D. Basics in convex analysis

Lemma D.1 (duality)**.** *Let $\mathcal{X}$ be a locally convex topological vector space. If $f : \mathcal{X} \to (-\infty, \infty]$ is convex and lower semicontinuous, then $f^{**}(x) = f(x)$.*

Proof. We should assume, without loss of generality, that f is not identically ∞, for the lemma holds obviously otherwise. Define

$$\mathcal{E} = \{(x, \alpha) \in \mathcal{X} \times \mathbb{R} : f(x) \leq \alpha\},$$
$$\mathcal{E}^* = \{(\lambda, \alpha) \in \mathcal{X}^* \times \mathbb{R} : f^*(\lambda) \leq \alpha\},$$

which are convex subsets in the locally convex topological vector spaces $\mathcal{X} \times \mathbb{R}$ and $\mathcal{X}^* \times \mathbb{R}$, respectively.

It is not hard to verify by definition that $f^{**} \leq f$. Hence, it suffices to prove the other inequality. Equivalently, it asserts that if $f(x) > \alpha$, then there is $(\lambda, \beta) \in \mathcal{X}^*$ such that $\langle \lambda, x\rangle - \beta > \alpha$. Then, observe that $A = \{(x, \alpha)\}$ is compact and $B = \mathcal{E}$ is closed by lower semicontinuity of f and nonempty by the fact f is not identically ∞. Moreover, under the assumption $f(x) > \alpha$, $A \cap B = \emptyset$. These altogether allow us to apply Hahn-Banach separation theorem (specifically, Theorem C.3) to yield some $\eta \in \mathcal{X}^*$ and $\rho \in \mathbb{R}$ satisfying

$$\gamma := \sup_{(y,\beta)\in\mathcal{E}} \langle \eta, y\rangle - \rho\beta < \langle \eta, x\rangle - \rho\alpha \tag{D.1}$$

where $\rho \geq 0$ must hold as there exists $y \in \mathcal{X}$ with $f(y) < \infty$.

If $\rho > 0$, then (D.1) implies that $\alpha < \langle \rho^{-1}\eta, x\rangle - \rho^{-1}\gamma$. Moreover, $(\rho^{-1}\eta, \rho^{-1}\gamma) \in \mathcal{E}^*$ as desired, for if otherwise, $f^*(\rho^{-1}\eta) > \rho^{-1}\gamma$ contradicting the definition of γ.

If $\rho = 0$, then for all $(\mu, \beta) \in \mathcal{E}^*$, define $(\mu_\delta, \beta_\delta) := \left(\frac{\eta}{\delta} + \mu, \frac{\gamma}{\delta} + \beta\right)$ for $\delta > 0$. In fact, $(\mu_\delta, \beta_\delta) \in \mathcal{E}^*$ since

$$f^*(\mu_\delta) - \beta_\delta \leq \sup_{y\in\mathcal{X}}[\langle \mu_\delta, y\rangle - f(y) - \beta_\delta]$$
$$\leq \sup_{y\in\mathcal{X}}\left[\frac{1}{\delta}(\langle \eta, y\rangle - \gamma) + (\langle \mu, y\rangle - f^*(\mu)) - f(y)\right] \leq 0.$$

Moreover,

$$[\langle \eta_\delta, x\rangle - f^*(\eta_\delta)] \geq \lim_{\delta\to 0}\left[\frac{1}{\delta}(\langle \eta, x\rangle - \gamma) + (\langle \mu, x\rangle - f^*(\mu))\right] = \infty,$$

yielding the desired $\lambda = \eta_\delta$ when δ is sufficiently small. □

Lemma D.2. *Suppose $f : \mathbb{R}^d \to [0, \infty]$ is a convex, lower semicontinuous function with $\inf_{\lambda\in\mathbb{R}^d} f(\lambda) = 0$ and $0 \in \operatorname{ri} \mathcal{D}_{f^*}$, then $f(\eta) = 0$ for some $\eta \in \mathbb{R}^d$*

Proof. Note that $f^*(0) = 0$, and hence the lemma, by duality (Lemma D.1), is equivalent to the existence of $\eta \in \mathbb{R}^d$ such that $\langle \eta, x\rangle \leq f^*(x)$ for all $x \in \mathbb{R}^d$. This latter claim, by convexity of f^*, is equivalent to

$$\langle \eta, x\rangle \leq \lim_{\delta\to 0+} \frac{f^*(\delta x) - f^*(0)}{\delta} = \inf_{\delta>0} \frac{f^*(\delta x) - f^*(0)}{\delta} =: g(x) \text{ for all } x \in \mathbb{R}^d.$$

To prove the claim, define

$$A = \overline{\{(x, \alpha) \in \mathbb{R}^d \times \mathbb{R} : g(x) \leq \alpha\}} \text{ and } B = \{(0, -1)\},$$

Obviously, B is convex, compact, and nonempty. On the other hand, A is nonempty, convex, and $A \cap B = \emptyset$. Obviously, $(0,0) \in A$. To prove convexity, note that g is a convex function and closure of a convex set is convex. Indeed,

$$g(tx + (1-t)y) = \lim_{\delta \to 0+} \frac{f^*(t\delta x + (1-t)\delta y)}{\delta}$$
$$\leq \lim_{\delta \to 0+} t \cdot \frac{f^*(\delta x)}{\delta} + (1-t) \cdot \frac{f^*(\delta y)}{\delta} = tg(x) + (1-t)g(y).$$

Finally, to prove $A \cap B = \emptyset$, we first show that $0 \in \text{ri } \mathcal{D}_g = \mathcal{D}_g$. Under the circumstances, we deduce that $g|_{\mathcal{D}_g}$ is continuous at $0 \in \text{ri } \mathcal{D}_g = \mathcal{D}_g$ and thus $A \cap B = \emptyset$ follows naturally. To this end, note that $0 \in \text{ri } \mathcal{D}_{f^*}$ implies $0 \in \text{ri } \mathcal{D}_g$. Indeed, $g(0) = 0$ and if $x \in \mathcal{D}_g$, then $f^*(x) < \infty$ and for all small $\varepsilon > 0$,

$$g(-\varepsilon x) = \inf_{\delta > 0} \frac{f^*(\delta(-\varepsilon)x)}{\delta} \leq f^*(-\varepsilon x) < \infty.$$

In particular, the above implies $0 \in \text{ri } \mathcal{D}_g$. Moreover, $0 \in \text{ri } \mathcal{D}_g$ implies also $\text{ri } \mathcal{D}_g = \mathcal{D}_g$ since $x \in \mathcal{D}_g$ if and only if $tx \in \mathcal{D}_g$ for all $t > 0$.

We now may apply Theorem C.3 to A and B to yield some $\mu \in \mathcal{X}$ such that

$$\langle \mu, 0 \rangle + \rho = \rho > \sup_{(x,\alpha) \in A} \langle \mu, x \rangle - \rho\alpha,$$

where $\rho > 0$ since $(0,0) \in A$. Hence, for every $x \in \mathcal{X}$ with $f^*(x) < \infty$, we have that

$$\langle \rho^{-1}\mu, \varepsilon x \rangle \leq g(\varepsilon x) + 1 \text{ for all } \varepsilon > 0 \Rightarrow \langle \rho^{-1}\mu, x \rangle \leq g(x).$$

The proof is concluded by choosing $\eta = \rho^{-1}\mu$. □

Lemma D.3. *Let f be an essentially smooth, convex function. If $f(0) = 0$ and $f^*(x) = 0$ for some $x \in \mathbb{R}^d$, then $0 \in \mathring{\mathcal{D}}_f$.*

Proof. Since $f(0) = 0$, it follows by convexity of f that

$$f(t\lambda) \leq tf(\lambda), \text{for all } t \in [0,1], \lambda \in \mathbb{R}^d.$$

Moreover, since $f^*(x) = 0$,

$$f(t\lambda) \geq \langle t\lambda, x \rangle - f^*(x) \geq -t|\lambda||x|.$$

Because f is essentially smooth, there exists a closed ball $\overline{B}_r(z) \subset \mathring{\mathcal{D}}_f$ in which f is differentiable. Hence,

$$M = \sup_{\lambda \in \overline{B}_r(z)} \{f(\lambda) \vee |\lambda||x|\} < \infty$$

Hence, for any $t \in (0,1]$ and any $\theta \in \overline{B}_{tr}(tz)$ different from tz, by the convexity of f,

$$f(\theta) - f(tz) \leq \frac{|\theta - tz|}{tr}(f(y) - f(tz)) \leq \frac{2tM}{tr}|\theta - tz|,$$

where $y = tz + \frac{tr}{|\theta - tz|}(\theta - tz) \in \overline{B}_{tr}(tz)$. Similarly, $f(\theta) - f(tz) \geq -\frac{2tM}{tr}|\theta - tz|$. Hence,

$$|f(\theta) - f(tz)| \leq \frac{2tM}{tr}|\theta - tz| \text{ for all } \theta \in \overline{B}_{tr}(tz).$$

Observe that $tz \in \mathring{\mathcal{D}}_f$ because of the convexity of $\mathcal{D}_f$. Hence, by assumption, $\nabla f(tz)$ exists, and by the preceding inequality, $|\nabla f(tz)| \le 2\frac{M}{r}$. Since f is steep, it follows by considering $t \to 0$, in which case $tz \to 0$, that $0 \in \mathring{\mathcal{D}}_f$. $\square$

We define the *approximate subgradient* as follows:

$$\begin{aligned}\partial_\varepsilon f(x) &= \{x^* \in \mathcal{X}^* : f(z) \ge (f(x) - \varepsilon) + \langle x^*, z - x\rangle \text{ for all } z \in \mathcal{X}\} \\ &= \{x^* \in \mathcal{X}^* : f(z) + f^*(x^*) - \langle x^*, x\rangle \le \varepsilon \text{ for all } z \in \mathcal{X}\}\end{aligned}$$

By Hahn-Banach separation theorem (Theorem C.2), such set is always non-empty as long as $x \in \mathcal{D}_f$.

Theorem D.4 (Brønsted–Rockafellar)**.** *Suppose $\mathcal{X}$ is a Banach space. Then, every lower semicontinuous convex function taking value in $(-\infty, \infty]$ which is not identically $+\infty$ satisfies the following:*

$$f(y) = \liminf_{x \to y} \overline{f}(x) \text{ for all } y, \text{where } \overline{f}(x) = \begin{cases} f(x) \text{ if } x \in \mathcal{D}_{\partial f}, \\ +\infty \;\text{ otherwise,}\end{cases}$$

$$f(y) = \sup\big\{ f(\overline{x}) + (y - \overline{x}, x^*) \mid \in \overline{x} \in \mathcal{D}_f, x^* \in \mathcal{D}_{\partial f} \big\} \text{ for all } y \in E.$$

Moreover, the conjugates f^* of such functions actually satisfy the stronger conditions derived from the above by restricting attention to the existence of subgradients of f^* belonging to $\mathcal{X}$ instead of $\mathcal{X}^{**}$.

Lemma D.5. *Assume that $\mathcal{X}$ is a Banach space and that $x^* \in \partial_\varepsilon f(x)$. Then, for any $\lambda > 0$, there exist vectors $\overline{x}$ and $\overline{x}^*$ such that $\|\overline{x} - x\| \le \lambda$, $\|\overline{x}^* - x^*\| \le \frac{\varepsilon}{\lambda}$ and $\overline{x}^* \in \partial f(x)$.*

Proof. Define a partial ordering $y \prec z$ on $\mathcal{D}_f$ if

$$\frac{\varepsilon}{\lambda}\, \|y - z\| \le [f(y) - \langle x^*, y\rangle] - [f(z) - \langle x^*, z\rangle].$$

By Zorn's lemma, there exists a maximal totally ordered subset M of $\big\{z \in \mathcal{D}_f : x\big\}$. Therefore,

$$f(z) - \langle x^*, z\rangle \searrow \rho \ge f(x) - \langle x^*, x\rangle - \varepsilon > -\infty \text{ as } z \nearrow \text{in } M.$$

implying that $z \to \overline{x}$ as $\overline{x} \nearrow$ by completeness, that $\overline{x} \in \mathcal{D}_f$, and that

$$\frac{\varepsilon}{\lambda}\, \|\overline{x} - z\| > [f(\overline{x}) - \langle x^*, \overline{x}\rangle] - [f(z) - \langle x^*, z\rangle] \text{ if } z \ne \overline{x}.$$

We now apply Hahn-Banach separation theorem (Theorem C.2) to

$$A = \{(y, \mu) : \mu \ge f(\overline{x} + y) - f(\overline{x}) - \langle x^*, y\rangle\} \text{ and } B = \Big\{(y, \mu) : \mu < -\frac{\varepsilon}{\lambda}\|y\|\Big\},$$

we find some $z^* \in \mathcal{X}^*$ such that

$$-\frac{\varepsilon}{\lambda}\, \|y\| \le \langle z^*, y\rangle \le f(\overline{x} + y) - f(\overline{x}) - \langle x^*, y\rangle,$$

which completes the proof if we take $\overline{x}^* = x^* + z^*$, assuring $\|\overline{x}^* - z^*\| \le \frac{\varepsilon}{\lambda}$ and $\overline{x}^* \in \mathcal{D}_{\partial f}$. $\square$

Proof of Theorem D.4. It suffices to show that $\liminf_{y \to x} \overline{f}(y) \le f(x)$ whenever $x \in \mathcal{D}_f$. Given any $\delta > 0$, choose any $x^* \in \partial_\varepsilon f$, where $\varepsilon = \frac{\delta}{2}$. Choose $\lambda > 0$ so small that $\lambda < \delta$ and $\lambda\|x^*\| < \frac{\delta}{2}$. Now let $\overline{x}$ and $\overline{x}^*$ be the vectors whose existence is guaranteed by Lemma D.5, meaning

$$f(\overline{x}) - f(x) \leq -(x^*, x - \overline{x}) \leq \|\overline{x} - x\| \, \|x^*\| \leq \lambda\left(\|x^*\| + \frac{\varepsilon}{\lambda}\right) < \delta.$$

This implies $\overline{x} \in \mathcal{D}_f$, $\|\overline{x} - x\| < \delta$ and $f(\overline{x}) < f(x) + \delta$. These proves the first part. The modified condition is proved using Lemma D.5 similarly.

The second equality is proved as follows. □

References

1. Komiya, H.: Elementary proof for Sion's minimax theorem. Kodai Mathematical Journal. 11, (1988). https://doi.org/10.2996/kmj/1138038812

Department of Mathematics, University of Exampleville, Tennessee, TN 59341
Email address: jdoe@math.ue.edu
URL: math.ue.edu/~jdoe