langle z, x - x_0\rangle \tag{4.34}$$

(1) $\hat{f}(x_0) = f_{s_0}(x_0) = f_{s_k}(x_0)$ for any $k \in \mathbb{N}_{\geq 1}$, as $s_0, s_k \in S_0(x_0)$.
(2) Consider that s_0 is a cluster point of $\{s_k\}_{k\in\mathbb{N}_{\geq 1}}$ and the upper semicontinuity of $f|_{\{s_0\}\times X}$, as in previous arguments.
so $z \in \partial f_{s_0}(x_0) \subseteq P$, concluding the statement.

■

Remark 4.11. The choices of r and the parameters α_i stem from the use of Lemma 4.5, a corollary of Carathéodory's Theorem.

References

[1] K. Jänich, *Topologie*, ISBN: 9783540213932.
[2] A.D. Ioffe and V.M. Tihomirov, *Theory of Extremal Problems*, ISBN: 9780444851673.
[3] D. Werner, *Funktionalanalysis*, ISBN: 9783662554067.
[4] R.T. Rockafellar, *Convex Analysis*, ISBN: 9780691015866.
[5] J. R. Munkres, *Topology*, ISBN: 9780131816299.