5

The Fenchel conjugate

Ουκ έστ' εραστὴσ όστισ ουκ αεὶ φιλεῖ
(He is not a lover who does not love forever)
(Euripides, "The Trojan Women")

In the study of a (constrained) minimum problem it often happens that another problem, naturally related to the initial one, is useful to study. This is the so-called duality theory, and will be the subject of the next chapter.

In this one, we introduce a fundamental operation on convex functions that allows building up a general duality theory. Given an extended real valued function f defined on a Banach space X, its Fenchel conjugate f^* is a convex and lower semicontinuous function, defined on the dual space X^* of X. After defining it, we give several examples and study its first relevant properties. Then we observe that we can apply the Fenchel conjugation to f^* too, and this provides a new function, again defined on X, and minorizing everywhere the original function f. It coincides with f itself if and only if $f \in \Gamma(X)$, and is often called the convex, lower semicontinuous relaxation (or regularization) of f. Moreover, there are interesting connections between the subdifferentials of f and f^*; we shall see that the graphs of the two subdifferentials are the same. Given the importance of this operation, a relevant question is to evaluate the conjugate of the sum of two convex functions. We then provide a general result in this sense, known as the Attouch–Brézis theorem.

5.1 Generalities

As usual, we shall denote by X a Banach space, and by X^* its topological dual.

Definition 5.1.1 Let $f\colon X \to (-\infty, \infty]$ be an arbitrary function. The *Fenchel conjugate* of f is the function $f^*\colon X^* \to [-\infty, \infty]$ defined as

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

We have that

$$(x^*, \alpha) \in \operatorname{epi} f^* \iff f(x) \geq \langle x^*, x \rangle - \alpha, \ \forall x \in X,$$

which means that the points of the epigraph of f^* parameterize the affine functions minorizing f. In other words, if the affine function $l(x) = \langle x^*, x \rangle - \alpha$ minorizes f, then the affine function $m(x) = \langle x^*, x \rangle - f^*(x^*)$ fulfills

$$l(x) \leq m(x) \leq f(x).$$

We also have that

$$\operatorname{epi} f^* = \bigcap_{x \in X} \operatorname{epi}\{\langle \cdot\,, x \rangle - f(x)\}.$$

Observe that even if f is completely arbitrary, its conjugate is a convex function, since $\operatorname{epi}\{\langle \cdot\,, x \rangle - f(x)\}$ is clearly a convex set for every $x \in X$. Furthermore, as $\operatorname{epi}\{\langle \cdot\,, x \rangle - f(x)\}$ is for all x, a closed set in $X^* \times \mathbb{R}$ endowed with the product topology inherited by the weak* topology on X^* and the natural topology on $\mathbb{R}$, it follows that for any arbitrary $f, \operatorname{epi} f^* \subset X^* \times \mathbb{R}$ is a closed set in the above topology.

A geometrical way to visualize the definition of f^* can be captured by observing that

$$-f^*(x^*) = \sup\{\alpha : \alpha + \langle x^*, x \rangle \leq f(x), \forall x \in X\}.$$

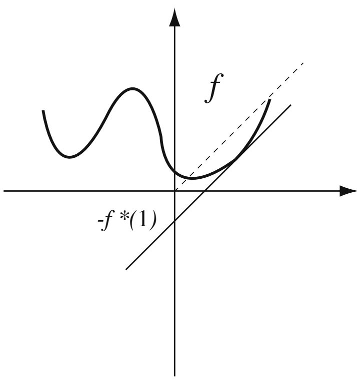


Figure 5.1.

For,

$$\begin{aligned} f^*(x^*) &= \inf\{-\alpha : \alpha + \langle x^*, x \rangle \leq f(x), \forall x \in X\} \\ &= -\sup\{\alpha : \alpha + \langle x^*, x \rangle \leq f(x), \forall x \in X\}. \end{aligned}$$

Example 5.1.2 Here we see some examples of conjugates.

(a) The conjugate of an affine function: for $a \in X^*$, $b \in \mathbb{R}$, let $f(x) = \langle a, x \rangle + b$; then

$$f^*(x^*) = \begin{cases} -b & \text{if } x^* = a, \\ \infty & \text{otherwise.} \end{cases}$$

(b) $f(x) = \|x\|$, $f^*(x^*) = I_{B^*}(x^*)$.

(c) Let X be a Hilbert space and $f(x) = \frac{1}{2}\|x\|^2$, then $f^*(x^*) = \frac{1}{2}\|x^*\|_*^2$, as one can see by looking for the maximizing point in the definition of the conjugate.

(d) $f(x) = I_C(x)$, $f^*(x^*) = \sup_{x \in C} \langle x^*, x \rangle := \sigma_C(x^*)$; σ_C is a positively homogeneous function, called the *support* function of C. If C is the unit ball of the space X, then $f^*(x^*) = \|x^*\|_*$. If C is a cone, the support function of C is the indicator function of the cone C°, the polar cone of C, which is defined as $C^\circ = \{x^* \in X^* : \langle x^*, x \rangle \leq 0, \forall x \in C\}$. Observe that C° is a weak*-closed convex cone.

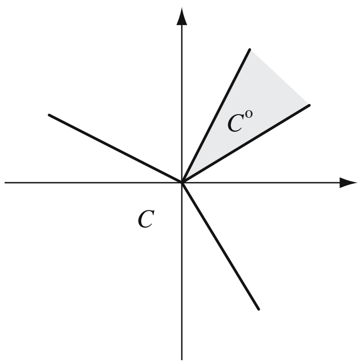


Figure 5.2. A cone C and its polar cone C°.

Exercise 5.1.3 Find f^*, for each f listed: (a) $f(x) = e^x$, (b) $f(x) = x^4$, (c) $f(x) = \sin x$, (d) $f(x) = \max\{0, x\}$, (e) $f(x) = -x^2$, (f) $f(x, y) = xy$,

(g) $f(x) = \begin{cases} e^x & \text{if } x \geq 0, \\ \infty & \text{otherwise,} \end{cases}$ (h) $f(x) = \begin{cases} x \ln x & \text{if } x \geq 0, \\ \infty & \text{otherwise,} \end{cases}$

(i) $f(x) = \begin{cases} 1 & \text{if } x \geq 0, \\ -1 & \text{otherwise;} \end{cases}$ (j) $f(x) = (x^2 - 1)^2$,

(k) $f(x) = \begin{cases} 0 & \text{if } |x| \leq 1, \\ (x^2 - 1)^2 & \text{otherwise.} \end{cases}$

The next proposition summarizes some elementary properties of f^*; we leave the easy proofs as an exercise.

Proposition 5.1.4 *We have:*

(i) $f^*(0) = -\inf f$;
(ii) $f \le g \Rightarrow f^* \ge g^*$;
(iii) $(\inf_{j\in J} f_j)^* = \sup_{j\in J} f_j^*$;
(iv) $(\sup_{j\in J} f_j)^* \le \inf_{j\in J} f_j^*$;
(v) $\forall r > 0, (rf)^*(x^*) = rf^*(\frac{x^*}{r})$;
(vi) $\forall r \in \mathbb{R}, (f+r)^*(x^*) = f^*(x^*) - r$;
(vii) $\forall \hat{x} \in X$, *if* $g(x) := f(x-\hat{x})$, *then* $g^*(x^*) = f^*(x^*) + \langle x^*, \hat{x}\rangle$.

Example 5.1.5 Let $f(x) = x$, $g(x) = -x$. Then $(\max\{f,g\})^*(x^*) = I_{[-1,1]}$, $\min\{f^*, g^*\}(x^*) = 0$ if $|x| = 1$, ∞ elsewhere. Thus the inequality in the fourth item above can be strict, which is almost obvious from the fact that in general $\inf_{j\in J} f_j^*$ need not be convex.

Example 5.1.6 Let $g\colon \mathbb{R} \to (-\infty, \infty]$ be an even function. Let $f\colon X \to \mathbb{R}$ be defined as $f(x) = g(\|x\|)$. Then

$$f^*(x^*) = g^*(\|x^*\|_*).$$

For,

$$\begin{aligned} f^*(x^*) &= \sup_{x\in X}\{\langle x^*, x\rangle - g(\|x\|)\} = \sup_{t\ge 0}\sup_{\|x\|=t}\{\langle x^*, x\rangle - g(\|x\|)\} \\ &= \sup_{t\ge 0}\{t\|x^*\|_* - g(t)\} = \sup_{t\in\mathbb{R}}\{t\|x^*\|_* - g(t)\} = g^*(\|x^*\|_*). \end{aligned}$$

Exercise 5.1.7 Let X be a Banach space, $f(x) = \frac{1}{p}\|x\|^p$, with $p > 1$. Then $f^*(x^*) = \frac{1}{q}\|x^*\|_*^q$ $(\frac{1}{p} + \frac{1}{q} = 1)$.

The case $p = 2$ generalizes Example 5.1.2 (c).

Exercise 5.1.8 Let X be a Banach space, let $A\colon X \to X$ be a linear, bounded and invertible operator. Finally, let $f \in \Gamma(X)$ and $g(x) = f(Ax)$. Evaluate g^*.

Hint. $g^*(x^*) = f^*((A^{-1})^*)(x^*)$.

Exercise 5.1.9 Evaluate f^* when f is

$$f(x) = \begin{cases} -\sqrt{x} & \text{if } x \ge 0, \\ \infty & \text{otherwise,} \end{cases} \qquad f(x,y) = \begin{cases} -2\sqrt{xy} & \text{if } x \ge 0, y \ge 0, \\ \infty & \text{otherwise.} \end{cases}$$

Exercise 5.1.10 Let X be a Banach space. Suppose $\lim_{\|x\|\to\infty} \frac{f(x)}{\|x\|} = \infty$. Prove that $\operatorname{dom} f^* = X^*$ and that the supremum in the definition of the conjugate of f is attained if X is reflexive.

Exercise 5.1.11 Let X be a Banach space and let $f \in \Gamma(X)$. Then the following are equivalent:

(i) $\lim_{\|x\|\to\infty} f(x) = \infty$;

(ii) there are $c_1 > 0, c_2$ such that $f(x) \geq c_1\|x\| - c_2$;
(iii) $0 \in \operatorname{int}\operatorname{dom} f^*$.

Find an analogous formulation for the function $f(x) - \langle x^*, x\rangle$, where $x^* \in X^*$.

Hint. Suppose $f(0) = 0$, and let r be such that $f(x) \geq 1$ if $\|x\| \geq r$. Then, for x such that $\|x\| > r$, we have that $f(x) \geq \frac{\|x\|}{r}$. Moreover, there exists $\hat{c} < 0$ such that $f(x) \geq \hat{c}$ if $\|x\| \leq r$. Then $f(x) \geq \frac{\|x\|}{r} + \hat{c} - 1$ for all x. This shows that (i) implies (ii).

Exercise 5.1.12 Let $f \in \Gamma(X)$. Then $\lim_{\|x^*\|_* \to \infty} \frac{f^*(x^*)}{\|x^*\|_*} = \infty$ if and only if f is upper bounded on all the balls. In particular this happens in finite dimensions, if and only if f is real valued. On the contrary, in infinite dimensions there are continuous real valued convex functions which are not bounded on the unit ball.

Hint. Observe that the condition $\lim_{\|x^*\|_* \to \infty} \frac{f^*(x^*)}{\|x^*\|_*} = \infty$ is equivalent to having that for each $k > 0$, there is c_k such that $f^*(x^*) \geq k\|x^*\|_* - c_k$. On the other hand, f is upper bounded on kB if and only if there exists c_k such that $f(x) \leq I_{kB}(x) + c_k$.

5.2 The bijection between $\Gamma(X)$ and $\Gamma^*(X^*)$

Starting from a given arbitrary function f, we have built its conjugate f^*. Of course, we can apply the same conjugate operation to f^*, too. In this way, we shall have a new function f^{**}, defined on X^{**}. But we are not interested in it. We shall instead focus our attention to its *restriction* to X, and we shall denote it by f^{**}. Thus

$$f^{**}\colon X \to [-\infty, \infty]; f^{**}(x) = \sup_{x^* \in X^*} \{\langle x^*, x\rangle - f^*(x^*)\}.$$

In this section, we study the connections between f and f^{**}.

Proposition 5.2.1 *We have $f^{**} \leq f$.*

Proof. $\forall x \in X$, $\forall x^* \in X^*$,

$$\langle x^*, x\rangle - f^*(x^*) \leq f(x).$$

Taking the supremum over $x^* \in X^*$ in both sides provides the result. □

Definition 5.2.2 We define the *convex, lower semicontinuous regularization* of $f\colon X \to (-\infty, \infty]$ to be the function $\hat{f}$ such that

$$\operatorname{epi} \hat{f} = \operatorname{cl}\operatorname{co}\operatorname{epi} f.$$

The definition is consistent because the convex hull of an epigraph is still an epigraph. Clearly, $\hat{f}$ is the largest convex (the closure of a convex set is convex) and lower semicontinuous function minorizing f: if $g \le f$ and g is convex and lower semicontinuous, then $g \le \hat{f}$. For, $\operatorname{epi} g$ is a closed convex set containing $\operatorname{epi} f$, hence it contains $\operatorname{cl}\operatorname{co}\operatorname{epi} f$.

Remark 5.2.3 If f is convex, then $\hat{f} = \bar{f}$. If $f \in \Gamma(X)$, then $f = \hat{f}$. This easily follows from

$$\operatorname{epi} f = \operatorname{cl}\operatorname{co}\operatorname{epi} f.$$

Observe that we always have $\hat{f} \ge f^{**}$, as $f^{**} \le f$ and f^{**} is convex and lower semicontinuous.

The next theorem provides a condition to ensure that $\hat{f}$ and f^{**} coincide. Exercise 5.2.5 shows that such a condition is not redundant.

Theorem 5.2.4 *Let $f\colon X \to (-\infty, \infty]$ be such that there are $x^* \in X^*, \alpha \in \mathbb{R}$ with $f(x) \ge \langle x^*, x\rangle + \alpha, \forall x \in X$. Then $\hat{f} = f^{**}$.*

Proof. The claim is obviously true if f is not proper, as in such a case, both f^{**} and $\hat{f}$ are constantly ∞. Then we have that $\forall x \in X$,

$$\hat{f}(x) \ge f^{**}(x) \ge \langle x^*, x\rangle + \alpha.$$

The last inequality follows from the fact that $f \ge g \Longrightarrow f^{**} \ge g^{**}$ and that the biconjugate of an affine function coincides with the affine function itself. Thus $f^{**}(x) > -\infty$ for all x. Let us suppose now, for the sake of contradiction, that there is $x_0 \in X$ such that $f^{**}(x_0) < \hat{f}(x_0)$. It is then possible to separate $(x_0, f^{**}(x_0))$ and $\operatorname{epi} \hat{f}$. If $\hat{f}(x_0) < \infty$, we then get the existence of $y^* \in X^*$ such that

$$\langle y^*, x_0\rangle + f^{**}(x_0) < \langle y^*, x\rangle + \hat{f}(x) \le \langle y^*, x\rangle + f(x), \forall x \in X.$$

(To be sure of this, take a look at the proof of Theorem 2.2.21). This implies

$$f^{**}(x_0) < \langle -y^*, x_0\rangle - \sup_{x\in X}\{\langle -y^*, x\rangle - f(x)\} = \langle -y^*, x_0\rangle - f^*(-y^*),$$

which is impossible. We then have to understand what is going on when $\hat{f}(x_0) = \infty$. In the case that the separating hyperplane is not vertical, one concludes as before. In the other case, we have the existence of $y^* \in X^*$, $c \in \mathbb{R}$ such that

(i) $\langle y^*, x\rangle - c < 0 \,\forall x \in \operatorname{dom} f$;
(ii) $\langle y^*, x_0\rangle - c > 0$.

Then

$$f(x) \ge \langle x^*, x\rangle + \alpha + t(\langle y^*, x\rangle - c),\ \forall x \in X, t > 0,$$

and this in turn implies, by conjugating twice, that

$$f^{**}(x) \ge \langle x^*, x\rangle + \alpha + t(\langle y^*, x\rangle - c),\ \forall x \in X, t > 0.$$

But then

$$f^{**}(x_0) \geq \langle x^*, x_0 \rangle + \alpha + t(\langle y^*, x_0 \rangle - c), \forall t > 0,$$

which implies $f^{**}(x_0) = \infty$. □

Exercise 5.2.5 Let

$$f(x) = \begin{cases} -x^2 & \text{if } x \leq 0, \\ \infty & \text{otherwise.} \end{cases}$$

Find f^{**} and $\hat{f}$.

Proposition 5.2.6 *Let $f\colon X \to [-\infty, \infty]$ be a convex function and suppose $f(x_0) \in \mathbb{R}$. Then f is lower semicontinuous at x_0 if and only if $f(x_0) = f^{**}(x_0)$.*

Proof. We always have $f^{**}(x_0) \leq f(x_0)$ (Proposition 5.2.1). Now, suppose f is lower semicontinuous at x_0. Let us see first that $\bar{f}$ cannot assume value $-\infty$ at any point. On the contrary, suppose there is z such that $\bar{f}(z) = -\infty$. Then $\bar{f}$ is never real valued, and so $\bar{f}(x_0) = -\infty$, against the fact that f is lower semicontinuous and real valued at x_0. It follows that $\bar{f}$ has an affine minorizing function; thus

$$\bar{f} = \hat{\bar{f}} = (\bar{f})^{**} \leq f^{**}.$$

As $\bar{f}(x_0) = f(x_0)$, we finally have $f(x_0) = f^{**}(x_0)$. Suppose now $f(x_0) = f^{**}(x_0)$. Then

$$\liminf f(x) \geq \liminf f^{**}(x) \geq f^{**}(x_0) = f(x_0),$$

and this shows that f is lower semicontinuous at x_0. □

The function

$$f(x) = \begin{cases} -\infty & \text{if } x = 0, \\ \infty & \text{otherwise,} \end{cases}$$

shows that the assumption $f(x_0) \in \mathbb{R}$ is *not* redundant in the above proposition. A more sophisticated example is the following one. Consider an infinite dimensional Banach space X, take $x^* \in X^*$ and a linear discontinuous functional l on X. Define

$$f(x) = \begin{cases} l(x) & \text{if } \langle x^*, x \rangle \geq 1, \\ \infty & \text{otherwise.} \end{cases}$$

Then f is continuous at zero, and it can be shown that $f^{**}(x) = -\infty$ for all x. Observe that f is lower semicontinuous at no point of its effective domain. This is the case because it can be shown that if there is at least a point of the effective domain of f where f is lower semicontinuous, then $f(x) = f^{**}(x)$

for all x such that f is lower semicontinuous (not necessarily real valued) at x ([Si2, Theorem 3.4]).

The next proposition shows that iterated application of the conjugation operation *does not* provide new functions.

Proposition 5.2.7 *Let* $f\colon X \to (-\infty, \infty]$*. Then* $f^* = f^{***}$*.*

Proof. As $f^{**} \le f$, one has $f^* \le f^{***}$. On the other hand, by definition of f^{***}, we have $f^{***}(x^*) = \sup_x \{\langle x^*, x\rangle - f^{**}(x)\}$, while, for all $x \in X$, $f^*(x^*) \ge \langle x^*, x\rangle - f^{**}(x)$, and this allows to conclude. □

Denote by $\Gamma^*(X^*)$ the functions of $\Gamma(X^*)$ which are conjugate of some function of $\Gamma(X)$. Then, from the previous results we get:

Theorem 5.2.8 *The operator* * *is a bijection between* $\Gamma(X)$ *and* $\Gamma^*(X^*)$*.*

Proof. If $f \in \Gamma(X)$, f^* cannot be $-\infty$ at any point. Moreover, f^* cannot be identically ∞ as there is an affine function $l(\cdot)$ of the form $l(x) = \langle x^*, x\rangle - r$ minorizing f (Corollary 2.2.17), whence $f^*(x^*) \le r$. These facts imply that * actually acts between $\Gamma(X)$ and $\Gamma^*(X^*)$. To conclude, it is enough to observe that if $f \in \Gamma(X)$, then $f = f^{**}$ (Proposition 5.2.4). □

Remark 5.2.9 If X is not reflexive, then $\Gamma^*(X^*)$ is a proper subset of $\Gamma(X^*)$. It is enough to consider a linear functional on X^* which is the image of no element of X via the canonical embedding of X into X^{**}; it belongs to $\Gamma(X^*)$, but it is not the conjugate of any function $f \in \Gamma(X)$.

5.3 The subdifferentials of f and f^*

Let us see, by a simple calculus in a special setting, how it is possible to evaluate the conjugate f^* of a function f, and the connection between the derivative of f and that of f^*. Let $f\colon \mathbb{R}^n \to (-\infty, \infty]$ be a convex function. Since $f^*(x^*) = \sup_{x\in X}\{\langle x^*, x\rangle - f(x)\}$, we start by supposing that f is superlinear ($\lim_{\|x\|\to\infty} \frac{f(x)}{\|x\|} = \infty$) and thus we have that the supremum in the definition of the conjugate is attained, for every x^*. To find a maximum point, like every student we assume that the derivative of f is zero at the maximum point, called $\bar{x}$. We get $x^* - \nabla f(\bar{x}) = 0$. We suppose also that ∇f has an inverse. Then $\bar{x} = (\nabla f)^{-1}(x^*)$. By substitution we get

$$f^*(x^*) = \langle x^*, (\nabla f)^{-1}(x^*)\rangle - f((\nabla f)^{-1}(x^*)).$$

We try now to determine $\nabla f^*(x^*)$. We get

$$\begin{aligned}\nabla f^*(x^*) &= (\nabla f)^{-1}(x^*) + \langle J_{(\nabla f)^{-1}}(x^*), x^*\rangle - \langle J_{(\nabla f)^{-1}}(x^*), \nabla f(\nabla f)^{-1})(x^*)\rangle \\ &= (\nabla f)^{-1}(x^*),\end{aligned}$$

where $J_{(\nabla f)^{-1}}$ denotes the jacobian matrix of the function $(\nabla f)^{-1}$. Then we have the interesting fact that the derivative of f is the inverse of the derivative of f^*. This fact can be fully generalized to subdifferentials, as we shall see in a moment.

Proposition 5.3.1 *Let $f\colon X \to (-\infty, \infty]$. Then $x^* \in \partial f(x)$ if and only if $f(x) + f^*(x^*) = \langle x^*, x\rangle$.*

Proof. We already know that

$$f(x) + f^*(x^*) \geq \langle x^*, x\rangle,\ \forall x \in X, x^* \in X^*.$$

If $x^* \in \partial f(x)$, then

$$f(y) - \langle x^*, y\rangle \geq f(x) - \langle x^*, x\rangle,\ \forall y \in X,$$

whence, $\forall y \in X$,

$$\langle x^*, y\rangle - f(y) + f(x) \leq \langle x^*, x\rangle.$$

Taking the supremum over all y in the left side provides one implication. As to the other one, if $f(x) + f^*(x^*) = \langle x^*, x\rangle$, then from the definition of f^*, we have that

$$f(x) + \langle x^*, y\rangle - f(y) \leq \langle x^*, x\rangle,\ \forall y \in X,$$

which shows that $x^* \in \partial f(x)$. □

Proposition 5.3.1 has some interesting consequences. At first,

Proposition 5.3.2 *Let $f\colon X \to (-\infty, \infty]$. If $\partial f(x) \neq \emptyset$, then $f(x) = f^{**}(x)$. If $f(x) = f^{**}(x)$, then $\partial f(x) = \partial f^{**}(x)$.*

Proof. $\forall x \in X, \forall x^* \in X^*$, we have

$$f^*(x^*) + f^{**}(x) \geq \langle x, x^*\rangle.$$

If $x^* \in \partial f(x)$, by Proposition 5.3.1 we get

$$f^*(x^*) + f(x) = \langle x^*, x\rangle.$$

It follows that $f^{**}(x) \geq f(x)$, and this shows the first part of the claim. Suppose now $f(x) = f^{**}(x)$. Then, using the equality $f^* = (f^{**})^*$,

$$\begin{aligned} x^* \in \partial f(x) &\Longleftrightarrow \langle x^*, x\rangle = f(x) + f^*(x^*) = f^{**}(x) + f^{***}(x^*) \\ &\Longleftrightarrow x^* \in \partial f^{**}(x). \end{aligned}$$

□

Another interesting consequence is the announced connection between the subdifferentials of f and f^*.

Corollary 5.3.3 *Let $f\colon X \to (-\infty, \infty]$. Then*

$$x^* \in \partial f(x) \Longrightarrow x \in \partial f^*(x^*).$$

If $f(x) = f^{**}(x)$, then

$$x^* \in \partial f(x) \text{ if and only if } x \in \partial f^*(x^*).$$

Proof. $x^* \in \partial f(x) \Longleftrightarrow \langle x^*, x\rangle = f(x) + f^*(x^*)$. Thus $x^* \in \partial f(x)$ implies $f^{**}(x) + f^*(x^*) \le \langle x^*, x\rangle$, and this is equivalent to saying that $x \in \partial f^*(x^*)$. If $f(x) = f^{**}(x)$,

$$\begin{aligned} x^* \in \partial f(x) &\Longleftrightarrow \langle x^*, x\rangle = f(x) + f^*(x^*) = f^{**}(x) + f^*(x^*) \\ &\Longleftrightarrow x \in \partial f^*(x^*). \end{aligned}$$

□

Thus, for a function $f \in \Gamma(X)$, it holds that $x^* \in \partial f(x)$ if and only if $x \in \partial f^*(x^*)$.

The above conclusion suggests how to draw the graph of the conjugate of a given function $f\colon \mathbb{R} \to \mathbb{R}$. We can construct the graph of its subdifferential, we "invert" it and we "integrate", remembering that, for instance, $f^*(0) = -\inf f$. See Figures 5.3–5.5 below.

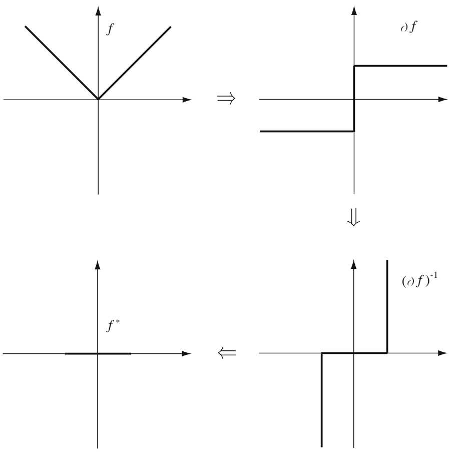


Figure 5.3. From the function to its conjugate through the subdifferentials.

A similar relation holds for approximate subdifferentials. For the following generalization of Proposition 5.3.1 holds:

Proposition 5.3.4 *Let $f \in \Gamma(X)$. Then $x^* \in \partial_\varepsilon f(x)$ if and only if $f^*(x^*) + f(x) \le \langle x^*, x\rangle + \varepsilon$. Hence, $x^* \in \partial_\varepsilon f(x)$ if and only if $x \in \partial_\varepsilon f^*(x^*)$.*

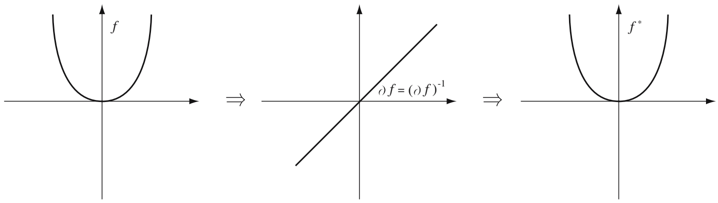


Figure 5.4. Another example.

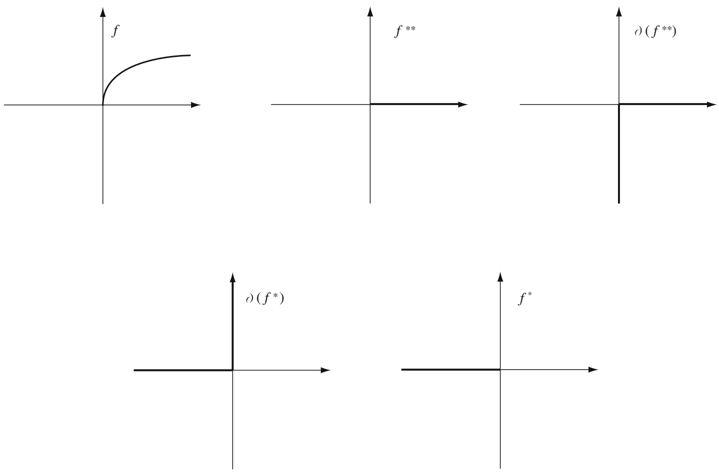


Figure 5.5. . . . and yet another one.

Proof. $x^* \in \partial_\varepsilon f(x)$ if and only if

$$f(x) + \langle x^*, y\rangle - f(y) \leq \langle x^*, x\rangle + \varepsilon, \; \forall y \in X,$$

if and only if $f(x) + f^*(x^*) \leq \langle x^*, x\rangle + \varepsilon$. The second claim follows from $f = f^{**}$. □

The previous proposition allows us to show that only in exceptional cases can the approximate subdifferential be a singleton (a nonempty, small set indeed).

Proposition 5.3.5 *Let $f \in \Gamma(X)$ and suppose there are $x \in \operatorname{dom} f$, $x^* \in X^*$ and $\bar{\varepsilon} > 0$ such that $\partial_{\bar{\varepsilon}} f(x) = \{x^*\}$. Then f is an affine function.*

Proof. As a first step one verifies that $\partial_\varepsilon f(x) = \{x^*\}$ for all $\varepsilon > 0$. This is obvious if $\varepsilon < \bar{\varepsilon}$, because $\partial_\varepsilon f(x) \neq \emptyset$, and due to monotonicity. Furthermore, the convexity property described in Theorem 3.7.2 implies that $\partial_\varepsilon f(x)$ is a singleton also for $\varepsilon > \bar{\varepsilon}$. For, take $\sigma < \bar{\varepsilon}$ and suppose $\partial_\varepsilon f(x) \ni y^* \neq x^*$, for some $\varepsilon > \bar{\varepsilon}$. An easy but tedious calculation shows that being $\partial_\sigma f(x) \ni x^*$,

$\partial_{\bar{\varepsilon}} f(x) \ni \frac{\varepsilon-\bar{\varepsilon}}{\varepsilon-\sigma}x^* + \frac{\bar{\varepsilon}-\sigma}{\varepsilon-\sigma}y^* \neq x^*$, a contradiction. It follows, by Proposition 5.3.4, that if $y^* \neq x^*$,

$$f^*(y^*) > \langle y^*, x\rangle - f(x) + \varepsilon,\ \forall \varepsilon > 0,$$

and this implies $\operatorname{dom} f^* = \{x^*\}$. We conclude that f must be an affine function. □

5.4 The conjugate of the sum

Proposition 5.4.1 *Let $f, g \in \Gamma(X)$. Then*

$$(f\nabla g)^* = f^* + g^*.$$

Proof.

$$\begin{aligned}(f\nabla g)^*(x^*) &= \sup_{x\in X}\big\{\langle x^*, x\rangle - \inf_{x_1+x_2=x}\{f(x_1)+g(x_2)\}\big\}\\ &= \sup_{\substack{x_1\in X\\ x_2\in X}}\{\langle x^*, x_1\rangle + \langle x^*, x_2\rangle - f(x_1) - g(x_2)\} = f^*(x^*) + g^*(x^*).\end{aligned}$$

□

Proposition 5.4.1 offers a good idea for evaluating $(f+g)^*$. By applying the above formula to f^*, g^* and conjugating, we get that

$$(f^*\nabla g^*)^{**} = (f^{**} + g^{**})^* = (f+g)^*.$$

So that if $f^*\nabla g^* \in \Gamma(X^*)$, then

$$(f+g)^* = f^*\nabla g^*.$$

Unfortunately we know that the inf-convolution operation between functions in $\Gamma(X)$ does not always produce a function belonging to $\Gamma(X)$; besides the case when at some point it is valued $-\infty$, it is not always lower semicontinuous. The next important theorem, due to Attouch–Brézis (see [AB]), provides a sufficient condition to get the result.

Theorem 5.4.2 *Let X be a Banach space and X^* its dual space. Let $f, g \in \Gamma(X)$. Moreover, let*

$$F := \mathbb{R}^+(\operatorname{dom} f - \operatorname{dom} g)$$

be a closed vector subspace of X. Then

$$(f+g)^* = f^*\nabla g^*,$$

and the inf*-convolution is exact.*

Proof. From the previous remark, it is enough to show that the inf-convolution is lower semicontinuous; in proving this we shall also see that it is exact (whence, in particular, it never assumes the value $-\infty$). We start by proving the claim in the particular case when $F = X$. From Exercise 2.2.4 it is enough to show that the level sets $(f^*\nabla g^*)^a$ are weak* closed for all $a \in \mathbb{R}$. On the other hand,

$$(f^*\nabla g^*)^a = \bigcap_{\varepsilon>0} C_\varepsilon := \{y^* + z^* : f^*(y^*) + g^*(z^*) \le a + \varepsilon\}.$$

It is then enough to show that the sets C_ε are weak* closed. Fixing $r > 0$, let us consider

$$K_{\varepsilon r} := \{(y^*, z^*) : f^*(y^*) + g^*(z^*) \le a + \varepsilon \text{ and } \|y^* + z^*\|_* \le r\}.$$

Then $K_{\varepsilon r}$ is a closed set in the weak* topology. Setting $T(y^*, z^*) = y^* + z^*$, we have that

$$C_\varepsilon \cap rB_{X^*} = T(K_{\varepsilon r}).$$

Since T is continuous from $X^* \times X^*$ to X^* (with the weak* topologies), if we show that $K_{\varepsilon r}$ is bounded (hence weak* compact), then $C_\varepsilon \cap rB^*$ is a weak* compact set, for all $r > 0$. The Banach–Dieudonné–Krein–Smulian theorem then guarantees that C_ε is weak* closed (See Theorem A.2.1 in Appendix B). Let us then show that $K_{\varepsilon r}$ is bounded. To do this, we use the uniform boundedness theorem. Thus, it is enough to show that $\forall y, z \in X$, there is a constant $C = C(y, z)$ such that

$$|\langle (y^*, z^*), (y, z)\rangle| = |\langle y^*, y\rangle + \langle z^*, z\rangle| \le C, \forall (y^*, z^*) \in K_{\varepsilon r}.$$

By assumption there is $t \ge 0$ such that $y - z = t(u - v)$ with $u \in \operatorname{dom} f$ and $v \in \operatorname{dom} g$. Then

$$\begin{aligned} |\langle y^*, y\rangle + \langle z^*, z\rangle| &= |t\langle y^*, u\rangle + t\langle z^*, v\rangle + \langle y^* + z^*, z - tv\rangle| \\ &\le |t(f(u) + f^*(y^*) + g(v) + g^*(z^*))| + r\|z - tv\| \\ &\le |t(a + \varepsilon + f(u) + g(v))| + r\|z - tv\| = C(y, z). \end{aligned}$$

The claim is proved in the case when $F = X$. Let us now turn to the general case. Suppose $u \in \operatorname{dom} f - \operatorname{dom} g$. Then $-u \in F$ and so there are $t \ge 0$ and $v \in \operatorname{dom} f - \operatorname{dom} g$ such that $-u = tv$. It follows that

$$0 = \frac{1}{1+t}u + \frac{t}{1+t}v \in \operatorname{dom} f - \operatorname{dom} g.$$

Hence $\operatorname{dom} f \cap \operatorname{dom} g \ne \emptyset$ and after a suitable translation, we can suppose that $0 \in \operatorname{dom} f \cap \operatorname{dom} g$, whence $\operatorname{dom} f \subset F$, $\operatorname{dom} g \subset F$. Let $i\colon F \to X$ be the canonical injection of F in X and let $i^*\colon X^* \to F^*$ be its adjoint operator: $\langle i^*(x^*), d\rangle = \langle x^*, i(d)\rangle$. Let us consider the functions

$$\tilde{f}\colon F \to (-\infty, \infty],\ \tilde{f} := f \circ i, \quad \tilde{g}\colon F \to (-\infty, \infty],\ \tilde{g} := g \circ i.$$

We can apply the first step of the proof to them. We have

$$(\tilde{f} + \tilde{g})^*(z^*) = (\tilde{f}^* \nabla \tilde{g}^*)(z^*),$$

for all $z^* \in F^*$. It is now easy to verify that if $x^* \in X^*$,

$$f^*(x^*) = \tilde{f}^*(i^*(x^*)), \quad g^*(x^*) = \tilde{g}^*(i^*(x^*)),$$
$$(f+g)^*(x^*) = (\tilde{f} + \tilde{g})^*(i^*(x^*)), \quad (f^* \nabla g^*)(x^*) = (\tilde{f}^* \nabla \tilde{g}^*)(i^*(x^*)),$$

(in the last one we use that i^* is onto).

For instance, we have

$$\begin{aligned}\tilde{f}^*(i^*(x^*)) &= \sup_{z \in F}\{\langle i^*(x^*), z\rangle - \tilde{f}(z)\} = \sup_{z \in F}\{\langle x^*, i(z)\rangle - f(i(z))\}\\ &= \sup_{x \in X}\{\langle x^*, x\rangle - f(x)\},\end{aligned}$$

where the last inequality holds as $\operatorname{dom} f \subset F$. The others follow in the same way. Finally, the exactness at a point $x^* \in \operatorname{dom} f^* \nabla g^*$ follows from the compactness, previously shown, of $K_{\varepsilon, \|x^*\|_*}$, with $a = (f^* \nabla g^*)(x^*)$ and $\varepsilon > 0$ arbitrary. This allows us to conclude. □

Besides its intrinsic interest, the previous theorem yields the following sum rule for the subdifferentials which generalizes Theorem 3.4.2.

Theorem 5.4.3 *Let $f, g \in \Gamma(X)$. Moreover, let*

$$F := \mathbb{R}_+(\operatorname{dom} f - \operatorname{dom} g)$$

be a closed vector space. Then

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

Proof. Let $x^* \in \partial(f+g)(x)$. We must find $y^* \in \partial f(x)$ and $z^* \in \partial g(x)$ such that $y^* + z^* = x^*$. By the previous result there are y^*, z^* such that $y^* + z^* = x^*$ and fulfilling $f^*(y^*) + g^*(z^*) = (f+g)^*(x^*)$. As $x^* \in \partial(f+g)(x)$ we have (Proposition 5.3.1)

$$\begin{aligned}\langle y^*, x\rangle + \langle z^*, x\rangle = \langle x^*, x\rangle &= (f+g)(x) + (f+g)^*(x^*)\\ &= f(x) + f^*(y^*) + g(x) + g^*(z^*).\end{aligned}$$

This implies (why?)

$$\langle y^*, x\rangle = f(x) + f^*(y^*) \text{ and } \langle z^*, x\rangle = g(x) + g^*(z^*),$$

and we conclude. □

The previous generalization is useful, for instance, in the following situation: suppose we have a Banach space Y, a (proper) closed subspace X and two continuous functions $f, g \in \Gamma(X)$ fulfilling the condition $\operatorname{int}\operatorname{dom} f \cap \operatorname{dom} g \neq \emptyset$. It can be useful sometimes to consider the natural extensions $\tilde{f}, \tilde{g} \in \Gamma(Y)$ of f and g (by defining them ∞ outside X). In such a case the previous theorem can be applied, while Theorem 3.4.2 obviously cannot.

Exercise 5.4.4 Let

$$f(x,y) = \begin{cases} -\sqrt{xy} & \text{if } x \leq 0, y \leq 0, \\ \infty & \text{otherwise,} \end{cases}$$

$$g(x,y) = \begin{cases} -\sqrt{-xy} & \text{if } x \geq 0, y \leq 0, \\ \infty & \text{otherwise.} \end{cases}$$

Find $(f+g)^*$ and $f^* \nabla g^*$.

Exercise 5.4.5 Given a nonempty closed convex set K,

$$d^*(\,\cdot\,, K) = \sigma_K + I_{B^*}.$$

Hint. Remember that $d(\,\cdot\,, K) = (\|\ \| \nabla I_K)(\cdot)$ and apply Proposition 5.4.1.

Exercise 5.4.6 Let X be a reflexive Banach space. Let $f, g \in \Gamma(X)$. Let

$$\lim_{\|x\| \to \infty} \frac{f(x)}{\|x\|} = \infty.$$

Then $(f \nabla g) \in \Gamma(X)$.

Hint. Try to apply the Attouch–Brézis theorem to f^*, g^*.

5.5 Sandwiching an affine function between a convex and a concave function

In this section we deal with the following problem: suppose we are given a Banach space X and two convex, lower semicontinuous extended real valued functions f and g such that $f(x) \geq -g(x)\ \forall x \in X$. The question is: when is it possible to find an affine function m with the property that

$$f(x) \geq m(x) \geq -g(x),$$

for all $x \in X$? It is clear that the problem can be restated in an equivalent, more geometric, way: suppose we can separate the sets $\operatorname{epi} f$ and $\operatorname{hyp}(-g)$ with a nonvertical hyperplane. With a standard argument this provides the affine function we are looking for. And, clearly, the condition $f \geq -g$ gives some hope to be able to make such a separation.

In order to study the problem, let us first observe the following simple fact.

Proposition 5.5.1 *Let $y^* \in X^*$. Then $y^* \in \{p : f^*(p) + g^*(-p) \leq 0\}$ if and only if there exists $a \in \mathbb{R}$ such that*

$$f(x) \geq \langle y^*, x \rangle + a \geq -g(x),$$

for all $x \in X$.

Proof. Suppose $f^*(y^*) + g^*(-y^*) \leq 0$. Then, for all $x \in X$,

$$\langle y^*, x \rangle - f(x) + g^*(-y^*) \leq 0,$$

i.e.,

$$f(x) \geq \langle y^*, x \rangle + a,$$

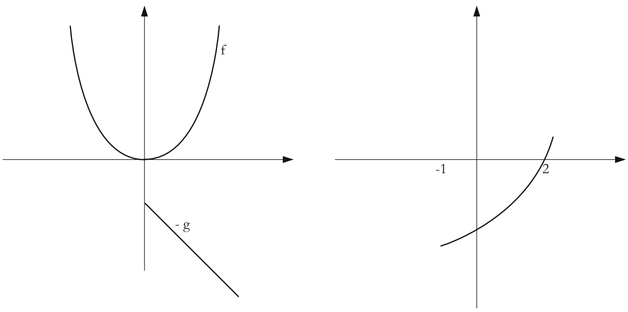
with $a = g^*(-y^*)$. Moreover

$$a = g^*(-y^*) \geq \langle -y^*, x \rangle - g(x),$$

for all $x \in X$, implying $\langle y^*, x \rangle + a \geq -g(x)$, for all $x \in X$. Conversely, if $f(x) \geq \langle y^*, x \rangle + a$ and $\langle y^*, x \rangle + a \geq -g(x)$ for all x, then

$$-a \geq f^*(y^*), \quad a \geq \langle -y^*, x \rangle - g(x),$$

for all x, implying $f^*(y^*) + g^*(-y^*) \leq 0$. □



$f(x) = \frac{1}{2}x^2, g(x) = \begin{cases} x+3 & \text{if } x \geq 0, \\ \infty & \text{otherwise.} \end{cases}$ $\qquad f^*(\cdot) + g^*(-\,\cdot).$

Figure 5.6.

It follows in particular that the set of the "slopes" of the affine functions sandwiched between f and $-g$ is a weak* closed and convex set, as it is the zero level set of the function $h(\cdot) = f^*(\cdot) + g^*(-\,\cdot)$. Now, observe that $\inf_x (f+g)(x) \geq 0$ if and only if $(f+g)^*(0^*) \leq 0$. Thus, *if*

$$(f+g)^*(0^*) = (f^* \nabla g^*)(0^*)$$

and the epi-sum is exact, then $\inf_x (f+g)(x) \geq 0$ is equivalent to saying that there exists $y^* \in X^*$ such that

$$(f^* \nabla g^*)(0^*) = f^*(y^*) + g^*(-y^*) \leq 0.$$

Thus a sufficient condition to have an affine function sandwiched between f and $-g$ is that the assumption of the Attouch–Brezis theorem be satisfied.

Now we specialize to the case when X is a Euclidean space. In this case the condition $f \geq -g$ implies that

$$\operatorname{ri}\operatorname{epi} f \cap \operatorname{ri}\operatorname{hyp}(-g) = \emptyset.$$

Then we can apply Theorem A.1.13 to separate the sets $\operatorname{epi} f$ and $\operatorname{hyp}(-g)$. However, this does not solve the problem, as it can happen that the separating hyperplane is vertical. So, let us now see a sufficient condition in order to assure that the separating hyperplane is not vertical, which amounts to saying that the affine function we are looking for is finally singled out.

Proposition 5.5.2 *Suppose*

$$\operatorname{ri}\operatorname{dom} f \cap \operatorname{ri}\operatorname{dom}(-g) \neq \emptyset.$$

Then there exists y^* such that $f^*(y^*) + g^*(-y^*) \leq 0$.

Proof. Let us use the Attouch–Brezis theorem, as suggested at the beginning of the section. Thus, we must show that

$$F := \mathbb{R}_+(\operatorname{dom} f - \operatorname{dom} g)$$

is a subspace. As is suggested in the next exercise, it is enough to show that if $x \in F$, then $-x \in F$. We can suppose, without loss of generality, that $0 \in \operatorname{ri}\operatorname{dom} f \cap \operatorname{ri}\operatorname{dom} g$. As $x \in F$, there are $l > 0$, $u \in \operatorname{dom} f$ and $v \in \operatorname{dom} g$ such that $x = l(u-v)$. As $0 \in \operatorname{ri}\operatorname{dom} f \cap \operatorname{ri}\operatorname{dom} g$, there is $c > 0$ small enough such that $-cu \in \operatorname{dom} f$, $-cv \in \operatorname{dom} g$. Thus $-cu - (-cv) \in \operatorname{dom} f - \operatorname{dom} g$. Then

$$\frac{l}{c}(-cu - (-cv)) = -x \in F.$$

□

Exercise 5.5.3 Let A be a convex set containing zero. Then $\bigcup_{\lambda>0} \lambda A$ is a convex cone. Moreover, if $x \in \bigcup_{\lambda>0} \lambda A$ implies $-x \in \bigcup_{\lambda>0} \lambda A$, then $\bigcup_{\lambda>0} \lambda A$ is a subspace.

Hint. Call $F = \bigcup_{\lambda>0} \lambda A$. It has to be shown that $x, y \in F$ implies $x + y \in F$. There are positive l_1, l_2 and $u, v \in A$ such that $x = l_1 u$, $y = l_2 v$. Then $x/l_1 \in A$, $y/l_2 \in A$ and $\frac{1}{l_1+l_2}(x+y)$ is a convex combination of x/l_1 and y/l_2.

We now give some pretty examples showing that the affine function separating $\operatorname{epi} f$ and $\operatorname{hyp}(-g)$ need not exist, unless some extra condition is imposed.

Example 5.5.4

$$f(x) = \begin{cases} -\sqrt{x} & \text{if } x \geq 0, \\ \infty & \text{otherwise,} \end{cases}$$

$$g(x) = \begin{cases} 0 & \text{if } x = 0, \\ \infty & \text{otherwise.} \end{cases}$$

Here $\inf(f+g) = 0$, and $\operatorname{ri}(\operatorname{dom} f) \cap \operatorname{ri}(\operatorname{dom} g) = \emptyset$.

Example 5.5.5

$$f(u,v) = \begin{cases} -1 & \text{if } uv \geq 1, u \geq 0, \\ \infty & \text{otherwise,} \end{cases}$$

$$g(u,v) = \begin{cases} 0 & \text{if } u \geq 0, v = 0, \\ \infty & \text{otherwise.} \end{cases}$$

Here we have $\operatorname{dom} f \cap \operatorname{dom} g = \emptyset$.

Example 5.5.6

$$f(u,v) = \begin{cases} u & \text{if } v = -1, \\ \infty & \text{otherwise,} \end{cases}$$

$$g(u,v) = \begin{cases} 0 & \text{if } v = 0, \\ \infty & \text{otherwise.} \end{cases}$$

The Example 5.5.4 can induce the idea that the separator must be vertical as the two effective domains do intersect at a point. So, it could be argued that, if the two domain are far apart, the property could hold. But in Example 5.5.6 the distance between $\operatorname{dom} f$ and $\operatorname{dom} g$ is 1.

In the last two examples the domains of f and g do not intersect, while in the first example a crucial role is played by the fact that $\inf(f+g) = 0$. In the following example $\inf(f+g) > 0$, and yet there is no affine separator. Observe that such example could not be provided in one dimension (see Remark 2.2.15).

Example 5.5.7

$$f(u,v) = \begin{cases} 1 - 2\sqrt{uv} & \text{if } u, v \geq 0, \\ \infty & \text{otherwise,} \end{cases}$$

$$g(u,v) = \begin{cases} 1 - 2\sqrt{-uv} & \text{if } u \leq 0, v \geq 0, \\ \infty & \text{otherwise.} \end{cases}$$

A straightforward calculation shows

$$f^*(u^*, v^*) = \begin{cases} -1 & \text{if } u^* \le 0, u^* v^* \ge 1, \\ \infty & \text{otherwise,} \end{cases}$$

$$g^*(u^*, v^*) = \begin{cases} -1 & \text{if } u^* \ge 0, u^* v^* \le -1, \\ \infty & \text{otherwise.} \end{cases}$$

Our finite dimensional argument actually holds, without any changes in the proof, provided we assume that at least one of the sets $\operatorname{epi} f$, $\operatorname{hyp}(-g)$ has an interior point. In particular, the assumption in Proposition 5.5.2 becomes, in infinite dimensions, $\operatorname{int}\operatorname{dom} f \cap \operatorname{dom} g \neq \emptyset$. To conclude, let me mention that this section is inspired by my work with Lewis [LeL], where we studied the more general problem of giving sufficient conditions under which the slope of the affine function between f and $-g$ is in the range (or in the closure of the range) of the Clarke subdifferential of a locally Lipschitz function h such that $f \ge h \ge -g$.