11. Dualization

In the realm of convexity, almost every mathematical object can be paired with another, said to be *dual* to it. The pairing between convex cones and their polars has already been fundamental in the variational geometry of Chapter 6 in relating tangent vectors to normal vectors. The pairing between convex sets and sublinear functions in Chapter 8 has served as the vehicle for expressing connections between subgradients and subderivatives. Both correspondences are rooted in a deeper principle of duality for 'conjugate' pairs of convex functions, which will emerge fully here.

On the basis of this duality, close connections between otherwise disparate properties are revealed. It will be seen for instance that the level boundedness of one function in a conjugate pair corresponds to the finiteness of the other function around the origin. A catalog of such surprising linkages can be put together, and lists of dual operations and constructions to go with them.

In this way the analysis of a given situation can often be translated into an equivalent yet very different context. This can be a major source of insights as well as a means of unifying seemingly divergent parts of theory. The consequences go far beyond situations ruled by pure convexity, because many problems, although nonconvex, have crucial aspects of convexity in their structure, and the dualization of these can already be very fruitful. Among other things, we'll be able to apply such ideas to the general expression of optimality conditions in terms of a Lagrangian function, and even to the dualization of optimization problems themselves.

A. Legendre-Fenchel Transform

The general framework for duality is built around a 'transform' that gives an operational form to the envelope representations of convex functions. For any function $f : I\!R^n \to \overline{I\!R}$, the function $f^* : I\!R^n \to \overline{I\!R}$ defined by

$$f^*(v) := \sup_x \big\{ \langle v, x\rangle - f(x) \big\} \qquad 11(1)$$

is *conjugate* to f, while the function $f^{**} = (f^*)^*$ defined by

$$f^{**}(x) := \sup_v \big\{ \langle v, x\rangle - f^*(v) \big\} \qquad 11(2)$$

is *biconjugate* to f. The mapping $f \mapsto f^*$ from $\operatorname{fcns}(\mathbb{R}^n)$ into $\operatorname{fcns}(\mathbb{R}^n)$ is the *Legendre-Fenchel transform*.

The significance of the conjugate function f^* can easily be understood in terms of epigraph relationships. Formula 11(1) says that

$$(v,\beta) \in \operatorname{epi} f^* \iff \beta \geq \langle v, x\rangle - \alpha \text{ for all } (x,\alpha) \in \operatorname{epi} f.$$

If we write the inequality as $\alpha \geq \langle v, x\rangle - \beta$ and think of the affine functions on $\mathbb{R}^n$ as parameterized by pairs $(v,\beta) \in \mathbb{R}^n \times \mathbb{R}$, we can express this as

$$(v,\beta) \in \operatorname{epi} f^* \iff l_{v,\beta} \leq f, \text{ where } l_{v,\beta}(x) := \langle v, x\rangle - \beta.$$

Since the specification of a function on $\mathbb{R}^n$ is tantamount to the specification of its epigraph, this means that f^* *describes the family of all affine functions majorized by* f. Simultaneously, though, our calculation reveals that

$$\beta \geq f^*(v) \iff \beta \geq l_{x,\alpha}(v) \text{ for all } (x,\alpha) \in \operatorname{epi} f,$$

In other words, f^* *is the pointwise supremum of the family of all affine functions* $l_{x,\alpha}$ *for* $(x,\alpha) \in \operatorname{epi} f$. By the same token then, formula 11(2) means that f^{**} *is the pointwise supremum of all the affine functions majorized by* f.

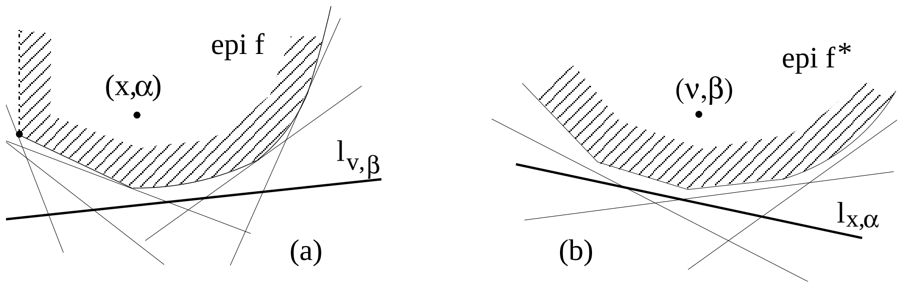

Fig. 11–1. (a) Affine functions majorized by f. (b) Affine functions majorized by f^*.

Recalling the facts about envelope representations in Theorem 8.13 and making use of the notion of the *convex hull* $\operatorname{con} f$ of an arbitrary function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ (see 2.31), we can summarize these relationships as follows.

11.1 Theorem (Legendre-Fenchel transform)**.** *For any function* $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ *with* $\operatorname{con} f$ *proper, both* f^* *and* f^{**} *are proper, lsc and convex, and*

$$f^{**} = \operatorname{cl} \operatorname{con} f.$$

Thus $f^{**} \leq f$, *and when* f *is itself proper, lsc and convex, one has* $f^{**} = f$. *Anyway, regardless of such assumptions, one always has*

$$f^* = (\operatorname{con} f)^* = (\operatorname{cl} f)^* = (\operatorname{cl} \operatorname{con} f)^*.$$

Proof. In the light of the preceding explanation of the meaning of the Legendre-Fenchel transform, this is immediate from Theorem 8.13; see 2.32 for the properness of $\operatorname{cl} f$ when f is convex and proper. $\square$

11.2 Exercise (improper conjugates)**.** *For a function $f : I\!R^n \to \overline{I\!R}$ with* $\operatorname{con} f$ *improper in the sense of taking on $-\infty$, one has $f^* \equiv \infty$ and $f^{**} \equiv -\infty$, while* $\operatorname{cl}\operatorname{con} f$ *has the value $-\infty$ on the set* $\operatorname{cl}\operatorname{dom}(\operatorname{con} f)$ *but the value ∞ outside this set. For the improper function $f \equiv \infty$, one has $f^* \equiv -\infty$ and $f^{**} \equiv \infty$.*

Guide. Make use of 2.5. □

The Legendre-Fenchel transform obviously reverses ordering among the functions to which it is applied:

$$f_1 \le f_2 \implies f_1^* \ge f_2^*.$$

The fact that $f = f^{**}$ when f is proper, lsc and convex means that the Legendre-Fenchel transform sets up a one-to-one correspondence in the class of all such functions: if g is conjugate to f, then f is conjugate to g:

$$f \underset{*}{\longleftrightarrow} g \quad \text{when} \quad \begin{cases} g(v) = \sup_x\big\{\langle v, x\rangle - f(x)\big\}, \\ f(x) = \sup_v\big\{\langle v, x\rangle - g(v)\big\}. \end{cases}$$

This is called the *conjugacy* correspondence. Every property of one function in a conjugate pair must mirror some property of the other function. Every construction or operation must have its conjugate counterpart. This far-reaching principle of duality allows everything to be viewed from two different angles, often with remarkable consequences.

An initial illustration of the duality of operations is seen in the following relations, which immediately fall out of the definition of conjugacy. In each case the expression on the left gives a function of x while the one on the right gives the corresponding function of v under the assumption that $f \underset{*}{\longleftrightarrow} g$:

$$\begin{aligned} f(x) - \langle a, x\rangle \quad &\underset{*}{\longleftrightarrow} \quad g(v + a), \\ f(x + b) \quad &\underset{*}{\longleftrightarrow} \quad g(v) - \langle v, b\rangle, \\ f(x) + c \quad &\underset{*}{\longleftrightarrow} \quad g(v) - c, \\ \lambda f(x) \quad &\underset{*}{\longleftrightarrow} \quad \lambda g(\lambda^{-1}v) \quad (\text{for } \lambda > 0), \\ \lambda f(\lambda^{-1}x) \quad &\underset{*}{\longleftrightarrow} \quad \lambda g(v) \quad (\text{for } \lambda > 0). \end{aligned} \tag{11(3)}$$

Interestingly, the last two relations pair *multiplication* with *epi-multiplication*:

$$(\lambda f)^* = \lambda \star f^*, \qquad (\lambda \star f)^* = \lambda f^*,$$

for positive scalars λ. (An extension to $\lambda = 0$ will come out in Theorem 11.5.) Later we'll see a similar duality between *addition* and *epi-addition* of functions (in Theorem 11.23(a)).

One of the most important dualization rules operates on subgradients. It stems from the fact that the subgradients of a convex function correspond to its affine supports (as described after 8.12). To say that the affine function $l_{\bar v,\bar\beta}$ supports f at $\bar x$, with $\bar\alpha = f(\bar x)$, is to say that the affine function $l_{\bar x,\bar\alpha}$ supports

f^* at $\bar{v}$, with $\bar{\beta} = f^*(\bar{v})$; cf. Figure 11–1 again. This gives us a relationship between subgradients of f and those of f^*.

11.3 Proposition (inversion rule for subgradient relations)**.** *For any proper, lsc, convex function f, one has $\partial f^* = (\partial f)^{-1}$ and $\partial f = (\partial f^*)^{-1}$. Indeed,*

$$\bar{v} \in \partial f(\bar{x}) \iff \bar{x} \in \partial f^*(\bar{v}) \iff f(\bar{x}) + f^*(\bar{v}) = \langle \bar{v}, \bar{x} \rangle,$$

whereas $f(x) + f^*(v) \geq \langle v, x \rangle$ for all x, v. Hence $\operatorname{gph} \partial f$ is closed and

$$\partial f(\bar{x}) = \operatorname{argmax}_v \big\{ \langle v, \bar{x} \rangle - f^*(v) \big\}, \qquad \partial f^*(\bar{v}) = \operatorname{argmax}_x \big\{ \langle \bar{v}, x \rangle - f(x) \big\}.$$

Proof. The argument just given would suffice, but here's another view of why the relations hold. From the first formula in 11(3) we know that for any $\bar{v}$ the points $\bar{x}$ furnishing the minimum of the convex function $f_{\bar{v}}(x) := f(x) - \langle \bar{v}, x \rangle$, if any, are the ones such that $f(\bar{x}) - \langle \bar{v}, x \rangle = -f^*(\bar{v})$, finite. But by the version of Fermat's principle in 10.1 they are also the ones such that $0 \in \partial f_{\bar{v}}(\bar{x})$, this subgradient set being the same as $\partial f(\bar{x}) - \bar{v}$ (cf. 8.8(c)). Thus, $f(\bar{x}) + f^*(\bar{v}) = \langle \bar{v}, \bar{x} \rangle$ if and only if $\bar{v} \in \partial f(\bar{x})$. The rest follows now by symmetry. □

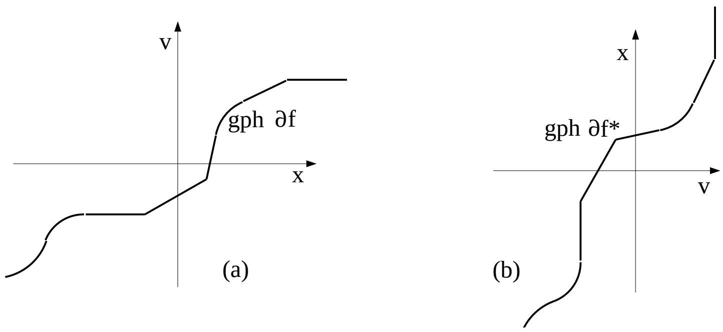

Fig. 11–2. Subgradient inversion for conjugate functions.

Subgradient relations, with normal cone relations as a special case, are widespread in the statement of optimality conditions. The inversion rule in 11.3 (illustrated in Figure 11–2) is therefore a key to writing such conditions in alternative ways and gaining other interpretations of them. That pattern will be prominent in our work with generalized Lagrangian functions and dual problems of optimization later in this chapter.

B. Special Cases of Conjugacy

Before proceeding with other features of the Legendre-Fenchel transform, let's observe that Theorem 11.1 covers, as special instances of the conjugacy correspondence, the fundamentals of cone polarity in 6.21 and support function theory in 8.24. This confirms that those earlier modes of dualization fit squarely in the picture now being unveiled.

11.4 Example (support functions and cone polarity).

(a) *For any set $C \subset \mathbb{R}^n$, the conjugate of the indicator function δ_C is the support function σ_C. On the other hand, for any positively homogeneous function h on $\mathbb{R}^n$ the conjugate h^* is the indicator δ_C of the set $C = \{x \mid \langle v, x\rangle \le h(v) \text{ for all } v\}$. In this sense, the correspondence between closed, convex sets and their support functions is imbedded within conjugacy:*

$$\delta_C \underset{*}{\longleftrightarrow} \sigma_C \quad \text{for } C \text{ a closed, convex, set.}$$

Under this correspondence one has

$$\bar{v} \in N_C(\bar{x}) \iff \bar{x} \in \partial\sigma_C(\bar{v}) \iff \bar{x} \in C,\ \langle \bar{v}, \bar{x}\rangle = \sigma_C(\bar{v}). \tag{11(4)}$$

(b) *For a cone $K \subset \mathbb{R}^n$, the conjugate of the indicator function δ_K is the indicator function δ_{K^*}. In this sense, the polarity correspondence for closed, convex cones is imbedded within conjugacy:*

$$\delta_K \underset{*}{\longleftrightarrow} \delta_{K^*} \quad \text{for } K \text{ a closed, convex cone.}$$

Under this correspondence one has

$$\bar{v} \in N_K(\bar{x}) \iff \bar{x} \in N_{K^*}(\bar{v}) \iff \bar{x} \in K,\ \bar{v} \in K^*,\ \bar{x} \perp \bar{v}. \tag{11(5)}$$

Detail. The formulas for the conjugate functions immediately reduce in these ways, and then 11.3 can be applied. □

In particular, the orthogonal subspace correspondence $M \leftrightarrow M^\perp$ is imbedded within conjugacy through $\delta_M^* = \delta_{M^\perp}$.

The support function correspondence has a bearing on the Legendre-Fenchel transform from a different angle too, namely in characterizing the effective domains of functions conjugate to each other.

11.5 Theorem (horizon functions as support functions). *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper, lsc and convex. The horizon function f^∞ is then the support function of* $\operatorname{dom} f^*$, *whereas $f^{*\infty}$ is the support function of* $\operatorname{dom} f$.

Proof. We have $(f^*)^* = f$ (by 11.1), and because of this symmetry it suffices to prove that the function $f^{*\infty} = (f^*)^\infty$ is the support function of $D := \operatorname{dom} f$. Fix any $v_0 \in \operatorname{dom} f^*$. We have for arbitrary $v \in \mathbb{R}^n$ and $\tau > 0$ that

$$\begin{aligned} f^*(v_0 + \tau v) &= \sup_{x \in D}\big\{\langle v_0 + \tau v, x\rangle - f(x)\big\} \\ &\le \sup_{x \in D}\big\{\langle v_0, x\rangle - f(x)\big\} + \tau \sup_{x \in D}\langle v, x\rangle = f^*(v_0) + \tau\sigma_D(v), \end{aligned}$$

hence $\big[f^*(v_0 + \tau v) - f^*(v_0)\big]/\tau \le \sigma_D(v)$ for all $v \in \mathbb{R}^n$, $\tau > 0$. Through 3(4) this guarantees that $f^{*\infty} \le \sigma_D$. On the other hand, for $v \in \mathbb{R}^n$ and $\beta \in \mathbb{R}$ with $f^{*\infty}(v) \le \beta$ one has $f^*(v_0 + \tau v) \le f^*(v_0) + \tau\beta$ for all $\tau > 0$, hence for any $x \in \mathbb{R}^n$ that

$$\begin{aligned} f(x) &\geq \langle v_0 + \tau v, x\rangle - f^*(v_0 + \tau v) \\ &\geq \langle v_0, x\rangle - f^*(v_0) + \tau\big(\langle v, x\rangle - \beta\big) \ \text{ for all } \ \tau > 0. \end{aligned}$$

This implies that $\langle v, x\rangle \leq \beta$ for all x with $f(x) < \infty$, so $D \subset \{x \mid \langle v, x\rangle \leq \beta\}$. Thus $\sigma_D(v) \leq \beta$, and we conclude that also $\sigma_D \leq f^{*\infty}$, so $\sigma_D = f^{*\infty}$. □

Another way that support functions come up in the dualization of properties of f and f^* is seen in connection with level sets. For simplicity in the following, we look only at 0-level sets, since $\operatorname{lev}_{\leq\alpha} f$ can be studied as $\operatorname{lev}_{\leq 0} f_\alpha$ for $f_\alpha = f - \alpha$, with $f_\alpha^* = f^* + \alpha$ by 11(3).

11.6 Exercise (support functions of level sets)**.** *If* $C = \{x \mid f(x) \leq 0\}$ *for a finite, convex function* f *such that* $\inf f < 0$, *then*

$$\sigma_C(v) \;=\; \inf_{\lambda > 0} \lambda f^*(\lambda^{-1} v) \ \text{ for all } \ v \neq 0.$$

Guide. Let h denote the function of v on the right side of the equation; take $h(0) = 0$. Show in terms of the 'pos' operation defined ahead of 3.48 that h is a positively homogeneous, convex function for which the points x satisfying $\langle v, x\rangle \leq h(v)$ for all v are the ones such that $f^{**}(x) \leq 0$. Argue that $f^{**} = f$ and hence via support function theory that $\sigma_C = \operatorname{cl} h$. Verify through 11.5 that $f^{*\infty} = \delta_{\{0\}}$ and in this way deduce from 3.48(b) that $\operatorname{cl} h = h$. □

Note that the roles of f and f^* in 11.6 could be reversed: the support functions for the level sets of f^*, when that function is finite, can be derived from f (as long as the convex function f is proper and lsc, so that $(f^*)^* = f$.)

While the polarity of cones is a special case of conjugacy of functions, the opposite is true as well, in a certain sense. This is not only interesting but valuable for certain theoretical purposes.

11.7 Exercise (conjugacy as cone polarity)**.** *For proper functions* f *and* g *on* $\mathbb{R}^n$, *consider in* $\mathbb{R}^{n+2}$ *the cones*

$$\begin{aligned} K_f &= \Big\{(x, \alpha, -\lambda) \,\Big|\, \lambda > 0,\ (x, \alpha) \in \lambda \operatorname{epi} f; \ \text{ or } \ \lambda = 0,\ (x, \alpha) \in \operatorname{epi} f^\infty\Big\}, \\ K_g &= \Big\{(v, -\mu, \beta) \,\Big|\, \mu > 0,\ (v, \beta) \in \mu \operatorname{epi} g; \ \text{ or } \ \mu = 0,\ (v, \beta) \in \operatorname{epi} g^\infty\Big\}. \end{aligned}$$

Then f *and* g *are conjugate to each other if and only if* K_f *and* K_g *are polar to each other.*

Guide. Verify that K_f is convex and closed if and only if f is convex and lsc; indeed, K_f is then the cone representing $\operatorname{epi} f$ in the ray space model for $\operatorname{csm} \mathbb{R}^{n+1}$. The cone K_g has a similar interpretation with respect to g, except for a reversal in the roles of last two components. Use the definition of conjugacy along with the relationships in 11.5 to analyze polarity. □

How does the duality between f and f^* affect situations where f represents a problem of optimization? Here are the central facts.

11.8 Theorem (dual properties in minimization)**.** *The properties of a proper, lsc, convex function* $f : I\!R^n \to \overline{I\!R}$ *are paired with those of its conjugate function* f^* *in the following manner.*

(a) $\inf f = -f^*(0)$ *and* $\operatorname{argmin} f = \partial f^*(0)$.

(b) $\operatorname{argmin} f = \{\bar{x}\}$ *if and only* f^* *is differentiable at* 0 *with* $\nabla f^*(0) = \bar{x}$.

(c) f *is level-coercive (or level-bounded) if and only if* $0 \in \operatorname{int}(\operatorname{dom} f^*)$.

(d) f *is coercive if and only if* $\operatorname{dom} f^* = I\!R^n$.

Proof. The first property in (a) simply re-expresses the definition of $f^*(0)$, while the second comes from 11.3 and the fact that $\operatorname{argmin} f$ consists of the points x such that $0 \in \partial f(x)$; cf. 10.1. In (b) this is elaborated through the fact that $\partial f^*(0) = \{\bar{x}\}$ if and only if f^* is strictly differentiable at 0, this by 9.18 and the fact that because f^* is itself proper, lsc and convex, f is regular with $\partial^\infty f^*(0) = \partial f^*(0)^\infty$ (see 7.27 and 8.11). The regularity of f^* implies further that f^* is strictly differentiable wherever it's differentiable (cf. 9.20).

In (c) we recall that f is level-coercive if and only if $f^\infty(w) > 0$ for all $w \neq 0$ (see 3.26(a)), whereas the convex set $D = \operatorname{dom} f^*$ has $0 \in \operatorname{int} D$ if and only if $\sigma_D(w) > 0$ for all $w \neq 0$ (see 8.29(a)). The equivalence comes from 11.5, where $f^\infty(w)$ is identified with $\sigma_D(w)$. (Recall too that a convex function is level-bounded if and only if it is level-coercive; 3.27.) Similarly, in (d) we are seeing an instance of the fact that a convex set is the whole space if and only if it isn't contained in any closed half-space, i.e., its support function is $\delta_{\{0\}}$. □

The dualizations in 11.8 can be extended through elementary conjugacy relations like the ones in 11(3). Thus, one has

$$\inf_x\bigl\{f(x) - \langle a, x\rangle\bigr\} = -f^*(a), \quad \operatorname{argmin}_x\bigl\{f(x) - \langle a, x\rangle\bigr\} = \partial f^*(a), \tag{11(6)}$$

the argmin being $\{b\}$ if and only if f^* is differentiable at a with $\nabla f^*(a) = b$. The function $f - \langle a, \cdot\rangle$ is level-coercive if and only if $a \in \operatorname{int}(\operatorname{dom} f^*)$.

So far, little has been said about how f^* can effectively be determined when f is given. Because of conjugacy's abstract uses in analysis and the dualization of properties for purposes of understanding them better, a formula for f beyond the defining one in 11(1) isn't always needed, but what about the times when it is? Ways of constructing f^* out of the conjugates of other functions that are part of the make up of f can be very helpful (and will be occupy our attention in due course), but somewhere along the line it's crucial to have a repertory of examples that can serve as building blocks, much as power functions, exponentials, logarithms and trigonometric expressions serve in classical differentiation and integration.

We've observed in 11.4(a) that f^* can sometimes be identified as a support function (here examples like 8.26 and 8.27 can be kept in mind), or in reverse as the indicator of a convex set defined by the system of linear constraints associated with a sublinear function (cf. 8.24) when f exhibits sublinearity. In this respect the results in 11.5 and 11.6 can be useful, and further also the subderivative-subgradient relations in Chapter 8 for functions that are subdifferentially regular. Then again, as in 11.4(b), f^* might be the indicator of a polar cone. For instance, the polar of $\mathbb{R}^n_+$ is $\mathbb{R}^n_-$, and the polar of a subspace M is $M^\perp$. The Farkas lemma in 6.45 and the relations between tangent cones and polar cones can provide assistance as well.

C. The Role of Differentiability

Beyond special cases such as these, there is the possibility of generating examples directly from the formula in 11(1) for f^* in terms of f. This may be intimidating, though, because it not only demands the calculation of a global supremum (the solution of a certain optimization problem), but requires this to be done parametrically—the supremum must be expressed as a function of the v element. For functions on $\mathbb{R}^1$, 'brute force' may succeed, but elsewhere some guidelines are needed. The next three examples will present important cases where f^* can be calculated from f by use of derivatives alone.

As long as f is convex and differentiable everywhere, one can hope to get the supremum in formula 11(1), and thereby the value of $f^*(v)$, by setting the gradient (with respect to x) of the expression $\langle v, x\rangle - f(x)$ equal to 0 and solving that equation for x. This is justified because the expression is concave with respect to x; the vanishing of its gradient corresponds therefore to the attainment of the global maximum. The equation in question is $v - \nabla f(x) = 0$, and its solutions are the vectors x, if any, belonging to $(\nabla f)^{-1}(v)$. An x identified in this manner can be substituted into $\langle v, x\rangle - f(x)$ to get $f^*(v)$. Pitfalls gape, however, in the fact that the range of the mapping ∇f might not be all of $\mathbb{R}^n$. For $v \notin \operatorname{rge} \nabla f$, the supremum would need to be determined through additional analysis. It might be ∞, with the meaning that $v \notin \operatorname{dom} f^*$, or it might be finite, yet not attained.

Putting such troubles aside to get a picture first of the nicest circumstances, one can ask what happens when ∇f is a one-to-one mapping from $\mathbb{R}^n$ onto $\mathbb{R}^n$, so that $(\nabla f)^{-1}$ is single-valued everywhere. Understandably, this is the historical case in which conjugate functions first attracted interest.

11.9 Example (classical Legendre transform)**.** *Let f be a finite, coercive, convex function of class $\mathcal{C}^2$ (twice continuously differentiable) on $\mathbb{R}^n$ whose Hessian matrix $\nabla^2 f(x)$ is positive-definite for every x. Then the conjugate $g = f^*$ is likewise a finite, coercive, convex function of class $\mathcal{C}^2$ on $\mathbb{R}^n$ with $\nabla^2 g(v)$ positive-definite for every v and $g^* = f$. The gradient mapping ∇f is one-to-one from $\mathbb{R}^n$ onto $\mathbb{R}^n$, and its inverse is ∇g; one has*

$$\begin{aligned} g(v) &= \big\langle (\nabla f)^{-1}(v), v\big\rangle - f\big((\nabla f)^{-1}(v)\big),\\ f(x) &= \big\langle (\nabla g)^{-1}(x), x\big\rangle - g\big((\nabla g)^{-1}(x)\big). \end{aligned} \tag{11(7)}$$

Moreover the matrices $\nabla^2 f(x)$ and $\nabla^2 g(v)$ are inverse to each other when $v = \nabla f(x)$, or equivalently $x = \nabla g(v)$ (then x and v are conjugate points).

Detail. The assumption on second derivatives makes f strictly convex (see 2.14). Then for a fixed $a \in \mathbb{R}^n$ in 11(6) we not only have through coercivity the attainment of the infimum but its attainment at a *unique* point x (by 2.6). Then $\nabla f(x) - a = 0$ by Fermat's principle. This line of reasoning demonstrates that for each $v \in \mathbb{R}^n$ there is a unique x with $\nabla f(x) = v$, i.e., the mapping ∇f is invertible. The first equation in 11(7) is immediate, and the rest of the assertions can then be obtained from 11.8(d) and differentiation of g, using the standard inverse mapping theorem. □

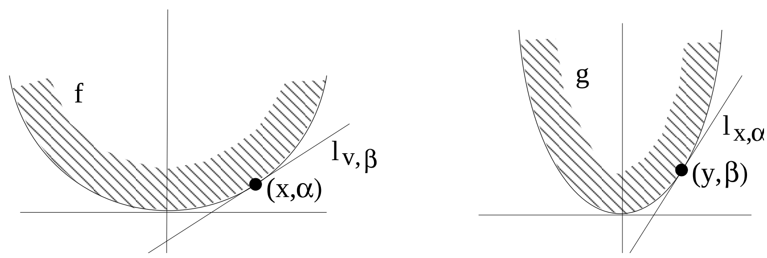

Fig. 11–3. Conjugate points in the classical setting.

The next example fits the pattern of the preceding one in part, but also illustrates how the approach to calculating f^* from the derivatives of f can be followed a bit more flexibly.

11.10 Example (linear-quadratic functions)**.** *Suppose*

$$f(x) = \tfrac{1}{2}\langle x, Ax\rangle + \langle a, x\rangle + \alpha$$

with $A \in \mathbb{R}^{n\times n}$ symmetric and positive-semidefinite, so that f is convex. If A is nonsingular, the conjugate function is

$$f^*(v) = \tfrac{1}{2}\big\langle v - a, A^{-1}(v-a)\big\rangle - \alpha.$$

At the other extreme, if $A = 0$, so that f is merely affine, the conjugate function is given by $f^*(v) = \delta_{\{a\}}(v) - \alpha$.

In general, the column space of A (the range of $x \mapsto Ax$) is a linear subspace L, and there is a unique symmetric, positive-semidefinite matrix $A^\dagger$ (the pseudo-inverse of A) having $A^\dagger A = AA^\dagger =$ [orthogonal projector on L]. The conjugate function is given then by

$$f^*(v) = \begin{cases} \frac{1}{2}\big\langle v-a, A^\dagger(v-a)\big\rangle - \alpha & \text{when } v - a \in L,\\ \infty & \text{when } v - a \notin L.\end{cases}$$

Detail. The nonsingular case fits the pattern of 11.9, while the affine case is obvious on its own. The general case is made simple by reducing to $a = 0$ and $\alpha = 0$ through the relations 11(3) and invoking a change of coordinates that diagonalizes the matrix A. □

11.11 Example (self-conjugacy). *The function $f(x)=\frac{1}{2}|x|^2$ on $I\!R^n$ has $f^*=f$ and is the only function with this property.*

Detail. The self-conjugacy of this function is evident as the special case of 11.10 in which $A=I$, $a=0$. Its uniqueness in this respect is seen as follows. If $f=f^*$, then $f^{**}=f$ by 11.1, and f is proper by 11.2. Formula 11(1) gives in this case $f(v)+f(x)\geq\langle v,x\rangle$ for all x and v, hence with $x=v$ that $f(x)\geq\frac{1}{2}|x|^2$ for all x. Passing to conjugates in this inequality, one sees on the other hand that $f^*(v)\leq\frac{1}{2}|v|^2$ for all v, hence from $f^*=f$ that $f(x)\leq\frac{1}{2}|x|^2$ for all x. Therefore, $f(x)=\frac{1}{2}|x|^2$ for all x. □

The formula in 11.6 for the support function of a level set can be illustrated through Example 11.10. For a convex set of the form

$$C \;=\; \big\{x \,\big|\, \tfrac{1}{2}\langle x,Ax\rangle+\langle a,x\rangle+\alpha\leq 0\big\}$$

with A symmetric and positive-definite, and such that the inequality is satisfied strictly by at least one x, one necessarily has $\langle a,A^{-1}a\rangle-2\alpha>0$ (by 11.8(a) because this quantity is $2f^*(0)$), and thus the expression

$$\sigma_C(v)=\beta\sqrt{\langle v,A^{-1}v\rangle}-\langle b,v\rangle \quad\text{for}\quad b=A^{-1}a,\;\; \beta=\sqrt{\langle a,A^{-1}a\rangle-2\alpha}.$$

In general, if one of the functions in a general conjugate pair is finite and coercive, so too must be the other function; this is clear from 11.8(d). Otherwise, at least one of the two functions in a conjugate pair must take on the value ∞ somewhere and thus have some convex set other than $I\!R^n$ itself as its effective domain. The support function relation in 11.5 shows in these cases how $\operatorname{dom} f^*$ relates to properties of f through the horizon function f^∞. For the same reason, since $f^{**}=f$ (when f is proper, lsc and convex), $\operatorname{dom} f$ relates to properties of f^* through the way it determines $f^{*\infty}$. Information about effective domains facilitates the calculation of conjugates in many cases.

The following example illustrates this principle as an extension of the method for calculating f^* from the derivatives of f.

11.12 Example (log-exponential function and entropy). *For $f(x)=\operatorname{logexp}(x)$, the conjugate f^* is the entropy function g defined for $v=(v_1,\ldots,v_n)$ by*

$$g(v) \;=\; \begin{cases} \sum_{j=1}^n v_j\log v_j & \text{when } v_j\geq 0,\ \sum_{j=1}^n v_j=1,\\ \infty & \text{otherwise,}\end{cases}$$

with $0\log 0=0$. The support function of the set $C=\big\{x\,\big|\,\operatorname{logexp}(x)\leq 0\big\}$ is the function $h:I\!R^n\to\overline{I\!R}$ defined under the same convention by

$$h(v) \;=\; \begin{cases} \sum_{j=1}^n v_j\log v_j-\big(\sum_{j=1}^n v_j\big)\log\big(\sum_{j=1}^n v_j\big) & \text{when } v_j\geq 0,\\ \infty & \text{otherwise.}\end{cases}$$

Detail. The horizon function of $f(x)=\operatorname{logexp}(x)$ is $f^\infty(x)=\operatorname{vecmax} x$ by 3(5), and this is the support function of the unit simplex C consisting of the vectors $v\geq 0$ with $v_1+\cdots+v_n=1$, as already noted in 8.26. Since f is a finite, convex function (cf. 2.16), it's in particular a proper, lsc, convex function (cf. 2.36). We may conclude from 11.5 that $\operatorname{dom} f^*$ has the same support function as C and therefore has $\operatorname{cl}(\operatorname{dom} f^*) = C$ (cf. 8.24). Hence in terms of relative interiors (cf. 2.40),

$$\operatorname{rint}(\operatorname{dom} f^*) = \operatorname{rint} C = \big\{ v \,\big|\, v_j > 0,\ \textstyle\sum_{j=1}^n v_j = 1 \big\}.$$

From the formula $\nabla f(x) = \sigma(x)^{-1}(e^{x_1}, \ldots, e^{x_n})$ for $\sigma(x) := e^{x_1} + \cdots + e^{x_n}$ it is apparent that each $\bar{v} \in \operatorname{rint} C$ is of the form $\nabla f(\bar{x})$ for $\bar{x} = (\log \bar{v}_1, \ldots, \log \bar{v}_n)$. The inequality $f(x) \geq f(\bar{x}) + \langle \nabla f(\bar{x}), x - \bar{x} \rangle$ in 2.14(b) yields

$$\sup_x \big\{ \langle x, \nabla f(\bar{x}) \rangle - f(x) \big\} = \langle \bar{x}, \nabla f(\bar{x}) \rangle - f(\bar{x}),$$

which is the same as

$$\sup_x \big\{ \langle x, \bar{v} \rangle - f(x) \big\} = \textstyle\sum_{j=1}^n (\log \bar{v}_j) \bar{v}_j - \log\big(\sum_{j=1}^n \bar{v}_j \big) = g(\bar{v}).$$

Thus $f^* = g$ on $\operatorname{rint} C$. The closure formula in 2.35 as translated to the context of relative interiors shows then that these functions agree on all of C, therefore on all of $\mathbb{R}^n$.

The function h is $\operatorname{pos} g$, where $g(0) = \infty$ and $\inf f = -\infty$; cf. 11.8(a). According to 11.6, h is then the support function of $\operatorname{lev}_{\leq 0} f$. $\square$

With minor qualifications on the boundaries of domains, differentiability itself dualizes under the Legendre-Fenchel transform to strict convexity.

11.13 Theorem (strict convexity versus differentiability). *The following properties are equivalent for a proper, lsc, convex function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and its conjugate function f^*:*

(a) *f is almost differentiable, in the sense that f is differentiable on the open, convex set $\operatorname{int}(\operatorname{dom} f)$, which is nonempty, but $\partial f(x) = \emptyset$ for all points $x \in \operatorname{dom} f \setminus \operatorname{int}(\operatorname{dom} f)$, if any;*

(b) *f^* is almost strictly convex, in the sense that f^* is strictly convex on every convex subset of $\operatorname{dom} \partial f^*$ (hence on $\operatorname{rint}(\operatorname{dom} f^*)$, in particular).*

Likewise, the function f^* is almost differentiable if and only if f is almost strictly convex.

Proof. Since $f = (f^*)^*$ under our assumptions (by 11.1), there's symmetry between the first equivalence asserted and the one claimed at the end. We can just as well work at verifying the latter. As seen from 11.8 (and its extension explained after the proof of that result), f^* is almost differentiable if and only if, for every $a \in \mathbb{R}^n$ such that the set $\operatorname{argmin}_x \big\{ f(x) - \langle a, x \rangle \big\} = \partial f^*(a)$ is nonempty, it's actually a singleton. Our task is to show that this holds if and only if f is almost strictly convex.

Certainly if for some a this minimizing set, which is convex, contained two different points x_0 and x_1, it would contain $x_\tau := (1-\tau)x_0 + \tau x_1$ for all $\tau \in (0,1)$. Because $x_\tau \in \partial f^*(a)$ we would have $a \in \partial f(x_\tau)$ by 11.3, so the line segment joining x_0 and x_1 would lie in $\operatorname{dom} \partial f$. From the fact that

$\inf_x\{f(x) - \langle a, x\rangle\} = -f^*(a)$, we would have $f(x_\tau) - \langle a, x_\tau\rangle = -f^*(a)$ for $\tau \in (0,1)$. This implies $f(x_\tau) = (1-\tau)f(x_0) + \tau f(x_1)$ for $\tau \in (0,1)$, since $\langle a, x_\tau\rangle = (1-\tau)\langle a, x_0\rangle + \tau\langle a, x_1\rangle$. Then f isn't almost strictly convex.

Conversely, if f fails to be almost strictly convex there must exist $x_0 \neq x_1$ such that the points x_τ on the line segment joining them belong to $\operatorname{dom}\partial f$ and satisfy $f(x_\tau) = (1-\tau)f(x_0) + \tau f(x_1)$. Fix any $\bar\tau \in (0,1)$ and any $a \in \partial f(x_{\bar\tau})$. From 11.3 and formula 11(1) for f^* in terms of f, we have

$$f(x_\tau) \geq \langle a, x_\tau\rangle - f^*(a) \ \text{ for all } \ \tau \in (0,1), \ \text{ with equality for } \ \tau = \bar\tau.$$

The affine function $\varphi(\tau) := f(x_\tau)$ on $(0,1)$ thus attains its minimum at the intermediate point $\bar\tau$. But then φ has to be constant on $(0,1)$. In other words, for all $\tau \in (0,1)$ we must have $f(x_\tau) = \langle a, x_\tau\rangle - f^*(a)$, hence $x_\tau \in \partial f^*(a)$ by 11.3. In this event $\partial f^*(a)$ isn't a singleton. □

The property in 11.13(a) of being *almost* differentiable can be identified with the single-valuedness of the mapping ∂f relative to its domain (see 9.18, recalling from 7.27 that proper, lsc, convex functions are regular). It implies f is continuously differentiable—smooth—on $\operatorname{int}(\operatorname{dom} f)$; cf. 9.20.

D. Piecewise Linear-Quadratic Functions

Differentiability isn't the only tool available for understanding the nature of conjugate functions, of course. A major class of *nondifferentiable* functions with nice behavior under the Legendre-Fenchel transform consists of the convex functions that are piecewise linear (see 2.47) or more generally piecewise linear-quadratic (see 10.20).

11.14 Theorem (piecewise linear-quadratic functions in conjugacy)**.** *Suppose that $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper, lsc and convex. Then*

(a) *f is piecewise linear if and only if f^* has this property;*

(b) *f is piecewise linear-quadratic if and only if f^* has this property.*

For proving part (b) we'll need a lemma, which is of some interest in itself. We take care of this first.

11.15 Lemma (linear-quadratic test on line segments)**.** *In order that f be linear-quadratic relative to a convex set $C \subset \mathbb{R}^n$, in the sense of being expressible by a formula of type $f(x) = \frac{1}{2}\langle x, Ax\rangle + \langle a, x\rangle + \alpha$ for $x \in C$, it is necessary and sufficient that f be linear-quadratic relative to every line segment in C.*

Proof. The condition is trivially necessary, so the challenge is proving its sufficiency. Without loss of generality we can focus on the case where $\operatorname{int} C \neq \emptyset$ (cf. 2.40 and the discussion preceding it). It's enough actually to demonstrate that the condition implies f is linear-quadratic relative to $\operatorname{int} C$, because the formula obtained on $\operatorname{int} C$ must then extend to the rest of C through the fact that when a boundary point x of C is joined by a line segment to an interior point, all of the segment except x itself lies in $\operatorname{int} C$ (see 2.33).

We claim next that if f is linear-quadratic in some neighborhood of each point of $\operatorname{int} C$, then it's linear-quadratic relative to $\operatorname{int} C$. Consider any two points x_0 and x_1 of $\operatorname{int} C$. We'll show that the formula around x_0 must agree with the formula around x_1.

The line segment $[x_0, x_1]$ is a compact set, every point of which has an open ball relative to which f is linear-quadratic, and it can therefore be covered by a finite collection of such open balls, say O_k for $k = 1, \ldots, r$, each with a formula $f(x) = \frac{1}{2}\langle x, A_k x\rangle + \langle a_k, x\rangle + \alpha_k$. If two sets O_{k_1} and O_{k_2} overlap, their formulas have to agree on the intersection; this implies that $A_{k_1} = A_{k_2}$, $a_{k_1} = a_{k_2}$ and $\alpha_{k_1} = \alpha_{k_2}$. But as one moves along $[x_0, x_1]$ from x_0 to x_1, each transition out of one set O_k and into another passes through a region of overlap (again because of the line segment principle for convex sets, or more generally because line segments are connected sets). Thus, all the formulas for $k = 1, \ldots, r$ agree.

Having reduced the task to proving that f is linear-quadratic relative to a neighborhood of each point of $\operatorname{int} C$, we can take such neighborhoods to be cubes. The question then is whether, if f is linear-quadratic on every line segment in a certain cube, it must be linear-quadratic relative to the cube.

A cube in $I\!R^n$ is a product of n intervals, so an induction argument can be contemplated in which the product grows by one interval at a time until the cube is built up, and at each stage the linear-quadratic property of f relative to the partial product is verified. For a single interval, as the starter, the property holds by hypothesis.

To validate the induction argument we only have to show that *if U and V are convex neighborhoods of the origin in $I\!R^p$ and $I\!R^q$ respectively, and if a function f on $U \times V$ is such that $f(u, v)$ is linear-quadratic in u for fixed v, linear-quadratic in v for fixed u, and moreover f is linear-quadratic relative to all line segments in $U \times V$, then f is linear-quadratic relative to $U \times V$ as a whole.* We go about this by first using the property in u to express

$$f(u, v) \;=\; \tfrac{1}{2}\big\langle u, A(v)u\big\rangle + \big\langle a(v), u\big\rangle + \alpha(v) \;\text{ for }\; u \in U \;\text{ when }\; v \in V. \qquad 11(8)$$

We'll demonstrate next, invoking the linear-quadratic property of $f(u, v)$ in v, that $\alpha(v)$ and each component of the vector $a(v)$ and the matrix $A(v)$ must be linear-quadratic as a function of $v \in V$. For $\alpha(v)$ this is clear from having $\alpha(v) = f(0, v)$. In the case of $A(v)$ we observe that

$$\big\langle u', A(v)u\big\rangle \;=\; f(u + u', v) - f(u, v) - f(u', v) + \alpha(v)$$

for any u and u' in $I\!R^p$ small enough that they and $u + u'$ belong to U. Hence $\langle u', A(v)u\rangle$ is linear-quadratic in $v \in V$ for any such u and u'. Choosing u and u' among $\varepsilon e^1, \ldots, \varepsilon e^p$ for small $\varepsilon > 0$ and the canonical basis vectors e^k for $I\!R^p$ (where e^k has 1 in kth position and 0's elsewhere), we deduce that for every row i and column j the component $A_{ij}(v)$ is linear-quadratic in v. By writing

$$\langle a(v), u\rangle = f(u,v) - \tfrac{1}{2}\langle u, A(v)u\rangle - \alpha(v),$$

where the right side is now known to be linear-quadratic in $v \in V$, we see that $\langle a(v), u\rangle$ has this property for each u sufficiently near to 0. Again by choosing u from among $\varepsilon e^1, \ldots, \varepsilon e^p$ we are able to conclude that each component $a_j(v)$ of $a(v)$ is linear-quadratic in $v \in V$.

When such linear-quadratic expressions in v for $\alpha(v)$, $a(v)$ and $A(v)$ are introduced in 11(8), a polynomial formula for $f(u,v)$ is obtained in which there are no terms of degree higher than 4. We have to show that there really aren't any terms of degree higher than 2, i.e., that f is linear-quadratic relative to $U \times V$ as a whole. We already know that $\alpha(v)$ has no higher-order terms, so the issue concerns $a(v)$ and $A(v)$.

This is where we bring in the last of our assumptions, that f is linear-quadratic on all line segments in $U \times V$. We'll only need to look at line segments that join an arbitrary point $(\bar{u}, \bar{v}) \in U \times V$ to $(0,0)$. The assumption means then that $f(\theta\bar{u}, \theta\bar{v})$ is linear-quadratic in $\theta \in [0,1]$. The argument just presented for reducing to individual components can be repeated by looking at $f(\theta[\bar{u} + \bar{u}'])$ with the vectors $\bar{u}$ and $\bar{u}'$ chosen from among $\varepsilon e^1, \ldots, \varepsilon e^p$, and so forth, to see that $\theta^2 A_{ij}(\theta\bar{v})$ and $\theta a_j(\theta\bar{v})$ are polynomials of at most degree 2 in θ for every choice of $\bar{v} \in V$. In the linear-quadratic expressions for $A_{ij}(v)$ and $a_j(v)$ as functions of v, it is obvious then that $A_{ij}(v)$ has to be constant in v, while $a_j(v)$ can at most have first-order terms in v. This finishes the proof. $\square$

Proof of 11.14. The justification of (a) is relatively easy on the basis of earlier results. When the convex function f is piecewise linear, it can be expressed in the manner of 3.54: for some choice of vectors a_i and scalars c_i,

$$f(x) = \begin{cases} \text{infimum of } \; t_1 c_1 + \cdots + t_m c_m + t_{m+1} c_{m+1} + \cdots + t_r c_r \\ \text{subject to } \; t_1 a_1 + \cdots + t_m a_m + t_{m+1} a_{m+1} + \cdots + t_r a_r = x \\ \text{with } \; t_i \geq 0 \text{ for } i = 1, \ldots, r, \; \sum_{i=1}^m t_i = 1. \end{cases}$$

From $f^*(v) = \sup_x \{\langle v, x\rangle - f(x)\}$ we get $f^*(v) = \max_{i=1,\ldots,m}\{\langle v, a_i\rangle - c_i\} + \delta_C$ for the polyhedral set $C := \{v \mid \langle v, a_i\rangle \leq c_i \text{ for } i = m+1, \ldots, r\}$. This signifies by 2.49 that f^* is piecewise linear. On the other hand, if f^* is piecewise linear, then so is f^{**} by this argument; but $f^{**} = f$.

For (b), suppose now that f is piecewise linear-quadratic: for $D := \operatorname{dom} f$ there are polyhedral sets C_k, $k = 1, \ldots, r$, such that $D = \bigcup_{k=1}^r C_k$ and

$$f(x) = \tfrac{1}{2}\langle x, A_k x\rangle + \langle a_k, x\rangle + \alpha_k \text{ when } x \in C_k. \qquad 11(9)$$

Our task is to show that f^* has a similar representation. We'll base our argument on the fact in 11.3 that

$$f^*(v) = \langle v, x\rangle - f(x) \text{ for any } x \text{ with } v \in \partial f(x). \qquad 11(10)$$

This requires careful investigation of the structure of the mapping ∂f.

Recall from 10.21 that the convex set $D = \operatorname{dom} f$, as the union of finitely many polyhedral sets C_k, is itself polyhedral. Any polyhedral set may be represented as the intersection of a finite collection of closed half-spaces, so we can contemplate a finite collection $\mathcal{H}$ of closed half-spaces in $I\!R^n$ such that (1) each of the sets D, $C_1, \ldots, C_r$ is the intersection of a subcollection of the half-spaces in $\mathcal{H}$, and (2) for every $H \in \mathcal{H}$ the opposite closed half-space H' (meeting H in a shared hyperplane) is likewise in $\mathcal{H}$.

Let $J_x = \{H \in \mathcal{H} \,|\, x \in H$ for each $x \in D$. Let $\mathcal{J}$ consist of all $J \subset \mathcal{H}$ such that $J = J_x$ for some $x \in D$, and for each $J \in \mathcal{J}$ let D_J be the intersection of the half-spaces $H \in J$. It is clear that each D_J is a nonempty polyhedral set contained in D; in fact, the half-spaces in $\mathcal{H}$ that intersect to form C_k belong to J_x if $x \in C_k$), so that $D_J \subset C_k$ when $J = J_x$ for any $x \in C_k$.

For each $J \in \mathcal{J}$, let $F_J = \operatorname{rint} D_J$, recalling that then $D_J = \operatorname{cl} F_J$. We claim that $J = J_x$ if and only if $x \in F_J$. For the half-spaces $H \in J_x$, there are only two possibilities: either $x \in \operatorname{int} H$ or x lies on the boundary of H, which corresponds to having both H and the opposite half-space H' belong to J_x. Thus, for $J = J_x$, D_J is the intersection of various hyperplanes along with some closed half-spaces having x in their associated open half-spaces. That intersection is the relatively open set F_J. Hence $x \in F_J$. On the other hand, for any x' in this set F_J, and in particular D_J, we have $J_{x'} \supset J = J_x$. If there were a half-space H in $J_{x'} \backslash J_x$, then x would have to lie outside of H, or more specifically, in the interior of the opposite half-space H' (likewise belonging to $\mathcal{H}$). In that case, however, $\operatorname{int} H'$ is one of the open half-spaces that includes F_J, and hence contains x', in contradiction to x' being in H. Thus, any $x' \in F_J$ must have $J_{x'} = J$. Indeed, we see from this that $\{F_J \,|\, J \in \mathcal{J}\}$ is a finite partition of D, comprised in effect of the equivalence classes under the relation that $x' \sim x$ when $J_{x'} = J_x$. Moreover, if any F_J touches a sets C_k, it must lie entirely in C_k, and the same is true then for its closure, namely D_J. In other words, the index set $K(x) = \{k \,|\, x \in C_k\}$ is the same set $K(J)$ for all $x \in F_J$.

It was shown in the proof of 10.21 that $df(x)$ is piecewise linear with $\operatorname{dom} df(\bar{x}) = T_D(x) = \bigcup_{k \in K(x)} T_{C_k}(x)$ and $df(x)(w) = \langle A_k x + a_k, w\rangle$ when $w \in T_{C_k}(x)$. Because f, being a proper, lsc, convex function, is regular (cf. 7.27), we know that $\partial f(x)$ consists of the vectors v such that $\langle v, w\rangle \le df(x)(w)$ for all $w \in I\!R^n$ (see 8.30). Hence

$$\begin{aligned}\partial f(x) &= \bigcap\nolimits_{k \in K(x)} \Big\{ v \,\Big|\, \langle v - A_k x - a_k, w\rangle \le 0 \text{ for all } w \in T_{C_k}(x) \Big\} \\ &= \bigcap\nolimits_{k \in K(x)} \Big\{ v \,\Big|\, v - A_k x - a_k \in N_{C_k}(x) \Big\}.\end{aligned}$$

In this we appeal to the polarity between $T_{C_k}(x)$ and $N_{C_k}(x)$, which results from C_k being convex (cf. 6.24). Observe next that the normal cone $N_{C_k}(x)$ (polyhedral) must be the same for all x in a given F_J. That's because having $v \in N_{C_k}(x)$ corresponds to the maximum of $\langle v, x'\rangle$ over $x' \in C_k$ being attained at x, and by virtue of F_J being relatively open, that can't happen unless this linear function is constant on F_J (and therefore attains its maximum at *every* point of F_J). This common normal cone can be denoted by $N_k(J)$, and in terms of the common index set $K(x) = K(J)$ for $x \in F_J$, we then have

$$\langle v, x' \rangle = \langle v, x \rangle \text{ for all } x,\, x' \in F_J \text{ when } v \in N_k(J),\ k \in K(J), \qquad 11(11)$$

along with $\partial f(x) = \{ v \mid v - a_k - A_k x \in N_k(J) \text{ for all } k \in K(J) \}$ when $x \in F_J$. Consider for each $J \in \mathcal{J}$ the polyhedral set

$$G_J = \{ (x, v) \mid x \in D_J \text{ and } v - a_k - A_k x \in N_k(J) \text{ for all } k \in K(J) \},$$

which is the closure of the analogous set with F_J in place of D_J. Because $\operatorname{gph} \partial f$ is closed (cf. 11.3), it follows now that $\operatorname{gph} \partial f = \bigcap_{J \in \mathcal{J}} G_J$.

For each $J \in \mathcal{J}$ let E_J be the image of G_J under $(x, v) \mapsto v$, which like G_J is polyhedral by 3.55(a), therefore closed. Since $\operatorname{dom} \partial f^* = \operatorname{rge} \partial f$ (by the inversion rule in 11.3), it follows that $\operatorname{dom} \partial f^* = \bigcup_{J \in \mathcal{J}} E_J$. Hence $\operatorname{dom} \partial f^*$ is closed, because the union of finitely many closed sets is closed. But since f^* is lsc and proper, $\operatorname{dom} \partial f^*$ is dense in $\operatorname{dom} f^*$ (see 8.10). The union of the polyhedral sets E_J is thus $\operatorname{dom} f^*$.

All that's left now is to show f^* is linear-quadratic relative to each set E_J. We'll appeal to Lemma 11.15. Consider any v_0 and v_1 in a given E_J, coming from G_J, and choose any x_0 and x_1 such that (x_0, v_0) and (x_1, v_1) belong to G_J. Then the pair $(x_\tau, v_\tau) := (1 - \tau)(x_0, v_0) + \tau(x_1, v_1)$ belongs to G_J too, so that $v_\tau \in \partial f(x_\tau)$. From 11(9) and 11(10) we get, for any $k \in K(J)$, that $f^*(v_\tau) = \langle v_\tau, x_\tau \rangle - f(x_\tau) = \langle v_\tau - A_k x_\tau - a_k, x_\tau \rangle - \alpha_k + \frac{1}{2} \langle x_\tau, A_k x_\tau \rangle = \langle v_\tau - A_k x_\tau - a_k, x_0 \rangle - \alpha_k + \frac{1}{2} \langle x_\tau, A_k x_\tau \rangle$, where the last equation is justified through the fact that $x_\tau = x_0 + \tau(x_1 - x_0)$ but $\langle v_\tau - A_k x_\tau - a_k,\, x_1 - x_0 \rangle = 0$ by 11(11). This expression for $f^*(v_\tau)$, being linear-quadratic in $\tau \in [0, 1]$, gives us what was required.

The fact that f^* is piecewise linear-quadratic *only* if f is piecewise linear-quadratic follows now by symmetry, because $f = f^{**}$. □

11.16 Corollary (minimum of a piecewise linear-quadratic function)**.** *For any proper, convex, piecewise linear-quadratic function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, if $\inf f$ is finite, then $\operatorname{argmin} f$ is nonempty and polyhedral.*

Proof. We apply 11.8(a) with the knowledge that f^* is piecewise linear-quadratic like f, so that when $0 \in \operatorname{dom} f^*$ the set $\partial f^*(0)$ must be nonempty and polyhedral (cf. 10.21). □

11.17 Corollary (polyhedral sets in duality)**.**

(a) *A closed, convex set C is polyhedral if and only if its support function σ_C is piecewise linear.*

(b) *A closed, convex cone K is polyhedral if and only if its polar cone K^* is polyhedral.*

Proof. This specializes 11.14(b) to the correspondences in 11.4. A convex indicator δ_C is piecewise linear if and only if C is polyhedral. The cone fact could also be deduced right from the Minkowski-Weyl theorem in 3.52. □

The preservation of the piecewise linear-quadratic property in passing to the conjugate of a given function, as in Theorem 11.14(b), is illustrated in Figure 11–4. As the figure suggests, this duality is closely tied to a property of the associated subgradient mappings through the inversion rule in 11.3. It will later be established in 12.30 that a convex function is piecewise linear-quadratic if and only if its subgradient mapping is piecewise polyhedral as defined in 9.57. The inverse of a piecewise polyhedral mapping is obviously still piecewise polyhedral.

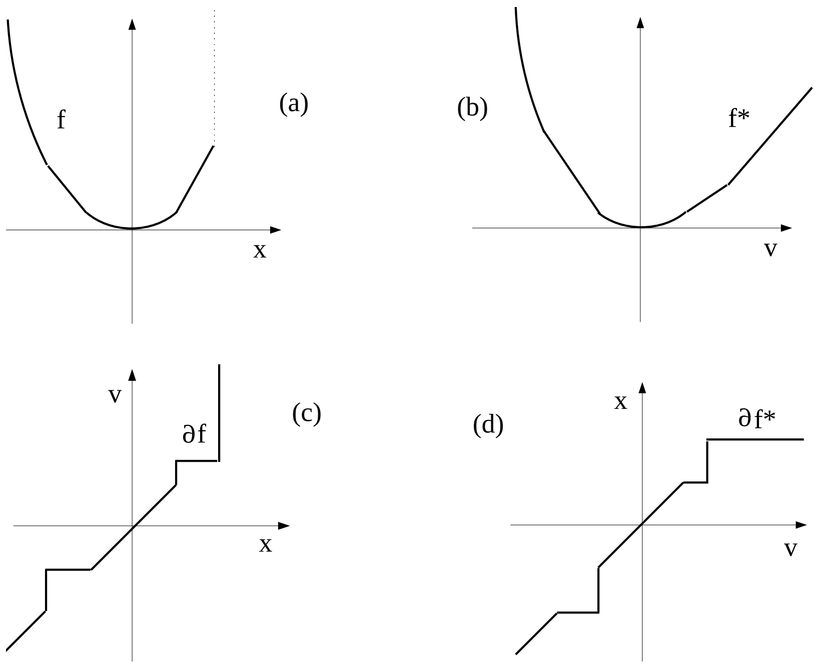

Fig. 11–4. Conjugate piecewise linear-quadratic functions.

An important application of the conjugacy in Theorem 11.14 comes up in the following class of functions $\theta : I\!R^m \to \overline{I\!R}$, which are useful in setting up 'penalty' expressions $\theta\big(f_1(x),\ldots,f_m(x)\big)$ in composite formats of optimization.

11.18 Example (piecewise linear-quadratic penalties)**.** *For a nonempty polyhedral set $Y \subset I\!R^m$ and a symmetric positive-semidefinite matrix $B \in I\!R^{m\times m}$ (possibly $B = 0$), the function $\theta_{Y,B} : I\!R^n \to \overline{I\!R}$ defined by*

$$\theta_{Y,B}(u) := \sup_{y\in Y}\Big\{\langle y,u\rangle - \tfrac{1}{2}\langle y,By\rangle\Big\}$$

is proper, convex and piecewise linear-quadratic. When $B = 0$, it is piecewise linear; $\theta_{Y,0} = \sigma_Y$ (support function). In general,

$$\operatorname{dom}\theta_{Y,B} = (Y^\infty \cap \ker B)^* =: D_{Y,B}, \quad \textit{where} \quad \ker B := \big\{y \,\big|\, By = 0\big\};$$

this is a polyhedral cone, and it is all of $\mathbb{R}^m$ *if and only if* $Y^\infty \cap \ker B = \{0\}$. *The subgradients of* $\theta_{Y,B}$ *are given by*

$$\begin{aligned}\partial\theta_{Y,B}(u) &= \operatorname*{argmax}_{y\in Y}\Big\{\langle y,u\rangle - \tfrac{1}{2}\langle y,By\rangle\Big\}\\ &= \big\{y \,\big|\, u - By \in N_Y(y)\big\} = (N_Y + B)^{-1}(u),\\ \partial^\infty\theta_{Y,B}(u) &= \big\{y \in Y^\infty \cap \ker B \,\big|\, \langle y,u\rangle = 0\big\} = N_{D_{Y,B}}(u),\end{aligned}$$

these sets being polyhedral as well, while the subderivative function $d\theta_{Y,B}$ *is piecewise linear and expressed by the formula*

$$d\theta_{Y,B}(u)(z) = \sup\big\{\langle y,z\rangle \,\big|\, y \in (N_Y + B)^{-1}(u)\big\}.$$

Detail. We have $\theta_{Y,B} = (\delta_Y + j_B)^*$ for $j_B(y) := \frac{1}{2}\langle y,By\rangle$. The function $\delta_Y + j_B$ is proper, convex and piecewise linear-quadratic, and $\theta_{Y,B}$ therefore has these properties as well by 11.14. In particular, the effective domain of $\theta_{Y,B}$ is a polyhedral set, hence closed. The support function of this effective domain is $(\delta_Y + j_B)^\infty$ by 11.5, and

$$(\delta_Y + j_B)^\infty \;=\; \delta_Y^\infty + j_B^\infty \;=\; \delta_{Y^\infty} + \delta_{\ker B} \;=\; \delta_{Y^\infty\cap\ker B}.$$

Hence by 11.4, $\operatorname{dom}\theta_{Y,B}$ must be the polar cone $(Y^\infty \cap \ker B)^*$, which is $\mathbb{R}^m$ if and only if $Y^\infty \cap \ker B$ is the zero cone.

The argmax formula for $\partial\theta_{Y,B}(u)$ specializes the argmax part of 11.3. The maximum of the concave function $h(y) = \langle y,u\rangle - \frac{1}{2}\langle y,By\rangle$ over Y is attained at y if and only if the gradient $\nabla h(y) = u - By$ belongs to $N_Y(y)$; cf. 6.12. That yields the other expressions for $\partial\theta_{Y,B}(u)$. We know from the convexity of $\theta_{Y,B}$ that $\partial^\infty\theta_{Y,B}(u) = N_{D_{Y,B}}(u)$; cf. 8.12. Since the cones $D_{Y,B}$ and $Y^\infty \cap \ker B$ are polar to each other, we have from 11.4(b) that $y \in N_{D_{Y,B}}(u)$ if and only if $u \in D_{Y,B}$, $y \in Y^\infty \cap \ker B$, and $u \perp y$.

On the basis of 10.21, the sets $\partial\theta_{Y,B}(u)$ and $\partial^\infty\theta_{Y,B}(u)$ are polyhedral and the function $d\theta_{Y,B}(u)$ is piecewise linear. The formula for $d\theta_{Y,B}(u)(z)$ merely expresses the fact that this is the support function of $\partial\theta_{Y,B}(u)$. □

E. Polar Sets and Gauges

While most of the major duality correspondences, like convex sets versus sublinear functions, or polarity of convex cones, fit directly within the framework of conjugate convex functions as in 11.4, others, like polarity of convex sets that aren't necessarily cones but contain the origin, fit obliquely. In the next example we draw on the notion of the gauge γ_C of a set C in 3.50.

11.19 Example (general polarity of sets)**.** *For any set* $C \subset \mathbb{R}^n$ *with* $0 \in C$, *the polar of* C *is the set*

$$C^\circ := \{v \mid \langle v,x\rangle \le 1 \text{ for all } x \in C\},$$

which is closed and convex with $0 \in C^\circ$*; when* C *is a cone,* C° *agrees with the polar cone* C^**. The bipolar of* C*, which is the set*

$$C^{\circ\circ} := (C^\circ)^\circ = \{x \mid \langle v,x\rangle \le 1 \text{ for all } v \in C^\circ\},$$

agrees always with $\operatorname{cl}(\operatorname{con} C)$*. Thus,* $C^{\circ\circ} = C$ *when* C *is a closed, convex set containing the origin, so the transformation* $C \mapsto C^\circ$ *maps that class of sets one-to-one onto itself. This correspondence is connected to conjugacy through the associated gauges, which obey the rule that*

$$\gamma_C = \sigma_{C^\circ} \underset{*}{\longleftrightarrow} \delta_{C^\circ}, \qquad \gamma_{C^\circ} = \sigma_C \underset{*}{\longleftrightarrow} \delta_C.$$

When two convex sets are polar to each other, one says that their gauges are polar to each other as well.

Detail. The facts about C° and $C^{\circ\circ}$ are evident from the envelope description of convex sets in 6.20. Clearly $C^{\circ\circ}$ is the intersection of all the closed half-spaces that include C and have the origin in their interior.

Because the gauge γ_C is proper, lsc and sublinear (cf. 3.50), we know from 8.24 that it's the support function of a certain nonempty, closed, convex set, namely the one consisting the vectors v such that $\langle v,x\rangle \le \gamma_C(x)$ for all x. But in view of the definition of γ_C (in 3.50) this set is C°. Thus $\gamma_C = \sigma_{C^\circ}$, and since $C^{\circ\circ} = C$ also by symmetry $\gamma_{C^\circ} = \sigma_C$. These functions are conjugate to δ_{C° and δ_C respectively by 11.4(a). □

11.20 Exercise (dual properties of polar sets)**.** *Let* C *be a closed, convex subset of* $\mathbb{R}^n$ *containing the origin, and let* C° *be its polar as defined in 11.19.*

(a) C *is bounded if and only if* $0 \in \operatorname{int} C^\circ$*; likewise,* C° *is bounded if and only if* $0 \in \operatorname{int} C$.

(b) C *is polyhedral if and only if* C° *is polyhedral.*

(c) $C^\infty = (\operatorname{pos} C^\circ)^*$ *and* $(C^\circ)^\infty = (\operatorname{pos} C)^*$.

Guide. In (a) and (b), rely on the gauge interpretation of polarity in 11.19; apply 11.8(c) and 11.14. Argue the second equation in (c) from the intersection rule in 3.9 and the definition of C° as an intersection of half-spaces. Obtain the other equation in (c) then by symmetry. □

Polars of convex sets other than cones are employed most notably in the study of norms. Any closed, bounded, convex set $B \subset \mathbb{R}^n$ that's *symmetric* ($-B = B$) with nonempty interior (and hence has the origin in this interior) corresponds to a certain norm $\|\cdot\|$, given by its gauge γ_B, cf. 3.50. The polar set B° is likewise a closed, bounded, convex set that's symmetric with nonempty interior. Its gauge γ_{B° gives the norm $\|\cdot\|^\circ$ *polar* to $\|\cdot\|$. Of particular note is the famous rule for polarity in the family of l_p norms in 2.17, namely

$$\begin{aligned} &\|\cdot\|_p^\circ = \|\cdot\|_q \text{ when } 1 < p < \infty,\ 1 < q < \infty,\ p^{-1} + q^{-1} = 1, \\ &\|\cdot\|_1^\circ = \|\cdot\|_\infty, \quad \|\cdot\|_\infty^\circ = \|\cdot\|_1. \end{aligned} \tag{11(12)}$$

This can be derived from the next result, which furnishes additional examples of conjugate convex functions. The argument will be sketched after the proof.

11.21 Proposition (conjugate composite functions from polar gauges)**.** *Consider the gauge γ_C of a closed, convex set $C \subset I\!R^n$ with $0 \in C$ and any lsc, convex function $\theta : I\!R \to \overline{I\!R}$ with $\theta(-r) = \theta(r)$. Under the convention $\theta(\infty) = \infty$ one has the conjugacy relation*

$$\theta\big(\gamma_C(x)\big) \;\underset{*}{\longleftrightarrow}\; \theta^*\big(\gamma_{C^\circ}(v)\big).$$

In particular, for any norm $\|\cdot\|$ and its polar norm $\|\cdot\|^\circ$, one has

$$\theta\big(\|x\|\big) \;\underset{*}{\longleftrightarrow}\; \theta^*\big(\|v\|^\circ\big).$$

Proof. Let $f(x) = \theta\big(\gamma_C(x)\big)$. The function θ has to be nondecreasing on $I\!R_+$ since for any $r > 0$ in $\operatorname{dom}\theta$ we have $\theta(-r) = \theta(r) < \infty$ and consequently

$$\theta\big((1-\tau)(-r) + \tau r\big) \;\le\; (1-\tau)\theta(-r) + \tau\theta(r) = \theta(r) \;\text{ for }\; 0 < \tau < 1,$$

so that $\theta(r') \le \theta(r)$ for all $r' \in (-r, r)$. This monotonicity ensures that f is convex and enables us to write $f(x) = \inf\big\{\theta(\lambda) \,\big|\, \lambda \ge \gamma_C(x)\big\}$. In calculating the conjugate we then have

$$\begin{aligned}
f^*(v) \;&=\; \sup\Big\{\langle v, x\rangle - \theta(\lambda) \,\Big|\, (x,\lambda) \in \operatorname{epi}\gamma_C\Big\} \\
&=\; \sup_{\substack{\lambda\ge 0\\ \lambda\in\operatorname{dom}\theta}} \;\sup\Big\{\langle v, x\rangle - \theta(\lambda) \,\Big|\, x \in \operatorname{lev}_{\le\lambda}\gamma_C\Big\} \\
&=\; \sup_{\substack{\lambda\ge 0\\ \lambda\in\operatorname{dom}\theta}} \begin{Bmatrix} \lambda\sigma_C(v) - \theta(\lambda) & \text{for } \lambda > 0 \\ \delta_{C^{\infty *}}(v) & \text{for } \lambda = 0 \end{Bmatrix} \\
&=\; \sup_{\substack{\lambda\ge 0\\ \lambda\in\operatorname{dom}\theta}} \begin{Bmatrix} \lambda\gamma_{C^\circ}(v) - \theta(\lambda) & \text{for } \lambda > 0 \\ \delta_{\operatorname{cl}(\operatorname{dom}\gamma_{C^\circ})}(v) & \text{for } \lambda = 0 \end{Bmatrix} \;=\; \theta^*\big(\gamma_{C^\circ}(v)\big),
\end{aligned}$$

relying here on $\operatorname{lev}_{\le 0}\gamma_C = C^\infty$ and $C^{\infty *} = \operatorname{cl}(\operatorname{dom}\sigma_C)$; cf. the end of 8.24. $\square$

An illustration of the possibilities in Proposition 11.21 is furnished by the case of composition with the dual functions

$$\theta_p(r) = \frac{1}{p}|r|^p \;\underset{*}{\longleftrightarrow}\; \theta_q^*(s) = \frac{1}{q}|s|^q \qquad\text{when }\; 1 < p < \infty,\; 1 < q < \infty,\; p^{-1} + q^{-1} = 1. \tag{11(13)}$$

This one-dimensional conjugacy can be used to deduce from Proposition 11.21 the polarity rule for the l_p-norms in 11(12). The argument proceeds as follows for $p \in (1, \infty)$. The function $f(x) := \theta_p(x_1) + \cdots + \theta_p(x_n)$ is convex, so the set $I\!B_p = \big\{x \,\big|\, f(x) \le p^{-1}\big\}$ is convex; also, it's closed and symmetric about 0. The function $h := \|\cdot\|_p = (pf)^{1/p}$ is nonnegative, lsc and positively homogeneous with $\mathrm{lev}_{\leq 1} h = I\!B_p$. This implies that $\|\cdot\|_p = \gamma_{\boldsymbol{B}_p}$ and also that $\|\cdot\|_p$ is convex (hence truly is a norm); furthermore $f = \theta_p \circ \gamma_{\boldsymbol{B}_p}$. In parallel, the function $g(v) := \theta_q(v_1) + \cdots \theta_q(v_n)$ agrees with $g = \theta_q \circ \gamma_{\boldsymbol{B}_q}$, with $\|\cdot\|_q = \gamma_{\boldsymbol{B}_q}$. But f and g are conjugate to each other, as seen directly through 11(13). It follows then from Proposition 11.21 that $\|\cdot\|_p$ and $\|\cdot\|_q$ must be polar to each other. (For $p=1$ and $p=\infty$ the polarity in 11(12) can be deduced more simply from 11.19 and the fact that $\|\cdot\|_1$ is the support function of $I\!B_\infty = [-1,1]^n$.)

F. Dual Operations

With a wealth of examples of conjugate convex functions now in hand, we turn to the question of how to dualize other functions generated from these by various operations. The effects of some elementary operations have already been compiled in 11(3), but we now take up the topic in earnest.

11.22 Proposition (conjugation in product spaces). *For proper functions f_i on $I\!R^{n_i}$, the function conjugate to $f(x_1,\ldots,x_m) = f_1(x_1) + \cdots + f_m(x_m)$ is $f^*(v_1,\ldots,v_m) = f_1^*(v_1) + \cdots + f_m^*(v_m)$.*

Proof. This is elementary from the definition of the transform. □

11.23 Theorem (dual operations).

(a) (addition/epi-addition). *For proper functions f_i, if $f = f_1 \mathbin{\#} f_2$, then $f^* = f_1^* + f_2^*$. Dually, if $f = f_1 + f_2$ for proper, lsc, convex functions f_i such that* $\mathrm{dom}\, f_1$ *meets* $\mathrm{dom}\, f_2$, *then $f^* = \mathrm{cl}\big(f_1^* \mathbin{\#} f_2^*\big)$. Here the closure operation is superfluous when $0 \in \mathrm{int}(\mathrm{dom}\, f_1 - \mathrm{dom}\, f_2)$, as is true in particular when* $\mathrm{dom}\, f_1$ *meets* $\mathrm{int}(\mathrm{dom}\, f_2)$ *or when* $\mathrm{dom}\, f_2$ *meets* $\mathrm{int}(\mathrm{dom}\, f_1)$.

(b) (composition/epi-composition). *If $g = Af$ for $f : I\!R^n \to \overline{I\!R}$ and $A \in I\!R^{m\times n}$, where $(Af)(u) := \inf\big\{f(x) \,\big|\, Ax = u\big\}$, then $g^* = f^*A^*$, where $(f^*A^*)(y) := f^*(A^*y)$ (with A^* the transpose of A). Dually, if $f = gA$ for a proper, lsc, convex function $g : I\!R^m \to \overline{I\!R}$ such that the subspace* $\mathrm{rge}\, A$ *meets* $\mathrm{dom}\, g$, *then $f^* = \mathrm{cl}(A^*g^*)$. Here the closure operation is superfluous when $0 \in \mathrm{int}(\mathrm{dom}\, g - \mathrm{rge}\, A)$, as is true in particular when* $\mathrm{rge}\, A$ *meets* $\mathrm{int}(\mathrm{dom}\, g)$.

(c) (restriction/inf-projection). *If $p(u) = \inf_x f(x,u)$ for a proper function $f : I\!R^n \times I\!R^m \to \overline{I\!R}$, then $p^*(y) = f^*(0,y)$. Dually, if f is also convex and lsc, then $\varphi(x) = f(x,\bar{u})$ for some $\bar{u} \in U := \big\{u \,\big|\, \exists\, x,\ f(x,u) < \infty\big\}$, one has $\varphi^* = \mathrm{cl}\, q$ for the function $q(v) = \inf_y\big\{f^*(v,y) - \langle y, \bar{u}\rangle\big\}$. Here the closure operation is superfluous when actually $\bar{u} \in \mathrm{int}\, U$.*

(d) (pointwise sup/inf). *For a family of functions f_i, if $f = \inf_{i\in I} f_i$, then $f^* = \sup_{i\in I} f_i^*$. Dually, if $f = \sup_{i\in I} f_i$ with f_i proper, lsc and convex, and if f is proper, then $f^* = \mathrm{cl\ con}\big(\inf_{i\in I} f_i^*\big)$.*

Proof. The first relation in (a) falls out of the definition of f^* and the formula for $f_1 \mathbin{\#} f_2$ in 1(12). It implies that $(f_1^* \mathbin{\#} f_2^*)^* = f_1^{**} + f_2^{**}$, the latter being the same as $f_1 + f_2$ when each f_i is proper, lsc and convex. When that is true and

$\operatorname{dom} f_1 \cap \operatorname{dom} f_2 \neq \emptyset$, so that $f = f_1 + f_2$ is proper, we get $f^* = (f_1^* \# f_2^*)^{**} = \operatorname{cl}\operatorname{con}(f_1^* \# f_2^*)$ by 11.1. The convex hull operation is superfluous because the convexity of f_i^* implies that of $f_1^* \# f_2^*$ (cf. 2.24). The closure operation can be omitted when $f_1^* \# f_2^*$ is lsc, which holds when the set $\operatorname{epi} f_1^* + \operatorname{epi} f_2^*$ is closed (cf. 1.28), a property guaranteed by the absence of any nonzero $(v, \beta) \in (\operatorname{epi} f_1^*)^\infty$ with $(-v, -\beta) \in (\operatorname{epi} f_2^*)^\infty$ (cf. 3.12).

Because $(\operatorname{epi} f_i^*)^\infty$ is the epigraph of $f_i^{*\infty}$, which we have identified in 11.5 with the support function of $D_i = \operatorname{dom} f_i$, this condition translates to the nonexistence of $v \neq 0$ such that $\sigma_{D_1}(v) \leq -\sigma_{D_2}(-v)$. But this is equivalent by 8.29(b) to having $0 \in \operatorname{int}(D_1 - D_2)$. Obviously $\operatorname{int}(D_1 - D_2)$ includes the sets $D_1 - (\operatorname{int} D_2)$ and $(\operatorname{int} D_1) - D_2$, since these are open.

For (b) the same pattern works. It's easily seen from the definitions that $(Af)^* = f^* A^*$. For the same reason, $(A^* g^*)^* = g^{**} A^{**} = gA$ when g is proper, lsc and convex. When $\operatorname{rge} A$ meets $\operatorname{dom} g$, so that gA is proper, we obtain from 11.1 that $(gA)^* = (A^* g^*)^{**} = \operatorname{cl}\operatorname{con} A^* g^*$. The convex hull operation can be omitted because the convexity of $A^* g^*$ accompanies that of g^* by 2.22(b). The closure operation can be omitted when $A^* g^*$ is lsc, which holds when the set $L(\operatorname{epi} g^*)$ is closed for the linear mapping $L(y, \alpha) = (A^* y, \alpha)$; cf. 1.31. This is implied by the absence of any nonzero (y, α) in $L^{-1}(0,0) \cap (\operatorname{epi} g^*)^\infty$, i.e., the absence of any $y \neq 0$ such that $A^* y = 0$, $g^{*\infty}(y) \leq 0$. Identifying $g^{*\infty}$ with the support function of $\operatorname{dom} g$ through 11.5, and identifying the indicator of the null space $\{ y \mid A^* y = 0 \}$ with the support function of the range space $\operatorname{rge} A$, we translate this condition into the absence of any $y \neq 0$ such that $\sigma_{\operatorname{dom} g}(y) \leq -\sigma_{\operatorname{rge} A}(-y)$. By 8.29(b), this means $0 \in \operatorname{int}(\operatorname{dom} g - \operatorname{rge} A)$. Here $\operatorname{int}(\operatorname{dom} g - \operatorname{rge} A)$ includes the open set $\operatorname{int}(\operatorname{dom} g) - \operatorname{rge} A$.

Likewise in (c), the definitions of p and p^* give

$$\begin{aligned} p^*(y) &= \sup_u \{ \langle y, u \rangle - \inf_x f(x, u) \} \\ &= \sup_{x,u} \{ \langle (0, y), (x, u) \rangle - f(x, u) \} = f^*(0, y). \end{aligned}$$

For parallel reasons, $q^*(x) = f^{**}(x, \bar{u})$. When f is proper, lsc and convex, we have $f^{**} = f$, so $q^* = \varphi$. But the convexity of f^* implies that of q by 2.22(a). Hence as long as $\bar{u} \in U$, so that φ is proper, we have $\varphi^* = \operatorname{cl} q$ by 11.1. To omit the closure operation as well, we can look to cases where q is known to be lsc. One such case, furnished by 3.31, is associated with having $f^{*\infty}(0, y) - \langle y, \bar{u} \rangle > 0$ when $y \neq 0$. But if f is proper and convex, $f^{*\infty}$ is the support function of $\operatorname{dom} f$ by 11.5, and $f^{*\infty}(0, \cdot)$ is then the support function of the image U of $\operatorname{dom} f$ under the projection $(x, u) \mapsto u$. The condition that $f^{*\infty}(0, y) - \langle y, \bar{u} \rangle > 0$ when $y \neq 0$ translates therefore to the condition that $\sigma_U(y) > 0$ when $y \neq 0$, which by 8.29(a) is equivalent to having $0 \in \operatorname{int} U$.

The first relation in (d) is immediate from the definition of f^*. It implies that $(\inf_{i \in I} f_i^*)^* = \sup_{i \in I} f_i^{**}$. When $f = \sup_{i \in I} f_i$ with f_i proper, lsc and convex, so that $f_i = f_i^{**}$ by 11.1, and f is proper (in addition to being convex by 2.9), we obtain from 11.1 that $f^* = (\inf_{i \in I} f_i^*)^{**} = \operatorname{cl}\operatorname{con}(\inf_{i \in I} f_i^*)$. □

In part (a) of Theorem 11.23, the condition $0 \in \operatorname{int}(\operatorname{dom} g - \operatorname{rge} A)$ is equivalent to the nonexistence of a 'separating hyperplane' for C_1 and C_2 (see 2.39). Likewise in part (b), the condition $0 \in \operatorname{int}(\operatorname{dom} g - \operatorname{rge} A)$ means that $\operatorname{dom} g$ can't be separated from $\operatorname{rge} A$, even improperly.

11.24 Corollary (rules for support functions).

(a) *If* $D = \lambda C$ *with* $C \neq \emptyset$ *and* $\lambda > 0$, *then* $\sigma_D = \lambda \sigma_C$.

(b) *If* $C = C_1 + C_2$ *with* $C_i \neq \emptyset$, *then* $\sigma_C = \sigma_{C_1} + \sigma_{C_2}$.

(c) *If* $D = \{Ax \mid x \in C\}$ *with* $A \in \mathbb{R}^{m \times n}$, *then* $\sigma_D(y) = \sigma_C(A^*y)$.

(d) *If* $C = \{x \mid Ax \in D\}$ *for a closed, convex set* $D \subset \mathbb{R}^m$, *and if* $C \neq \emptyset$, *then* $\sigma_C = \operatorname{cl} A^* \sigma_D$, *where* $(A^*\sigma_D)(v) := \inf\{\sigma_D(y) \mid A^*y = v\}$. *Here the closure operation is superfluous when* $0 \in \operatorname{int}(D - \operatorname{rge} A)$, *as is true in particular when the subspace* $\operatorname{rge} A$ *meets* $\operatorname{int} D$.

(e) *If* $C = C_1 \cap C_2 \neq \emptyset$ *with each set* C_i *convex and closed, then* $\sigma_C = \operatorname{cl}(\sigma_{C_1} \# \sigma_{C_2})$. *Here the closure operation is superfluous when* $0 \in \operatorname{int}(C_1 - C_2)$, *as is true in particular when* C_1 *meets* $\operatorname{int} C_2$ *or* C_2 *meets* $\operatorname{int} C_1$.

(f) *If* $C = \bigcup_{i \in I} C_i$, *then* $\sigma_C = \sup_{i \in I} \sigma_{C_i}$.

Proof. These six rules correspond to the cases where (a) $\delta_D = \lambda \star \delta_C$, (b) $\delta_C = \delta_{C_1} \# \delta_{C_2}$, (c) $\delta_D = A\delta_C$, (d) $\delta_C = \delta_D A$, (e) $\delta_C = \delta_{C_1} + \delta_{C_2}$, and (f) $\delta_C = \inf_{i \in I} \delta_{C_i}$. All except (a) are covered by Theorem 11.23 through 11.4(a), while (a) simply comes from 11(3)—but is best listed here. □

11.25 Corollary (rules for polar cones).

(a) *If* $K = K_1 + K_2$ *for cones* K_i, *then* $K^* = K_1^* \cap K_2^*$. *Likewise, if* $K = \bigcup_{i \in I} K_i$ *one has* $K^* = \bigcap_{i \in I} K_i^*$.

(b) *If* $K = K_1 \cap K_2$ *for closed, convex cones* K_i, *then* $K^* = \operatorname{cl}(K_1^* + K_2^*)$. *The closure operation is superfluous when* $0 \in \operatorname{int}(K_1 - K_2)$.

(c) *If* $H = \{Ax \mid x \in K\}$ *for* $A \in \mathbb{R}^{m \times n}$ *and a cone* $K \subset \mathbb{R}^n$, *then* $H^* = \{y \mid A^*y \in K^*\}$.

(d) *If* $K = \{x \mid Ax \in H\}$ *for* $A \in \mathbb{R}^{m \times n}$ *and a closed, convex cone* $H \subset \mathbb{R}^m$, *then* $K^* = \operatorname{cl}\{A^*y \mid y \in H^*\}$. *The closure operation is superfluous when* $0 \in \operatorname{int}(H - \operatorname{rge} A)$.

Proof. In (a) we apply 11.23(a) to $\delta_K = \delta_{K_1} \# \delta_{K_2}$, or 11.23(d) to $\delta_K = \inf_{i \in I} \delta_{K_i}$, whereas in (b) we apply 11.23(a) to $\delta_K = \delta_{K_1} + \delta_{K_2}$, each time making the observation in 11.4(b). In (c) and (d) it's the same story in applying 11.23(b) with $\delta_H = A\delta_K$ in the first case and $\delta_K = \delta_H A$ in the second. □

11.26 Example (distance functions, Moreau envelopes and proximal hulls).

(a) *For any nonempty, closed, convex set* $C \subset \mathbb{R}^n$, *the functions* d_C *and* $\sigma_C + \delta_{\mathbb{B}}$ *are conjugate to each other.*

(b) *For any proper, lsc, convex function* $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ *and any* $\lambda > 0$, *the functions* $e_\lambda f$ *and* $f^* + \frac{\lambda}{2}|\cdot|^2$ *are conjugate to each other. This entails*

$$e_\lambda f(x) + e_{\lambda^{-1}} f^*(\lambda^{-1}x) = \frac{1}{2\lambda}|x|^2 \quad \text{for all } x \in \mathbb{R}^n,\ \lambda > 0.$$

(c) *For any* $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, *not necessarily convex, the Moreau envelope* $e_\lambda f$ *and proximal hull* $h_\lambda f$ *for each* $\lambda > 0$ *are expressed by*

$$\left.\begin{aligned} e_\lambda f(x) &= \lambda^{-1} j(x) - (f + \lambda^{-1} j)^*(\lambda^{-1} x) \\ h_\lambda f(x) &= (f + \lambda^{-1} j)^{**}(x) - \lambda^{-1} j(x) \end{aligned}\right\} \text{ for } j = \tfrac{1}{2}|\cdot|^2.$$

Here $(f + \lambda^{-1} j)^{**}$ *can be replaced by* $\operatorname{con}(f + \lambda^{-1} j)$ *when* f *is lsc, proper and prox-bounded with threshold* $\lambda_f > \lambda$. *Then* $\operatorname{dom} h_\lambda f = \operatorname{con}(\operatorname{dom} f)$, *and on the interior of this convex set the function* $h_\lambda f$ *must be lower-*$\mathcal{C}^2$.

(d) *A proper function* $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ *is* λ*-proximal, in the sense that* $h_\lambda f = f$, *if and only if* f *is lsc and* $f + \frac{1}{2\lambda}|\cdot|^2$ *is convex.*

Detail. In (a) we have $d_C = \delta_C \,\#\, |\cdot|$ by 1(13), where the Euclidean norm $|\cdot|$ is the support function $\sigma_{\boldsymbol{B}}$, as recorded in 8.27, and $\sigma_{\boldsymbol{B}}^* = \delta_{\boldsymbol{B}}$ by 11.4(a). Then $d_C^* = \delta_C^* + \sigma_{\boldsymbol{B}}^*$ by 11.23(b), with $\delta_C^* = \sigma_C$ by 11.4(a). Because d_C is finite, hence continuous (cf. 2.36), we have $d_C^{**} = d_C$ (by 11.1).

In (b) the situation is similar. In terms of $j(x) = \frac{1}{2}|x|^2$ we have $e_\lambda f = f \,\#\, \lambda^{-1} j$ by 1(13), with $e_\lambda f$ finite and convex by 2.25. Here $\lambda^{-1} j$ is the same as $\lambda \star j$ and is conjugate to λj by 11.11 and the rules in 11(3). The conjugate function $(e_\lambda f)^*$ is calculated then from 11.23(a) to be $f^* + \lambda j$. But to say that $e_\lambda f$ is conjugate to $f^* + \lambda j$ is to say that for all x one has

$$\begin{aligned} e_\lambda f(x) &= \sup\nolimits_w \Big\{ \langle w, x \rangle - f^*(w) - \frac{\lambda}{2}|w|^2 \Big\} \\ &= \frac{1}{2\lambda}|x|^2 - \inf\nolimits_w \Big\{ f^*(w) + \frac{\lambda}{2}|w - \lambda^{-1} x|^2 \Big\} = \frac{1}{2\lambda}|x|^2 - e_{\lambda^{-1}} f^*(\lambda^{-1} x). \end{aligned}$$

This gives the identity claimed in (b).

The same calculation produces the first identity in (c), and the second then follows from the formula for $h_\lambda f$ in 1.44. Justification for replacing $(f + \lambda^{-1} j)^{**}$ by $\operatorname{con}(f + \lambda^{-1} j)$ comes from 3.28 and 3.47; $f + \lambda^{-1} j$ is coercive when $\lambda \in (0, \lambda_f)$. The domain assertion is then obvious. The claim about $h_\lambda f$ being lower-$\mathcal{C}^2$ on the interior is supported by 10.33. To get (d), we merely invoke the rule from (c) that $h_\lambda f = f$ if and only if $(f + \lambda^{-1} j)^{**} = (f + \lambda^{-1} j)$. □

11.27 Exercise (proximal mappings and projections as gradient mappings)**.**

(a) *For any proper, lsc, convex function* $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ *and any* $\lambda > 0$ *the proximal mapping* $P_\lambda f$ *is* ∇g *for* $g = \lambda \star \big(e_{\lambda^{-1}} f^* \big)$.

(b) *For any nonempty, closed, convex set* $C \subset \mathbb{R}^n$, *the projection mapping* P_C *is* ∇g *for* $g = e_1 \sigma_C = \sigma_C \,\#\, \frac{1}{2}|\cdot|^2$.

Guide. Derive the first expression from the formula for $\nabla e_\lambda f$ in 2.26 using the identity in 11.26(b). Then derive the second expression by specializing to $f = \delta_C$ and $\lambda = 1$; cf. 11.4. □

11.28 Example (piecewise linear-quadratic envelopes)**.**

(a) *For* $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ *proper, convex and piecewise linear-quadratic and for any* $\lambda > 0$, *the convex function* $e_\lambda f$ *is piecewise linear-quadratic.*

(b) *For $C \subset \mathbb{R}^n$ nonempty and polyhedral, the convex function d_C^2 is piecewise linear-quadratic.*

Detail. The assertion in (a) is justified by the conjugacy rules in 11.14(a) and 11.26(b). The one in (b) specializes this to $f = \delta_C$; then $e_\lambda f = \frac{1}{2\lambda} d_C^2$. □

The use of several dualizing rules during the course of a calculation is illustrated by the next example, which concerns adjoints of sublinear mappings as defined in 8(27) and 8(28). Relations are obtained between the outer and inner norms for such mappings that were introduced in 9(4) and 9(5).

11.29 Example (norm duality for sublinear mappings)**.** *The outer norm $|H|^+$ and inner norm $|H|^-$ of a sublinear, osc mapping $H : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ are related to those of its upper adjoint H^{*+} and lower adjoint H^{*-} by*

$$|H|^+ = |H^{*+}|^- = |H^{*-}|^-, \qquad |H|^- = |H^{*+}|^+ = |H^{*-}|^+.$$

In addition, one has $d\big(0, H(w)\big) = \sigma_{H^{*+}(\boldsymbol{B})}(w)$ and $\sigma_{H(\boldsymbol{B})}(y) = d\big(0, H^{*-}(y)\big)$.

As a special case, if a mapping $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is graphically regular at $\bar{x}$ for $\bar{u}$, one has

$$\big|D^*S(\bar{x}\,|\,\bar{u})\big|^+ = \big|DS(\bar{x}\,|\,\bar{u})\big|^-, \qquad \big|D^*S(\bar{x}\,|\,\bar{u})^{-1}\big|^+ = \big|DS(\bar{x}\,|\,\bar{u})^{-1}\big|^-.$$

Detail. This can be derived by using the support function rules in 11.24 along with the rules of cone polarity. From the definition of $|H|^+$ in 9(4) and the description of the Euclidean norm in 8.27 we have

$$|H|^+ = \sup_{z \in H(\boldsymbol{B})} |z| \;=\; \sup_{z \in H(\boldsymbol{B}),\, y \in \boldsymbol{B}} \langle y, z\rangle \;=\; \sup_{y \in \boldsymbol{B}} \sigma_{H(\boldsymbol{B})}(y).$$

To prove that this equals $|H^{*+}|^-$, hence also $|H^{*-}|^-$ (inasmuch as $H^{*-}(y) = -H^{*+}(-y)$), it's enough now to demonstrate that $\sigma_{H(\boldsymbol{B})}(y) = d\big(0, H^{*-}(y)\big)$. In terms of the projection $A : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ the set $H(\mathbb{B})$ has the representation AC for $C = C_1 \cap C_2$ with $C_1 = \mathbb{B} \times \mathbb{R}^m$ and $C_2 = \operatorname{gph} H$. From 11.24(c) we get $\sigma_{H(\boldsymbol{B})}(y) = \sigma_C(A^*y) = \sigma_C\big((0,y)\big)$. On the other hand we have $C_2 \cap \operatorname{int} C_1 \neq \emptyset$, so $\sigma_C\big((0,y)\big) = \min_{(w,u)}\big\{\sigma_{C_1}\big((0,y) - (w,u)\big) + \sigma_{C_2}\big((w,u)\big)\big\}$ by 11.24(e). Next we note that $\sigma_{C_1}\big((w,u)\big) = |w| + \delta_{\{0\}}(u)$ (cf. 8.27), whereas $\sigma_{C_2} = \delta_{(\operatorname{gph} H)^*}$ by 11.4(b), so that $\sigma_{C_2}\big((w,u)\big) = \delta_{\operatorname{gph} H^{*-}}\big((-w,u)\big)$ by the definition of H^{*-} in 8(28). Therefore,

$$\begin{aligned}\sigma_{H(\boldsymbol{B})}(y) \;&=\; \min_{(w,u)}\Big\{|0 - w| + \delta_{\{0\}}(y-u) + \delta_{\operatorname{gph} H^{*-}}\big((-w,u)\big)\Big\}\\ &=\; \min_{-w \in H^{*-}(y)} |w| \;=\; d\big(0, H^{*-}(y)\big).\end{aligned}$$

Thus, $|H|^+ = |H^{*+}|^-$ is confirmed. To get the remaining formulas, we merely have to apply the ones already obtained to $G = H^{*+}$, since $G^{*-} = H$.

The application at the end is based on having, in the presence of graphical regularity, $D^*S(\bar{x}\,|\,\bar{u}) = DS(\bar{x}\,|\,\bar{u})^{*-}$ and likewise $D^*S^{-1}(\bar{x}\,|\,\bar{u}) = DS^{-1}(\bar{x}\,|\,\bar{u})^{*-}$;

cf. 8.40. It's elementary that one always has $|D^*S^{-1}(\bar{x}\,|\,\bar{u})|^+ = |D^*S(\bar{x}\,|\,\bar{u})^{-1}|^+$ and $|DS^{-1}(\bar{x}\,|\,\bar{u})|^- = |DS(\bar{x}\,|\,\bar{u})^{-1}|^-$. □

11.30 Exercise (uniform boundedness of sublinear mappings). *If a family $\mathcal{H}$ of osc, sublinear mappings $H : I\!R^n \rightrightarrows I\!R^m$ has $\sup_{H\in\mathcal{H}} d\big(0, H(w)\big) < \infty$ for each $w \in I\!R^n$, it must actually have $\sup_{H\in\mathcal{H}} |H|^- < \infty$.*

Guide. Make use of 11.29, arguing that $h(w) = \sup_{H\in\mathcal{H}} d\big(0, H(w)\big)$ is the support function of the set $\bigcup_{H\in\mathcal{H}} H^{*+}(I\!B)$. □

11.31 Exercise (duality in calculating adjoints of sublinear mappings).

(a) *For sublinear mappings $H : I\!R^n \rightrightarrows I\!R^m$ and $G : I\!R^m \rightrightarrows I\!R^p$ with* $\operatorname{rint}(\operatorname{rge} H) \cap \operatorname{rint}(\operatorname{dom} G) \neq \emptyset$, *one has*

$$(G\circ H)^{*+} = H^{*+}\circ G^{*+}, \qquad (G\circ H)^{*-} = H^{*-}\circ G^{*-}.$$

(b) *For sublinear mappings $H : I\!R^n \rightrightarrows I\!R^m$ and $G : I\!R^n \rightrightarrows I\!R^m$ with* $\operatorname{rint}(\operatorname{dom} H) \cap \operatorname{rint}(\operatorname{dom} G) \neq \emptyset$, *one has*

$$(H + G)^{*+} = H^{*+} + G^{*+}, \qquad (H + G)^{*-} = H^{*-} + G^{*-}.$$

(c) *For a sublinear mapping $H : I\!R^n \rightrightarrows I\!R^m$ and arbitrary $\lambda > 0$, one has*

$$(\lambda H)^{*+} = \lambda H^{*+}, \qquad (\lambda H)^{*-} = \lambda H^{*-}.$$

Guide. In (a), $\operatorname{gph}(G\circ H) = L(K)$ for $K = \big[\operatorname{gph} H \times I\!R^p\big] \cap \big[I\!R^n \times \operatorname{gph} G\big]$ and L the projection of $I\!R^n \times I\!R^m \times I\!R^p$ onto $I\!R^n \times I\!R^p$. Apply the rules in 11.25(b)(c), relating the polars $\operatorname{gph} H$, $\operatorname{gph} G$ and $\operatorname{gph}(G\circ H)$, to the graphs of the adjoint mappings by way of definitions 8(27) and 8(28).

In (b) use the representation $\operatorname{gph}(H + G) = L_2\big(L_1^{-1}(K)\big)$ for the cone $K = (\operatorname{gph} H) \times (\operatorname{gph} G)$, the linear mapping $L_1 : (x, y, z) \mapsto (x, y, x, z)$ and the linear mapping $L_2 : (x, y, z) \mapsto (x, y + z)$. Apply the rules in 11.25(c)(d). Get the elementary fact in (c) straight from the definitions of H^{*+} and H^{*-}. □

For the pairs of operations in Theorem 11.23 to be fully dual to each other, closures had to be taken, but a readily verifiable condition was provided under which this was superfluous. Another common case where it can be omitted comes up when the functions are piecewise linear-quadratic.

11.32 Proposition (operations on piecewise linear-quadratic functions).

(a) *If $f = f_1 \# f_2$ with f_i proper, convex and piecewise linear-quadratic, then f is proper, convex and piecewise linear-quadratic, unless f is improper with $\operatorname{dom} f$ a polyhedral set on which $f \equiv -\infty$. Either way, f is lsc.*

(b) *If $g(u) = (Af)(u) = \inf_x\big\{f(x)\,\big|\,Ax = u\big\}$ for $A \in I\!R^{m\times n}$ and f proper, convex and piecewise linear-quadratic on $I\!R^n$, then g is proper, convex and piecewise linear-quadratic on $I\!R^m$, unless g is improper with $\operatorname{dom} g$ a polyhedral set on which $g \equiv -\infty$. Either way, g is lsc.*

(c) *If $p(u) = \inf_x f(x, u)$ for a proper, convex, piecewise linear-quadratic function $f : I\!R^n \times I\!R^m \to \overline{I\!R}$, then p is proper, convex and piecewise linear-*

quadratic, unless p *is improper with* $\operatorname{dom} p$ *a polyhedral set on which* $p \equiv -\infty$. *Either way,* p *is lsc.*

Proof. Let's deal with (c) first; it will be leveraged into the other assertions. Because f is convex, its inf-projection p is convex; cf. 2.22(a). The set $\operatorname{dom} p$ is the image of $\operatorname{dom} f$ under the projection mapping $(x, u) \mapsto u$, and $\operatorname{dom} f$ is polyhedral. Hence $\operatorname{dom} p$ is itself polyhedral and in particular closed; cf. 3.55(a). The function conjugate to p is known from 11.23(c) to be $f^*(0, \cdot)$, which is not only convex but piecewise linear-quadratic by 10.22(c), due to the fact that f^* inherits being piecewise linear-quadratic from f by 11.14(b). As long as $f^*(0, \cdot)$ is proper, which is equivalent by 11.1 to p being proper, it follows further by 11.14(b) that p^{**}, as the function conjugate to $f^*(0, \cdot)$, is piecewise linear-quadratic too. The claim in (c) can be established therefore in showing that $p^{**} = p$ unless p is improper with no values other than $\pm\infty$.

Suppose p is finite at $\bar{u}$. The function $\varphi := f(\cdot, \bar{u})$, whose infimum over $\mathbb{R}^n$ is $p(\bar{u})$, is convex and piecewise linear-quadratic, hence its minimum is attained at some $\bar{x}$; cf. 11.16. Then $0 \in \partial\varphi(\bar{x})$. But $\partial\varphi(\bar{x}) = \{v \mid \exists\, y,\ (v, y) \in \partial f(\bar{x}, \bar{u})\}$ by 10.22(c). Hence there exists $\bar{y}$ with $(0, \bar{y}) \in \partial f(\bar{x}, \bar{u})$. Through convexity this subgradient condition can be written as

$$f(x, u) \ \geq\ f(\bar{x}, \bar{u}) + \big\langle (0, \bar{y}),\, (x, u) - (\bar{x}, \bar{u}) \big\rangle \ \text{ for all } \ (x, u)$$

(see 8.12), which gives us $p(u) \geq p(\bar{u}) + \langle \bar{y},\, u - \bar{u} \rangle$ for all u, implying that p is proper on $\mathbb{R}^m$ and lsc at $\bar{u}$. Thus, unless $p \equiv -\infty$ on $\operatorname{dom} p$, p is a proper, convex function which is lsc at every point of $\operatorname{dom} p$. Since $\operatorname{dom} p$ is closed, we conclude that p is lsc everywhere, so $p^{**} = p$. This proves (c).

We get (b) now as the case of (c) where $Af = p$ for $p(u) = \inf_x \bar{f}(x, u)$ with $\bar{f}(x, u) = f(x) + \delta_M(x, u)$ and $M = \{(x, u) \mid Ax = u\}$. Because δ_M, like f, is convex and piecewise linear-quadratic (the set M being affine, hence polyhedral; cf. 2.10), $\bar{f}$ is piecewise linear-quadratic by 10.22.

Finally, (a) specializes (b) to $f_1 \# f_2 = Ag$ with $g(x_1, x_2) = f_1(x_1) + f_2(x_2)$ on $\mathbb{R}^n \times \mathbb{R}^n$ and A the matrix of the linear mapping $(x_1, x_2) \mapsto x_1 + x_2$. $\square$

11.33 Corollary (conjugate formulas in piecewise linear-quadratic case)**.**

(a) *If* $f = f_1 + f_2$ *with* f_i *proper, convex and piecewise linear-quadratic, and if* $f \not\equiv \infty$, *then* $f^* = f_1^* \# f_2^*$.

(b) *If* $f = gA$ *with* $A \in \mathbb{R}^m \times \mathbb{R}^n$ *and* g *proper, convex and piecewise linear-quadratic, and if* $f \not\equiv \infty$, *then* $f^* = A^* g^*$.

(c) *If* $\varphi(x) = f(x, \bar{u})$ *with* f *proper, convex and piecewise linear-quadratic on* $\mathbb{R}^n \times \mathbb{R}^m$, *and if* $\varphi \not\equiv \infty$, *then* $\varphi^*(v) = \inf_y \{f^*(v, y) - \langle y, \bar{u} \rangle\}$.

Proof. We apply 11.23 in the light of the additional conditions furnished by 11.32 for the dual operations to produce a lower semicontinuous function, so that the closure operation can be omitted. $\square$

G. Duality in Convergence

Switching to another topic, we look next at how the Legendre-Fenchel transform behaves with respect to the epi-convergence studied in Chapter 7.

11.34 Theorem (epi-continuity of the Legendre-Fenchel transform; Wijsman)**.** *If the functions f^ν and f on $I\!R^n$ are proper, lsc, and convex, one has*

$$f^\nu \xrightarrow{\text{e}} f \quad \Longleftrightarrow \quad f^{\nu*} \xrightarrow{\text{e}} f^*.$$

More generally, as long as $\text{e-}\liminf_\nu f^\nu$ *nowhere takes on* $-\infty$ *and a bounded set* B *exists with* $\limsup_\nu [\inf_B f^\nu] < \infty$, *one has*

$$\begin{aligned} \text{e-}\liminf_\nu f^\nu \geq f \quad &\Longleftrightarrow \quad \text{e-}\limsup_\nu f^{\nu*} \leq f^*, \\ \text{e-}\limsup_\nu f^\nu \leq f \quad &\Longleftrightarrow \quad \text{e-}\liminf_\nu f^{\nu*} \geq f^*. \end{aligned}$$

Proof. For any $\lambda > 0$ we have through 7.37 that $f^\nu \xrightarrow{\text{e}} f$ if and only if $e_\lambda f^\nu \xrightarrow{\text{p}} e_\lambda f$. Likewise, $f^{\nu*} \xrightarrow{\text{e}} f$ if and only if $e_\lambda f^{\nu*} \xrightarrow{\text{p}} e_\lambda f^*$. In particular we can take $\lambda = 1$ in these conditions. But from 11.26 we have

$$e_1 f^\nu(x) + e_1 f^{\nu*}(x) = \tfrac{1}{2}|x|^2 = e_1 f(x) + e_1 f^*(x).$$

Thus, the two conditions are equivalent.

The assumption about B in the more general case means the existence of $\beta \in I\!R$ such that $\operatorname{epi} f^\nu \cap (B \times (-\infty, \beta]) \neq \emptyset$ for all ν in some index set in $\mathcal{N}_\infty$. This ensures that no subsequence of $\{f^\nu\}_{\nu \in I\!N}$ can escape epigraphically to the horizon; cf. 7.5. Then every subsequence has an epi-convergent subsequence by 7.6, but on the other hand, the epi-limit of such a sequence can't take on $-\infty$ because of the assumption about $\text{e-}\liminf_\nu f^\nu$. We are therefore in a setting where every subsequence of $\{f^\nu\}_{\nu \in I\!N}$ has a subsequence that epi-converges to some proper, lsc function g, which must of course be convex. Through the cluster description of outer limits in 4.19 as applied to epigraphs, we see that $\text{e-}\liminf_\nu f^\nu \geq f$ if and only if $g \geq f$ for every such function g; likewise in terms of inner limits, we have $\text{e-}\limsup_\nu f^\nu \leq f$ if and only if $g \geq f$ for every such g. It remains only to invoke for these epi-convergent sequences the continuity property established in the main part of the theorem. □

11.35 Corollary (convergence of support functions and polar cones)**.**

(a) *For nonempty, closed, convex sets C^ν and C in $I\!R^n$, one has*

$$C^\nu \to C \quad \Longleftrightarrow \quad \sigma_{C^\nu} \xrightarrow{\text{e}} \sigma_C,$$

and if the sets are bounded this is equivalent to having $\sigma_{C^\nu}(v) \to \sigma_C(v)$ for each v. More generally, as long as $\limsup_\nu d(0, C^\nu) < \infty$ *one has*

$$\begin{aligned} \limsup_\nu C^\nu \subset C \quad &\Longleftrightarrow \quad \text{e-}\limsup_\nu \sigma_{C^\nu} \leq \sigma_C, \\ \liminf_\nu C^\nu \supset C \quad &\Longleftrightarrow \quad \text{e-}\liminf_\nu \sigma_{C^\nu} \geq \sigma_C. \end{aligned}$$

(b) *For closed, convex cones K^ν and K in $I\!R^n$, one has*

$$K^\nu \to K \quad \Longleftrightarrow \quad K^{\nu*} \to K^*,$$

and more generally,

$$\begin{aligned} \limsup_\nu K^\nu \subset K \quad &\Longleftrightarrow \quad \liminf_\nu K^{\nu*} \supset K^*, \\ \liminf_\nu K^\nu \supset K \quad &\Longleftrightarrow \quad \limsup_\nu K^{\nu*} \subset K^*. \end{aligned}$$

Proof. We apply Theorem 11.34 in the setting of 11.4. The case of bounded C^ν, in which the support functions are finite-valued, appeals also to 7.18. □

The Legendre-Fenchel transform isn't just continuous; it's an *isometry* with respect to a suitable choice of metric on the space of proper, lsc, convex functions on $I\!R^n$. We'll establish that by developing isometric properties of the polarity correspondence in 6.24 and 11.4(b) and then appealing to the way that conjugate functions can be associated with polar cones, as in 11.7. In this geometric context, the set metric $d\!l$ in 4(12) will be apt.

11.36 Theorem (cone polarity as an isometry; Walkup-Wets)**.** *For any cones $K_1,\, K_2 \subset I\!R^n$ that are convex, one has*

$$d\!l(K_1, K_2) = d\!l(K_1^*, K_2^*).$$

Proof. It can be assumed that K_1 and K_2 are closed, since $d\!l(K_1, K_2) = d\!l(\operatorname{cl} K_1, \operatorname{cl} K_2)$ with $\operatorname{cl} K_i$ convex and $[\operatorname{cl} K_i]^* = K_i^*$. Then $[K_i^*]^* = K_i$ by 11.4(b). We have $d\!l(K_1, K_2) = d\!l_1(K_1, K_2) = \hat{d\!l}_1(K_1, K_2)$ and $d\!l(K_1^*, K_2^*) = d\!l_1(K_1^*, K_2^*) = \hat{d\!l}_1(K_1^*, K_2^*)$ by 4.44. Thus, we need only demonstrate that $d\!l_1(K_1, K_2) \leq \hat{d\!l}_1(K_1^*, K_2^*)$ and $d\!l_1(K_1^*, K_2^*) \leq \hat{d\!l}_1(K_1, K_2)$; but the latter is just the former as applied to K_1^* and K_2^*, so the former suffices. Because of the way that K_1 and K_2 enter symmetrically in the definition of $d\!l_1(K_1, K_2)$ and $\hat{d\!l}_1(K_1^*, K_2^*)$ in 4(11), our task reduces simply to verifying for $\eta > 0$ that

$$d_{K_1} \leq d_{K_2} + \eta \text{ on } I\!B \quad \Longrightarrow \quad K_1^* \cap I\!B \subset K_2^* + \eta I\!B. \tag{11(14)}$$

The convex functions d_{K_1} and d_{K_2} are positively homogeneous, so the left side of 11(14) corresponds to the inequality $d_{K_1} \leq d_{K_2} + \eta|\cdot|$ holding on $I\!R^n$, or equivalently $d_{K_1}^* \geq (d_{K_2} + \eta|\cdot|)^*$. Here $d_{K_1} = \delta_{K_1} \# |\cdot|$ and $d_{K_2} = \delta_{K_2} \# |\cdot|$ (cf. 1.20), while $|\cdot| = \sigma_{\boldsymbol{B}}$ (cf. 8.27) and $\eta|\cdot| = \sigma_{\eta\boldsymbol{B}}$, so the inequality in 11(14) dualizes to $\delta_{K_1}^* + \sigma_{\boldsymbol{B}}^* \geq [\delta_{K_2}^* + \sigma_{\boldsymbol{B}}^*] \# \sigma_{\eta\boldsymbol{B}}^*$ through the rules in 11.23(a). By 11.4, this is the same as $\delta_{K_1^*} + \delta_{\boldsymbol{B}} \geq [\delta_{K_2^*} + \delta_{\boldsymbol{B}}] \# \delta_{\eta\boldsymbol{B}}$, or in other words $K_1^* \cap I\!B \subset [K_2^* \cap I\!B] + \eta I\!B$, which implies the right side of 11(14). □

Theorem 11.36 can be applied to convex cones in the ray space model of cosmic space, most fruitfully to the cones in $I\!R^{n+2}$ that represent the epigraphs in $I\!R^{n+1}$ of convex functions on $I\!R^n$. Since the cosmic metric $d\!l_{\rm csm}$ for functions on $I\!R^n$, as defined in 7(28), is based on the set distance between such cones, the following result is obtained.

11.37 Corollary (Legendre-Fenchel transform as an isometry)**.** *For functions $f_1,\, f_2 : I\!R^n \to \overline{I\!R}$ that are convex and proper, one has*

$$d\!l_{\rm csm}(f_1,f_2) = d\!l_{\rm csm}(f_1^*,f_2^*).$$

Proof. This comes out of the characterization of conjugate functions in terms of polar cones in 11.7. The cosmic epi-distances correspond to cone distances, and the isometry is apparent then from the one for cones in Theorem 11.36. □

It might be puzzling that this result is based on the cosmic epi-metric $d\!l_{\rm csm}$, which is known from 7.59 to characterize total epi-convergence, whereas the 'homeomorphism' in Theorem 11.34 refers to ordinary epi-convergence, which according to 7.58 is characterized by the ordinary epi-metric $d\!l$. The seeming discord vanishes when one recalls that, for sequences of convex functions, epi-convergence implies total epi-convergence (cf. 7.53). On the space of *convex* functions within $\text{lsc-fcns}_{\not\equiv\infty}(I\!R^n)$, the metrics $d\!l$ and $d\!l_{\rm csm}$ are actually equivalent topologically. Theorem 11.34 could equally well be stated in terms of total epi-convergence, but only $d\!l_{\rm csm}$ produces an isometry.

H. Dual Problems of Optimization

The rest of this chapter will be occupied with the important question of how *optimization problems* can be dualized. It will be shown that any optimization problem of convex type, when provided with a scheme of perturbation that respects convexity, is paired with a certain other optimization problem of convex type, which is provided in turn with a dual scheme of perturbation. The two problems are related to each other in remarkable ways. Even for problems that aren't of convex type, something analogous can be built up, although not as powerfully and not with full symmetry.

Hints of such duality are already present in the formulas we've been developing, and we'll work our way into the subject by looking there first. In principle, the value of the conjugate of a given function can be calculated at a given point by maximization in terms of the defining expression 11(1). But the formulas developed in 11.23, and more specially in 11.34, furnish an alternative approach to calculating the same value by minimization. This idea is captured most succinctly by the operation of inf-projection.

11.38 Lemma (dual calculations in parametric optimization)**.** *For any function $f : I\!R^n \times I\!R^m \to \overline{I\!R}$, one has*

$$\left.\begin{aligned} p(0) &= \inf_x \varphi(x) \\ p^{**}(0) &= \sup_y \psi(y) \end{aligned}\right\} \quad \textit{for} \quad \left\{\begin{aligned} &p(u) = \inf_x f(x,u) \\ &\varphi(x) = f(x,0) \\ &\psi(y) = -f^*(0,y). \end{aligned}\right.$$

Proof. This is immediate from 11.23(c). □

The circumstances under which $p(0) = p^{**}(0)$, and therefore $\inf\varphi = \sup\psi$ in this scheme, are of course governed in general by 11.1 and 11.2 and are rich in possibilities. The focus in 11.38 on $u = 0$ enhances symmetry and corresponds to interpreting u as a *perturbation* parameter. Much will be made of this perturbation idea as we proceed.

The next theorem develops out of 11.38 a symmetric framework in which some of the most distinguishing features of optimization problems of *convex* type find expression. Later we'll explore the extent to which dualization can be effective through 11.38 even for problems involving nonconvexities.

11.39 Theorem (dual problems of optimization)**.** *Let $f : I\!R^n \times I\!R^m \to \overline{I\!R}$ be proper, lsc and convex, and consider the primal problem*

$$\text{minimize } \varphi \text{ on } I\!R^n, \qquad \varphi(x) := f(x,0),$$

along with the dual problem

$$\text{maximize } \psi \text{ on } I\!R^m, \qquad \psi(y) := -f^*(0,y),$$

where φ is convex and lsc, but ψ is concave and usc. Let $p(u) = \inf_x f(x,u)$ and $U = \operatorname{dom} p$, while $q(v) = \inf_y f^*(v,y)$ and $V = \operatorname{dom} q$; these sets and functions are convex.

(a) *The inequality $\inf_x \varphi(x) \geq \sup_y \psi(y)$ holds always, and $\inf_x \varphi(x) < \infty$ if and only if $0 \in U$, whereas $\sup_y \psi(y) > -\infty$ if and only if $0 \in V$. Moreover,*

$$\inf_x \varphi(x) = \sup_y \psi(y) \quad \text{if either } 0 \in \operatorname{int} U \text{ or } 0 \in \operatorname{int} V.$$

(b) *The set $\operatorname{argmax}_y \psi(y)$ is nonempty and bounded if and only if $0 \in \operatorname{int} U$ and the value $\inf_x \varphi(x) = p(0)$ is finite, in which case $\operatorname{argmax}_y \psi(y) = \partial p(0)$.*

(c) *The set $\operatorname{argmin}_x \varphi(x)$ is nonempty and bounded if and only if $0 \in \operatorname{int} V$ and the value $\sup_y \psi(y) = -q(0)$ is finite, in which case $\operatorname{argmin}_x \varphi(x) = \partial q(0)$.*

(d) *Optimal solutions are characterized jointly through primal and dual forms of Fermat's rule:*

$$\left.\begin{array}{r} \bar{x} \in \operatorname{argmin}_x \varphi(x) \\ \bar{y} \in \operatorname{argmax}_y \psi(y) \\ \inf_x \varphi(x) = \sup_y \psi(y) \end{array}\right\} \iff (0,\bar{y}) \in \partial f(\bar{x},0) \iff (\bar{x},0) \in \partial f^*(0,\bar{y}).$$

$$\boxed{\begin{array}{lll} \varphi(x) := f(x,0) & p(u) := \inf_x f(x,u) & U := \operatorname{dom} p \\ \quad\uparrow * & \quad\downarrow * & \\ q(v) := \inf_y f^*(v,y) & -\psi(y) := f^*(0,y) & V := \operatorname{dom} q \end{array}}$$

Fig. 11–5. Notation for dual problems of convex type.

Proof. The convexity in the preamble is obvious from that of f and f^*. (The preservation of convexity under inf-projection is attested to by 2.22(a).)

We apply 11.23(c) in the context of 11.38, noting from 11.1 and 11.2 that $p(0) = p^{**}(0)$ in particular when $0 \in \operatorname{int} U$. The latter condition in combination with $p(0) > -\infty$ is equivalent by 11.8(c) to $-\psi$ being proper and level-bounded. But a proper, lsc, convex function is level-bounded if and only if its argmin set is nonempty and bounded; cf. 3.27 and 1.9. Then too, $\partial p(0) = \operatorname{argmin} p^* = \operatorname{argmax} \psi$ by 11.8(a).

Next we invoke the same facts with f replaced by f^*, using the relation $f^{**} = f$. This gives $-\sup \psi = q(0) \geq q^{**}(0) = -\inf \varphi$ with $\varphi = q^*$, where $q(0) = q^{**}(0)$ in particular when $0 \in \operatorname{int} V$. The latter condition in combination with $q(0) > -\infty$ corresponds by parallel argument to having $\operatorname{argmin} \varphi$ being nonempty and bounded. It also gives $\partial q(0) = \operatorname{argmin} q^* = \operatorname{argmin} \varphi$.

Turning to (d), we note that through 11.3 the relations $(0, \bar{y}) \in \partial f(\bar{x}, 0)$ and $(\bar{x}, 0) \in \partial f^*(0, \bar{y})$ are equivalent to each other and to having $\varphi(\bar{x}) = \psi(\bar{y})$. Since $\inf \varphi \geq \sup \psi$ in general by (a), they are equivalent further to having $\varphi(\bar{x}) = \inf \varphi = \sup \psi = \psi(\bar{y})$. □

11.40 Corollary (general best-case primal-dual relations)**.** *In the context of Theorem 11.39, the following conditions are equivalent to each other and serve to guarantee that* $-\infty < \min \varphi = \max \psi < \infty$*:*

(a) $0 \in \operatorname{int} U$ *and* $0 \in \operatorname{int} V$*;*

(b) $0 \in \operatorname{int} U$*, and* $\operatorname{argmin} \varphi$ *is nonempty and bounded;*

(c) $0 \in \operatorname{int} V$*, and* $\operatorname{argmax} \psi$ *is nonempty and bounded;*

(d) $\operatorname{argmin} \varphi$ *and* $\operatorname{argmax} \psi$ *are nonempty and bounded.*

The relation $\operatorname{argmax} \psi = \partial p(0)$ in 11.39(b) shows the significance of optimal solutions $\bar{y}$ to the dual problem. In the situation in 11.39(b), the convex function p is finite on a neighborhood of 0, so the relation tells us that

$$dp(0)(w) \;=\; \max\bigl\{\langle \bar{y}, w\rangle \bigm| \bar{y} \in \operatorname{argmax} \psi\bigr\}$$

(cf. 8.30, 7.27) and further that a unique optimal solution $\bar{y}$ to the dual problem corresponds to having $\nabla p(0) = \bar{y}$ (cf. 9.18, 9.14). All this has to be seen in the light of $p(0)$ being the optimal value in the given 'primal' problem of minimizing φ, with $p(u)$ the optimal value obtained when this problem is perturbed by the amount u in the way prescribed by the chosen parametric representation.

This kind of interpretation of the dual elements $\bar{y}$ accompanying the primal elements $\bar{x}$ resembles one that was discussed in parametric minimization more generally (cf. 10.14 and 10.15), but without a dual *problem* being brought in for a supplementary description of the vectors $\bar{y}$ as optimal in their own right: $\bar{y} \in \operatorname{argmax} \psi$. In the present setup, fortified by convexity and the relation $\operatorname{argmin} \varphi = \partial q(0)$ in 11.39(c), the roles of $\bar{x}$ and $\bar{y}$ can be interchanged. Remarkably, the solutions $\bar{x} \in \operatorname{argmin} \varphi$ to the primal problem gain a parallel interpretation relative to perturbations to the dual problem, namely:

$$dq(0)(z) \;=\; \max\bigl\{\langle \bar{x}, z\rangle \bigm| \bar{x} \in \operatorname{argmin} \varphi\bigr\}.$$

Because $f^{**} = f$, everything is completely symmetric between primal and dual, apart from the sign convention in designating whether a problem should be viewed in maximization or minimization mode. The dualization scheme proceeds from a primal problem with perturbation vector u to a dual problem with perturbation vector v, and it does so in such a way that *the dual of the dual problem is identical to the primal problem.*

11.41 Example (Fenchel-type duality scheme). *Consider the two problems*

$$\begin{array}{ll} \textit{minimize } \varphi \textit{ on } I\!R^n, & \varphi(x) := \langle c, x\rangle + k(x) + h(b - Ax), \\ \textit{maximize } \psi \textit{ on } I\!R^m, & \psi(y) := \langle b, y\rangle - h^*(y) - k^*(A^*y - c), \end{array}$$

where $k : I\!R^n \to \overline{I\!R}$ *and* $h : I\!R^m \to \overline{I\!R}$ *are proper, lsc and convex, and one has* $A \in I\!R^{m\times n}$, $b \in I\!R^m$, *and* $c \in I\!R^n$. *These problems fit the format of 11.39 with*

$$\begin{aligned} f(x, u) &= \langle c, x\rangle + k(x) + h(b - Ax + u), \\ f^*(v, y) &= -\langle b, y\rangle + h^*(y) + k^*(A^*y - c + v). \end{aligned}$$

The assertions of 11.39 and 11.40 are valid then in the context of having

$$\begin{aligned} 0 \in \operatorname{int} U &\iff b \in \operatorname{int}\big(A \operatorname{dom} k + \operatorname{dom} h\big), \\ 0 \in \operatorname{int} V &\iff c \in \operatorname{int}\big(A^*\operatorname{dom} h^* - \operatorname{dom} k^*\big). \end{aligned}$$

Furthermore, optimal solutions are characterized by

$$\left.\begin{array}{c} \bar{x} \in \operatorname{argmin} \varphi \\ \bar{y} \in \operatorname{argmax} \psi \\ \inf \varphi = \sup \psi \end{array}\right\} \iff \left\{\begin{array}{c} \bar{y} \in \partial h(b - A\bar{x}) \\ A^*\bar{y} - c \in \partial k(\bar{x}) \end{array}\right\} \iff \left\{\begin{array}{c} \bar{x} \in \partial k^*(A^*\bar{y} - c) \\ b - A\bar{x} \in \partial h^*(\bar{y}) \end{array}\right\}.$$

Detail. The function f is proper, lsc (by 1.39, 1.40) and convex (by 2.18, 2.20). The function p, defined by $p(u) := \inf_x f(x, u)$, has nonempty effective domain $U = A \operatorname{dom} k + \operatorname{dom} h - b$. Direct calculation of f^* yields

$$\begin{aligned} f^*(v, y) &= \sup\nolimits_{x,u}\Big\{\big\langle v, x\big\rangle + \big\langle y, u\big\rangle - \big\langle c, x\big\rangle - k(x) - h(b - Ax + u)\Big\} \\ &= \sup\nolimits_{x,w}\Big\{\big\langle v, x\big\rangle + \big\langle y,\, w - b + Ax\big\rangle - \big\langle c, x\big\rangle - k(x) - h(w)\Big\} \\ &= \sup\nolimits_{x}\Big\{\big\langle A^*y - c + v, x\big\rangle - k(x)\Big\} + \sup\nolimits_{w}\Big\{\big\langle y, w\big\rangle - h(w)\Big\} - \big\langle y, b\big\rangle \\ &= k^*(A^*y - c + v) + h^*(y) - \big\langle y, b\big\rangle \end{aligned}$$

as claimed. The effective domain of the function $q(v) = \inf_y f^*(v, y)$ is then $V = \operatorname{dom} k^* - A^* \operatorname{dom} h^* + c$.

To determine $\partial f(x, u)$ so as to verify the optimality condition claimed, we write $f = g\circ F$ for $g(x, w) = \langle c, x\rangle + k(x) + h(b+w)$ and $F(x, u) = (x, -Ax+u)$, noting that F is linear and nonsingular. We observe then that $\partial g(x, w) = \big\{(c + v, y) \,\big|\, v \in \partial k(x),\, y \in \partial h(b + w)\big\}$, and therefore through 10.7 that we have $\partial f(x, u) = \big\{(c + v - A^*y, y) \,\big|\, v \in \partial k(x),\, y \in \partial h(b - Ax + u)\big\}$. The condition $(0, \bar{y}) \in \partial f(\bar{x}, 0)$ in 11.39(d) reduces to having $A^*\bar{y} - c \in \partial k(\bar{x})$ for

$\bar{y} \in \partial h(b - A\bar{x})$. The alternate expression of optimality in terms of ∂k^* and ∂h^* comes immediately then from the inversion principle in 11.3. □

11.42 Theorem (piecewise linear-quadratic optimization). *For a proper, convex and piecewise linear-quadratic function $f : \mathbb{R}^n \times \mathbb{R}^m \to \overline{\mathbb{R}}$, consider the primal and dual problems*

$$\begin{aligned} &\textit{minimize } \varphi \textit{ on } \mathbb{R}^n, \quad \varphi(x) := f(x,0), \\ &\textit{maximize } \psi \textit{ on } \mathbb{R}^m, \quad \psi(y) := -f^*(0,y), \end{aligned}$$

along with the functions $p(u) = \inf_x f(x,u)$ and $q(v) = \inf_y f^*(v,y)$.

If either of the values $\inf \varphi$ or $\sup \psi$ is finite, then both are finite and both are attained. Moreover in that case one has $\inf \varphi = \sup \psi$ and

$$\begin{aligned} (\operatorname{argmin} \varphi) \times (\operatorname{argmax} \psi) &= \big\{ (\bar{x}, \bar{y}) \,\big|\, (0, \bar{y}) \in \partial f(\bar{x}, 0) \big\} \\ &= \big\{ (\bar{x}, \bar{y}) \,\big|\, (\bar{x}, 0) \in \partial f^*(0, \bar{y}) \big\} \;=\; \partial q(0) \times \partial p(0). \end{aligned}$$

Proof. This parallels Theorem 11.39, but in coming up with circumstances in which $p^{**}(0) = p(0)$, or $q^{**}(0) = q(0)$, it relies on the piecewise linear-quadratic nature of p and q in 11.32 instead of properties of general convex functions. □

11.43 Example (linear and extended linear-quadratic programming). *The problems in the Fenchel duality scheme in 11.41 fit the framework of piecewise linear-quadratic optimization in 11.42 when the convex functions k and h are piecewise linear-quadratic. A particular case of this, called extended linear-quadratic programming, concerns the primal and dual problems*

$$\begin{aligned} &\textit{minimize } \langle c, x \rangle + \tfrac{1}{2} \langle x, Cx \rangle + \theta_{Y,B}(b - Ax) \textit{ over } x \in X, \\ &\textit{maximize } \langle b, y \rangle - \tfrac{1}{2} \langle y, By \rangle - \theta_{X,C}(A^* y - c) \textit{ over } y \in Y, \end{aligned}$$

where the sets $X \subset \mathbb{R}^n$ and $Y \subset \mathbb{R}^m$ are nonempty and polyhedral, the matrices $C \in \mathbb{R}^{n \times n}$ and $B \in \mathbb{R}^{m \times m}$ are symmetric and positive-semidefinite, and the functions $\theta_{Y,B}$ and $\theta_{X,C}$ have expressions as in 11.18:

$$\theta_{Y,B}(u) = \sup_{y \in Y} \Big\{ \langle y, u \rangle - \tfrac{1}{2} \langle y, By \rangle \Big\}, \qquad \theta_{X,C}(v) = \sup_{x \in X} \Big\{ \langle v, x \rangle - \tfrac{1}{2} \langle x, Cx \rangle \Big\}.$$

When $C = 0$ and $B = 0$, while X and Y are cones, these primal and dual problems reduce to the linear programming problems

$$\begin{aligned} &\textit{minimize } \langle c, x \rangle \textit{ subject to } x \in X, \; b - Ax \in Y^*, \\ &\textit{maximize } \langle b, y \rangle \textit{ subject to } y \in Y, \; A^* y - c \in X^*. \end{aligned}$$

If in addition $X = \mathbb{R}^n_+$ and $Y = \mathbb{R}^m_+$, the linear programming problems have the form

$$\begin{aligned} &\textit{minimize } \langle c, x \rangle \textit{ subject to } x \geq 0, \; Ax \geq b, \\ &\textit{maximize } \langle b, y \rangle \textit{ subject to } y \geq 0, \; A^* y \leq c. \end{aligned}$$

More generally, if $X = I\!R^r_+ \times I\!R^{n-r}$ *and* $Y = I\!R^s_+ \times I\!R^{m-s}$ *these linear programming problems have mixed inequality and equality constraints.*

Detail. When k and h are piecewise linear-quadratic in the Fenchel scheme in 11.41, so too is f by the calculus in 10.22. Then Theorem 11.42 is applicable. In the special case described, we have $k = \delta_X + j_C$ for $j_C(x) := \frac{1}{2}\langle x, Cx\rangle$, whereas $h = (\delta_Y + j_B)^*$; cf. 11.18. □

Between linear programming and extended linear-quadratic programming in 11.43 is *quadratic programming*, where the primal problem takes the form

$$\text{minimize } \langle c, x\rangle + \tfrac{1}{2}\langle x, Cx\rangle \text{ subject to } x \in X,\ b - Ax \in Y^*,$$

for some choice of polyhedral cones $X \subset I\!R^n$, $Y \subset I\!R^m$ (such as $X = I\!R^n_+$, $Y = I\!R^m_+$) and a positive semidefinite matrix $C \in I\!R^{n\times n}$. This corresponds to the primal problem of extended linear-quadratic programming of 11.43 in the case of $B = 0$ and dualizes to

$$\text{maximize } \langle b, y\rangle - \theta_{X,C}(A^*y - c) \text{ over } y \in Y.$$

Thus, *the dual of a quadratic programming problem isn't another quadratic programming problem*, except in the linear programming subcase.

The general duality framework in Theorem 11.39 can be useful also in the derivation of conjugacy formulas. The next example shows this while demonstrating how the criteria in 11.39 can be verified in a particular case.

11.44 Example (dualized composition). *Let* $g : I\!R^n \to \overline{I\!R}$ *and* $\theta : I\!R \to \overline{I\!R}$ *be lsc, proper and convex with* $\operatorname{dom}\theta \subset I\!R_+$ *and* $\lim_{\lambda\to\infty} \theta(\lambda)/\lambda > -g(0)$. *With* $g^\infty(z)$ *replacing* $\lambda g(\lambda^{-1}z)$ *when* $\lambda = 0$, *one has for all* $z \in I\!R^n$ *that*

$$\inf_{\lambda \geq 0}\big\{\theta(\lambda) + \lambda g(\lambda^{-1}z)\big\} = (\theta^*\circ g^*)^*(z).$$

Detail. The assumption on $\operatorname{dom}\theta$ is equivalent to θ^* being nondecreasing; cf. 8.51. In particular, therefore, $\theta^*\circ g^*$ is another convex function; cf. 2.20(b).

Fixing any z, define $f(\lambda, u)$ to be $\theta(\lambda) + h(\lambda, z + u)$, with $h(\lambda, w) = \lambda g(\lambda^{-1}w)$ when $\lambda > 0$ but $g^\infty(w)$ when $\lambda = 0$; set $h(\lambda, w) = \infty$ when $\lambda < 0$. The left side of the claimed formula is then the optimal value in the problem of minimizing $\varphi(\lambda) = f(\lambda, 0)$ over $\lambda \in I\!R^1$.

The function f is lsc, proper and convex by 3.49(c), so we're in the territory of Theorem 11.39. To proceed, we have to determine f^*. Calculating from the definition (with μ as the variable dual to λ), we get

$$\begin{aligned} f^*(\mu, y) &= \sup_{\lambda,u}\Big\{\big\langle(\mu, y), (\lambda, u)\big\rangle - f(\lambda, u)\Big\} \\ &= \sup_{\lambda}\Big\{\mu\lambda - \theta(\lambda) + \sup_{w}\big\{\langle y, w - z\rangle - h(\lambda, w)\big\}\Big\}, \end{aligned} \tag{11(15)}$$

where the inner supremum only has to be analyzed for $\lambda \geq 0$, inasmuch as $\operatorname{dom}\theta \subset I\!R_+$. For $\lambda > 0$ it comes out as

$$\sup_w \big\{ \langle y,\, w - z \rangle - (\lambda \star g)(w) \big\} = \lambda g^*(y) - \langle y, z \rangle,$$

cf. 11(3). For $\lambda = 0$, we use the fact in 11.5 that g^∞ is the support function of $D = \operatorname{dom} g^*$ in order to see that the inner supremum is

$$\sup_w \big\{ \langle y,\, w - z \rangle - \sigma_D(w) \big\} = \delta_{\operatorname{cl} D}(y) - \langle y, z \rangle,$$

cf. 11.4(a). Substituting these into the last part of 11(15), we find that

$$f^*(\mu, y) = \theta^*\big(g^*(y) + \mu\big) - \langle y, z \rangle$$

(with $\theta^*(\infty)$ interpreted as ∞). The dual problem, which consists of maximizing $\psi(y) = -f^*(0, y)$ over $y \in \mathbb{R}^n$, therefore has optimal value $\sup \psi = (\theta^* \circ g^*)^*(z)$, which is the value on the right side of the claimed formula.

We can obtain the desired conclusion by verifying that $\inf \varphi = \sup \psi$. A criterion is provided for this purpose in 11.39(a): it suffices to know that $0 \in \operatorname{int}\big\{ \mu \,\big|\, \exists\, y \text{ with } g^*(y) + \mu \in \operatorname{dom} \theta^* \big\}$. Because θ^* is nondecreasing, $\operatorname{dom} \theta^*$ is an interval that's unbounded below; its right endpoint (maybe ∞) is $\theta^\infty(1)$ by 11.5. What we need to have is $\inf g^* < \theta^\infty(1)$. But $\inf g^* = -g(0)$ by 11.8(a) as applied to g^*. The criterion thus means $-g(0) < \theta^\infty(1)$. Since $\theta^\infty(1) = \lim_{\lambda \nearrow \infty} \theta(\lambda)/\lambda$, we've reached our goal. $\square$

I. Lagrangian Functions

Duality in the elegant style of Theorem 11.39 and its accompaniment is a feature of optimization problems of convex type only. In general, for an optimization problem represented as minimizing $\varphi = f(\cdot, 0)$ over $\mathbb{R}^n$ for a function $f : \mathbb{R}^n \times \mathbb{R}^m \to \overline{\mathbb{R}}$, regardless of convexity, we speak of the problem of maximizing $\psi = -f^*(0, \cdot)$ over $\mathbb{R}^n$ as the associated *dual* problem, in contrast to the given problem as the *primal* problem, but the relationships might not be as tight as the ones we've been seeing. The issue of whether $\inf \varphi = \sup \psi$ comes down to whether $p(0) = p^{**}(0)$ for $p(u) = \inf_x f(x, u)$, as observed in 11.38. The trouble is that in the absence of p being convex—which is hard to guarantee without simply assuming f is convex—there's no strong handle on whether $p(0) = p^{**}(0)$. We'll nonetheless eventually uncover some facts of considerable power about nonconvex duality and how it can be put to use.

An important step along the way is the study of general 'Lagrangian functions'. That has other motivations as well, most notably in the expression of optimality conditions and the development of methods for computing optimal solutions. Although tradition merely associates Lagrangian functions with constraint systems, their role can go far beyond just that.

The key idea is that a Lagrangian arises through *partial dualization* of a given problem in relation to a particular choice of perturbation parameters.

11.45 Definition (Lagrangians and dualizing parameterizations)**.** *For a problem of minimizing φ on $\mathbb{R}^n$, a dualizing parameterization is a representation $\varphi = f(\cdot, 0)$ in terms of a proper function $f : \mathbb{R}^n \times \mathbb{R}^m \to \overline{\mathbb{R}}$ such that $f(x, u)$ is lsc convex in u. The associated Lagrangian $l : \mathbb{R}^n \times \mathbb{R}^m \to \overline{\mathbb{R}}$ is given by*

$$l(x, y) \;:=\; \inf_u \Big\{ f(x, u) - \langle y, u \rangle \Big\}. \qquad 11(16)$$

This definition has its roots in the Legendre-Fenchel transform: for each $x \in \mathbb{R}^n$ the function $-l(x, \cdot)$ is conjugate to $f(x, \cdot)$ on $\mathbb{R}^m$. The conditions on f have the purpose of ensuring that $f(x, \cdot)$ is in turn conjugate to $-l(x, \cdot)$:

$$f(x, u) \;=\; \sup_y \Big\{ l(x, y) + \langle y, u \rangle \Big\}. \qquad 11(17)$$

Indeed, they require $f(x, \cdot)$ to be proper, lsc and convex, unless $f(x, \cdot) \equiv \infty$; either way, $f(x, \cdot)$ then coincides with its biconjugate by 11.1 and 11.2.

The convexity of $f(x, u)$ in u is a vastly weaker requirement than convexity in (x, u), although it's certainly implied by the latter. The parametric representations in 11.39, 11.41, and 11.42 are dualizing parameterizations in particular. Before going any further with those cases, however, let's look at how Definition 11.45 captures the notion of a Lagrangian function as a vehicle for conditions about Lagrange multipliers. It does so with such generality that the multipliers aren't merely associated with constraints but equally well with penalties and other features of composite modeling structure in optimization.

11.46 Example (multiplier rule in Lagrangian form)**.** *Consider the problem*

$$\textit{minimize} \;\; f_0(x) + \theta\big(f_1(x), \ldots, f_m(x)\big) \;\; \textit{over} \;\; x \in X$$

for a nonempty, closed set $X \subset \mathbb{R}^n$, smooth functions $f_i : \mathbb{R}^n \to \mathbb{R}$, and a proper, lsc, convex function $\theta : \mathbb{R}^m \to \overline{\mathbb{R}}$. Taking $F(x) = \big(f_1(x), \ldots, f_m(x)\big)$, identify this with the problem of minimizing $\varphi(x) = f(x, 0)$ over $x \in \mathbb{R}^n$ for

$$f(x, u) \;=\; \delta_X(x) + f_0(x) + \theta\big(F(x) + u\big).$$

This furnishes a dualizing parameterization for which the Lagrangian is

$$l(x, y) \;=\; \delta_X(x) + f_0(x) + \big\langle y, F(x) \big\rangle - \theta^*(y)$$

(with $\infty - \infty = \infty$ on the right). The optimality condition in the extended multiplier rule of 10.8, namely,

$$-\big[\nabla f_0(\bar{x}) + \nabla F(\bar{x})^* \bar{y}\big] \in N_X(\bar{x}) \;\; \textit{for some} \;\; \bar{y} \in \partial\theta\big(F(\bar{x})\big),$$

can then be written equivalently in the Lagrangian form

$$0 \in \partial_x l(\bar{x}, \bar{y}), \qquad 0 \in \partial_y[-l](\bar{x}, \bar{y}).$$

As a particular case, when $\theta = \theta_{Y,B}$ for a closed, convex set $Y \subset \mathbb{R}^m$ and a *symmetric, positive-semidefinite matrix* $B \in I\!R^{m\times m}$ *as in 11.18, possibly with* $B = 0$ *, the Lagrangian specializes to*

$$l(x,y) = \delta_X(x) + L(x,y) - \delta_Y(y) \quad \text{for } L(x,y) = f_0(x) + \langle y, F(x)\rangle - \tfrac{1}{2}\langle y, By\rangle,$$

and then the Lagrangian form of the multiplier rule specializes to

$$-\nabla_x L(\bar{x},\bar{y}) \in N_X(\bar{x}), \qquad \nabla_y L(\bar{x},\bar{y}) \in N_Y(\bar{y}).$$

The case of $\theta = \delta_K$ *for a closed, convex cone corresponds to* $B = 0$ *and* $Y = K^*$.

Detail. It's elementary that Definition 11.45 yields the expression claimed for $l(x,y)$. (The convention about infinities reconciles what happens when both $x \notin X$ and $y \notin \operatorname{dom}\theta^*$; we have $f(x,\cdot) \equiv \infty$ when $x \notin X$, so the conjugate function $-l(x,\cdot)$ is then correctly the constant function $-\infty$, i.e., we have $l(x,\cdot) \equiv \infty$.) Subgradient inversion through 11.3 turns the multiplier rule into the condition on subgradients of l.

When $\theta = \theta_{Y,B}$ as in 11.18, we have $\theta = (\delta_Y + j_B)^*$ for $j_B(y) := \frac{1}{2}\langle y, By\rangle$, hence $\theta^* = \delta_Y + j_B$ by 11.1. The subgradient condition calculates out then to the one stated in terms of $\nabla_x L(\bar{x},\bar{y})$ and $\nabla_y L(\bar{x},\bar{y})$; cf. 8.8(c). □

The possibility of expressing conditions for optimality in Lagrangian form is useful for many purposes, such as the design of numerical methods. The Lagrangian brings out properties of the problem that otherwise might be obscured. This is seen in Example 11.46 when θ is a function of type $\theta_{Y,B}$, which may lack continuous derivatives through the incorporation of various penalty terms. From some perspectives, such nonsmoothness could be a handicap in contemplating how to minimize $f_0(x) + \theta\big(F(x)\big)$ directly. But in working with the Lagrangian $L(x,y)$, one has a finite, smooth function on a relatively simple product set $X \times Y$, and that may be more approachable.

Problems of extended linear-quadratic programming fit Example 11.46 with $f_0(x) = \langle c, x\rangle + \frac{1}{2}\langle x, Cx\rangle$, $F(x) = b - Ax$ and $\theta = \theta_{Y,B}$. Their Lagrangians can also be viewed as specializing the ones associated with problems in the Fenchel scheme.

11.47 Example (Lagrangians in the Fenchel scheme)**.** *In the problem formulation of 11.41, the Lagrangian function is given by*

$$l(x,y) \ = \ \langle c, x\rangle + k(x) + \langle b, y\rangle - h^*(y) - \langle y, Ax\rangle$$

(with $\infty - \infty = \infty$ *on the right). The optimality conditions at the end of 11.41 can be written equivalently in the Lagrangian form*

$$0 \in \partial_x l(\bar{x},\bar{y}), \qquad 0 \in \partial_y[-l](\bar{x},\bar{y}).$$

For the special case of extended linear-quadratic programming described in 11.43, the Lagrangian reduces to

$$l(x,y) = \delta_X(x) + L(x,y) - \delta_Y(y)$$
$$\text{for } L(x,y) = \langle c,x\rangle + \tfrac{1}{2}\langle x,Cx\rangle + \langle b,y\rangle - \tfrac{1}{2}\langle y,By\rangle - \langle y,Ax\rangle,$$

and the optimality condition translates then to

$$-\nabla_x L(\bar{x},\bar{y}) \in N_X(\bar{x}), \qquad \nabla_y L(\bar{x},\bar{y}) \in N_Y(\bar{y}),$$

or in other words,

$$-c - C\bar{x} + A^*\bar{y} \in N_X(\bar{x}), \qquad b - A\bar{x} - B\bar{y} \in N_Y(\bar{y}).$$

Detail. The Lagrangian can be calculated right from Definition 11.45. The convention about $\infty - \infty$ comes in because the formula makes $l(x,y)$ have the value ∞ when $k(x) = \infty$, even if $h^*(y) = \infty$. The Lagrangian form of the optimality condition merely restates the conditions at the end of 11.41 in an alternative manner. □

Here $l(x,y)$ is convex in x and concave in y. This is characteristic of the convex duality framework in general.

11.48 Proposition (convexity properties of Lagrangians). *For any dualizing parameterization $\varphi = f(\cdot,0)$, the associated Lagrangian $l(x,y)$ is usc concave in y. It is convex in x besides if and only if $f(x,u)$ is convex in (x,u) rather than just with respect to u. In that case, one has*

$$(v,y) \in \partial f(x,u) \quad \Longleftrightarrow \quad v \in \partial_x l(x,y),\ u \in \partial_y[-l](x,y),$$

the value $l(x,y)$ in these circumstances being finite and equal to $f(x,u) - \langle y,u\rangle$.

Proof. The fact that $l(x,y)$ is usc concave in y is obvious from $-l(x,\cdot)$ being conjugate to $f(x,\cdot)$. If $f(x,u)$ is convex in (x,u), so too for any $y \in \mathbb{R}^m$ is the function $f_y(x,u) := f(x,u) - \langle y,u\rangle$, and then $\inf_u f_y(x,u)$ is convex because the inf-projection of any convex function is convex; cf. 2.22. But $\inf_u f_y(x,u) = l(x,y)$ by definition. Thus, the convexity of $f(x,u)$ in (x,u) implies the convexity of $l(x,y)$ in x.

Conversely, if $l(x,y)$ is convex in x, the function $l_y(x,u) := l(x,y) + \langle y,u\rangle$ is convex in (x,u). But according to 11(17), f is the pointwise supremum of the collection of functions l_y as y ranges over $\mathbb{R}^m$. Since the pointwise supremum of a collection of convex functions is convex, we conclude that in this case $f(x,u)$ is convex in (x,u).

To check the equivalence in subgradient relations for particular x_0, u_0, v_0 and y_0, observe from 11.3 and the conjugacy between $f(x_0,\cdot)$ and $-l(x_0,\cdot)$ that the conditions $y_0 \in \partial_u f(x_0,u_0)$ and $u_0 \in \partial_y[-l](x_0,y_0)$ are equivalent to each other and to having $l(x_0,y_0) = f(x_0,u_0) - \langle y_0,u_0\rangle$ (finite). When f is convex, we can appeal to 8.12 to write the full condition $(v_0,y_0) \in \partial f(x_0,u_0)$ as the inequality

$$f(x,u) \ \geq\ f(x_0,u_0) + \langle v_0,\, x - x_0\rangle + \langle y_0,\, u - u_0\rangle \ \text{ for all } x,\, u,$$

cf. 8.12. From the case of $x = x_0$ we see that this entails $y_0 \in \partial_u f(x_0, u_0)$ and therefore is equivalent to having $u_0 \in \partial_y[-l](x_0, y_0)$ along with

$$\inf_u \{ f(x,u) - \langle y_0, u\rangle \} \;\geq\; f(x_0,u_0) - \langle y_0, u_0\rangle + \langle v_0,\, x - x_0\rangle \;\text{ for all }\; x.$$

But the latter translates to $l(x, y_0) \geq l(x_0, y_0) + \langle v_0, x - x_0\rangle$ for all x, which by 8.12 and the convexity of $l(x, y_0)$ in x means that $v_0 \in \partial_x l(x_0, y_0)$. Thus, $(v_0, y_0) \in \partial f(x_0, u_0)$ if and only if $v_0 \in \partial_x l(x_0, y_0)$ and $u_0 \in \partial_y[-l](x_0, y_0)$. □

Whenever $l(x, y)$ is convex in x as well as concave in y, the Lagrangian condition in 11.46 and 11.47 can be interpreted through the following concept.

11.49 Definition (saddle points). *A vector pair $(\bar{x}, \bar{y})$ is said to be a saddle point of the function l on $\mathbb{R}^n \times \mathbb{R}^m$ (in the minimax sense, and under the convention of minimizing in the first argument and maximizing in the second) if $\inf_x l(x, \bar{y}) = l(\bar{x}, \bar{y}) = \sup_y l(\bar{x}, y)$, or in other words*

$$l(x,\bar{y}) \;\geq\; l(\bar{x},\bar{y}) \;\geq\; l(\bar{x},y) \;\text{ for all }\; x \;\text{ and }\; y. \tag{11(18)}$$

The set of all such saddle points $(\bar{x}, \bar{y})$ is denoted by argminimax l*, or in more detail by* $\operatorname{argminimax}_{x,y} l(x, y)$*.*

Likewise in the case of a function L on a product set $X \times Y$, a pair $(\bar{x}, \bar{y})$ is said to be a saddle point of L with respect to $X \times Y$ if

$$\begin{cases} \bar{x} \in X, \;\; \bar{y} \in Y, \;\; \text{and} \\ L(x,\bar{y}) \geq L(\bar{x},\bar{y}) \geq L(\bar{x},y) \;\text{ for all }\; x \in X,\; y \in Y. \end{cases} \tag{11(19)}$$

The notation for the saddle point set is then

$$\operatorname{argminimax}_{X,Y} L, \qquad \text{or} \;\; \operatorname*{argminimax}_{x\in X,\; y\in Y} L(x,y).$$

11.50 Theorem (minimax relations). *Let $l : \mathbb{R}^n \times \mathbb{R}^m \to \overline{\mathbb{R}}$ be the Lagrangian for a problem of minimizing φ on $\mathbb{R}^n$ with dualizing parameterization $\varphi = f(\cdot, 0)$, $f : \mathbb{R}^n \times \mathbb{R}^m \to \overline{\mathbb{R}}$. Let $\psi = -f^*(0, \cdot)$ on $\mathbb{R}^m$. Then*

$$\varphi(x) = \sup_y l(x,y), \qquad \psi(y) = \inf_x l(x,y), \tag{11(20)}$$

$$\inf_x \varphi(x) \;=\; \inf_x \big[\sup_y l(x,y)\big] \;\geq\; \sup_y \big[\inf_x l(x,y)\big] \;=\; \sup_y \psi(y), \tag{11(21)}$$

and furthermore

$$\left.\begin{aligned} \bar{x} &\in \operatorname{argmin}_x \varphi(x) \\ \bar{y} &\in \operatorname{argmax}_y \psi(y) \\ \inf_x \varphi(x) &= \sup_y \psi(y) \end{aligned}\right\} \iff (\bar{x},\bar{y}) \in \operatorname{argminimax}_{x,y} l(x,y) \tag{11(22)}$$

$$\iff \varphi(\bar{x}) = \psi(\bar{y}) = l(\bar{x},\bar{y}).$$

The saddle point condition $(\bar{x}, \bar{y}) \in \operatorname{argminimax}_{x,y} l(x, y)$ always entails the *subgradient condition*

$$0 \in \partial_x l(\bar{x}, \bar{y}), \qquad 0 \in \partial_y[-l](\bar{x}, \bar{y}), \tag{11(23)}$$

and is equivalent to it whenever $l(x, y)$ is convex in x and concave in y, in which case it is also the same as having $(0, \bar{y}) \in \partial f(\bar{x}, 0)$.

Proof. The expression for φ in 11(20) comes from 11(17) with $u = 0$, while the one for ψ is based on the observation that the formula

$$-f^*(v, y) \;=\; \inf_{x,u}\big\{f(x, u) - \langle v, x\rangle - \langle y, u\rangle\big\}$$

can be rewritten through 11(16) as

$$-f^*(v, y) \;=\; \inf_x\big\{l(x, y) - \langle v, x\rangle\big\}. \tag{11(24)}$$

From 11(20) we immediately get 11(21), since $\inf \varphi = p(0)$ and $\sup \psi = p^{**}(0)$ for $p(u) := \inf_x f(x, u)$; cf. 11.38 with $u = 0$. We then have 11(22), whose right side is now seen to be just another way of writing $\varphi(\bar{x}) = \psi(\bar{y})$. The final assertion about subgradients dualizes 11(23) through the relation in 11.48. □

An interesting consequence of Theorem 11.50 is the fact that the set of saddle points is always a product set:

$$\operatorname{argminimax} l \;=\; \begin{cases} (\operatorname{argmin} \varphi) \times (\operatorname{argmax} \psi) & \text{if } \inf \varphi = \sup \psi, \\ \emptyset & \text{if } \inf \varphi > \sup \psi. \end{cases} \tag{11(25)}$$

Moreover, l is constant on $\operatorname{argminimax} l$. This constant is called the *saddle value* of l and is denoted by $\operatorname{minimax} l$.

The crucial question of whether $\inf \varphi = \sup \psi$ in our Lagrangian setting is that of whether $p(0) = p^{**}(0)$ for $p(u) := \inf_x f(x, u)$, as we know already from 11.38. We'll return to this presently.

11.51 Corollary (saddle point conditions for convex problems)**.** *Consider a problem of minimizing $\varphi(x)$ over $x \in I\!R^n$ in the convex duality framework of 11.39, and let $l(x, y)$ be the corresponding Lagrangian. Suppose $0 \in \operatorname{int} U$, or merely that $0 \in U$ but f is piecewise linear-quadratic.*

Then for $\bar{x}$ to be an optimal solution it is necessary and sufficient that there exist a vector $\bar{y}$ such that $(\bar{x}, \bar{y})$ is a saddle point of the Lagrangian l on $I\!R^n \times I\!R^m$. Furthermore, $\bar{y}$ appears in this role with $\bar{x}$ if and only if $\bar{y}$ is an optimal solution to the dual problem of maximizing $\psi(y)$ over $y \in I\!R^m$.

Proof. From $0 \in \operatorname{int} U$ we have $\inf \varphi = \sup \psi < \infty$ by 11.39(a) along with the further fact in 11.39(b) that $\operatorname{argmax} \psi$ contains a vector $\bar{y}$ if $\inf \varphi > -\infty$. When $\bar{x} \in \operatorname{argmin} \varphi$, $\varphi(\bar{x})$ is finite (by the definition of 'argmin'). The claims follow then from the equivalences at the end of Theorem 11.50; cf. the preceding discussion. The piecewise linear-quadratic case substitutes the stronger results in 11.42 for those in 11.39. □

For instance, saddle points of the Lagrangian in 11.46 characterize optimality in the Fenchel scheme in 11.41 when $b \in \operatorname{int}\big(A \operatorname{dom} k + \operatorname{dom} h\big)$. When k and h are piecewise linear-quadratic, the sharper form of 11.51 is applicable—no constraint qualification is needed.

J*. Minimax Problems

Problems of minimizing φ and maximizing ψ, in which φ and ψ are derived from some function l on $\mathbb{R}^n \times \mathbb{R}^m$ by the formulas in 11(18), are well known in game theory. It's interesting to see that the circumstances in which such problems also form a primal-dual pair arising out of a dualizing parameterization are completely described by the foregoing. All we need to know is that $l(x,y)$ is concave and usc in y, and that for each x either $l(x,\cdot) < \infty$ or $l(x,\cdot) \equiv \infty$. These conditions ensure that in defining f by 11(17) we get $\varphi = f(\cdot,0)$ for a dualizing parameterization such that the associated Lagrangian is l.

11.52 Example (minimax problems). *Consider nonempty, closed, convex sets $X \subset \mathbb{R}^n$, $Y \subset \mathbb{R}^m$, and a continuous function $L : X \times Y \to \mathbb{R}$ with $L(x,y)$ convex in $x \in X$ for $y \in Y$ and concave in $y \in Y$ for $x \in X$. Let*

$$l(x,y) = \begin{cases} L(x,y) & \text{for } x \in X,\ y \in Y, \\ -\infty & \text{for } x \in X,\ y \notin Y, \\ \infty & \text{for } x \notin X, \end{cases}$$

$$\varphi(x) = \begin{cases} \sup_{y \in Y} L(x,y) & \text{for } x \in X, \\ \infty & \text{for } x \notin X, \end{cases}$$

$$\psi(x) = \begin{cases} \inf_{x \in X} L(x,y) & \text{for } y \in Y, \\ -\infty & \text{for } y \notin Y. \end{cases}$$

The saddle point set $\operatorname{argminimax}_{X,Y} L$*, which is closed and convex, coincides then with* $\operatorname{argminimax} l$ *and has the product structure in 11(25); on this set, L is constant, the saddle value* $\operatorname{minimax}_{X,Y} L$ *being* $\operatorname{minimax} l$*.*

Indeed, l is the Lagrangian for the problem of minimizing φ over $\mathbb{R}^n$ under the dualizing parameterization furnished by

$$f(x,u) = \begin{cases} \sup_{y \in Y} \big\{ L(x,y) + \langle y,u \rangle \big\} & \text{for } x \in X, \\ \infty & \text{for } x \notin X. \end{cases} .$$

The function f is lsc, proper, and convex with

$$-f^*(v,y) = \begin{cases} \inf_{x \in X} \big\{ L(x,y) - \langle v,x \rangle \big\} & \text{for } y \in Y, \\ -\infty & \text{for } y \notin Y, \end{cases} .$$

so the corresponding dual problem is that of maximizing ψ over $\mathbb{R}^m$.

If L is smooth on an open set that includes $X \times Y$, $\operatorname{argminimax}_{X,Y} L$ *consists of the pairs $(\bar{x},\bar{y}) \in X \times Y$ satisfying*

$$-\nabla_x L(\bar{x},\bar{y}) \in N_X(\bar{x}), \qquad \nabla_y L(\bar{x},\bar{y}) \in N_Y(\bar{y}).$$

The equivalent conditions in 11.40 are necessary and sufficient for this saddle point set to be nonempty and bounded. In particular, the saddle point set is nonempty and bounded when X and Y are bounded.

Detail. Obviously $f(x,0) = \varphi(x)$. For $x \in X$, the function $-l(x,\cdot)$ is lsc, proper and convex with conjugate $f(x,\cdot)$, while for $x \notin X$ it is the constant function $-\infty$, whereas $f(x,\cdot)$ is the constant function ∞. Hence $-l(x,\cdot)$ is the conjugate of $f(x,\cdot)$ for all $x \in I\!R^n$. Thus, f furnishes a dualizing parameterization for the problem of minimizing φ on $I\!R^n$, and the associated Lagrangian is l. We have $l(x,y)$ convex in x and concave in y, so f is convex by 11.48. The continuity of L on $X \times Y$ ensures that $-l(x,\cdot)$ depends epi-continuously on $x \in X$, and the same then holds for $f(x,\cdot)$ by Theorem 11.34. It follows that f is lsc on $I\!R^n \times I\!R^m$. The rest is clear then from 11.50. When X and Y are bounded, the sets U and V in 11.40 are all of $I\!R^n$ and $I\!R^m$. □

A powerful rule about the behavior of optimal values in parameterized problems of convex type comes out of this principle.

11.53 Theorem (perturbation of saddle values; Golshtein)**.** *Consider nonempty, closed, convex sets $X \subset I\!R^n$ and $Y \subset I\!R^m$, and an interval $[0,T] \subset I\!R$. Let $L : [0,T] \times X \times Y \to I\!R$ be continuous, with $L(t,x,y)$ convex in $x \in X$ and concave in $y \in Y$, and suppose that the saddle point set*

$$X_0 \times Y_0 = \operatorname*{argminimax}_{x\in X,\, y\in Y} L(0,x,y)$$

is nonempty and bounded. Then, relative to t in an interval $[0,\varepsilon]$ for some $\varepsilon > 0$, the saddle point set $\operatorname{argminimax}_{X,Y} L(t,\cdot,\cdot)$ is nonempty and bounded, the mapping $t \mapsto \operatorname{argminimax}_{X,Y} L(t,\cdot,\cdot)$ is osc and locally bounded at $t = 0$, and the saddle value

$$\lambda(t) := \operatorname*{minimax}_{x\in X,\, y\in Y} L(t,x,y)$$

converges to $\lambda(0)$ as $t \searrow 0$. If in addition the limit

$$L'_+(0,x,y) := \lim_{\substack{t \searrow 0 \\ x' \xrightarrow[X]{} x \\ y' \xrightarrow[Y]{} y}} \frac{L(t,x',y') - L(0,x',y')}{t} \qquad 11(26)$$

exists for all $(x,y) \in X_0 \times Y_0$, then the value $L'_+(0,x,y)$ is continuous relative to $(x,y) \in X_0 \times Y_0$, convex in $x \in X_0$ for each $y \in Y_0$, and concave in $y \in Y_0$ for each $x \in X_0$. Indeed, the right derivative $\lambda'_+(0)$ exists and is given by

$$\lambda'_+(0) = \operatorname*{minimax}_{x\in X_0,\, y\in Y_0} L'_+(0,x,y).$$

Proof. We put ourselves in the picture of Example 11.52 and its detail, but with everything depending additionally on $t \in [0,T]$, in particular $f(t,x,u)$.

In extension of the earlier argument, the mapping $(t,x) \mapsto f(t,x,\cdot)$ is epi-continuous relative to $[0,T] \times X$. From this it follows that the mapping $t \mapsto f(t,\cdot,\cdot)$ is epi-continuous as well. The convex functions $p(t,\cdot)$ defined by $p(t,u) = \inf_x f(t,x,u)$ and the convex sets $U(t) = \operatorname{dom} p(t,\cdot)$ then have

$$p(0,\cdot) = \operatorname*{e-lim}_{t \searrow 0} p(t,\cdot), \qquad U(0) \subset \liminf_{t \searrow 0} U(t),$$

by 7.57 and 7.4(h). Similarly, in using $f^*(t,\cdot,\cdot)$ to denote the convex function conjugate to $f(t,\cdot,\cdot)$, we have the epi-continuity of $t \mapsto f^*(t,\cdot,\cdot)$ by Theorem 11.34 and thus, for the convex functions $q(t,\cdot)$ defined by $q(t,v) = \inf_y f^*(t,v,y)$ and the convex sets $V(t) = \operatorname{dom} q(t,\cdot)$ that

$$q(0,\cdot) = \operatorname*{e-lim}_{t \searrow 0} q(t,\cdot), \qquad V(0) \subset \liminf_{t \searrow 0} V(t).$$

Our assumption that $\operatorname{argminimax}_{X,Y} L(0,\cdot,\cdot)$ is nonempty and bounded corresponds by 11.40 to having $0 \in \operatorname{int} U(0)$ and $0 \in \operatorname{int} V(0)$. The inner limit inclusions for these sets imply then by 4.15 that also $0 \in \operatorname{int} U(t)$ and $0 \in \operatorname{int} V(t)$ for all t in some interval $[0,\varepsilon]$. Hence by 11.40 we have $\operatorname{argminimax} L(t,\cdot,\cdot)$ nonempty and bounded for t in this interval. We know further from Theorem 11.39, via 11.50 and 11.52, that

$$\operatorname{argminimax}_{X,Y} L(t,\cdot,\cdot) = \partial_u p(t,0) \times \partial_v q(t,0) \quad \text{for } t \in [0,\varepsilon].$$

The epi-convergence of $p(t,\cdot)$ to $p(0,\cdot)$ guarantees by the convexity of these functions that, as $t \searrow 0$, uniform convergence on a neighborhood of $u = 0$; cf. 7.17. Likewise, $q(t,\cdot)$ converges uniformly to $q(0,\cdot)$ around $v = 0$. The subgradient bounds in 9.14 imply that the mapping

$$t \mapsto \partial_u p(t,0) \times \partial_v q(t,0) \quad \text{on } [0,\varepsilon]$$

is osc and locally bounded at $t = 0$. And thus, the constant value $\lambda(t)$ that $L(t,\cdot,\cdot)$ has on $\partial_u p(t,0) \times \partial_v q(t,0)$ must converge as $t \searrow 0$ to the constant value $\lambda(0)$ that $L(0,\cdot,\cdot)$ has on $\partial_u p(0,0) \times \partial_v q(0,0)$.

The fact that the limit in 11(26) is taken over $x' \underset{X_0}{\to} x$ and $y' \underset{Y_0}{\to} y$ guarantees that $L'_+(0,x,y)$ is continuous relative to $(x,y) \in X_0 \times Y_0$. The convexity-concavity of $L'_+(0,\cdot,\cdot)$ on $X_0 \times Y_0$ follows immediately from that of $L(t,\cdot,\cdot)$ and the constancy of $L(0,\cdot,\cdot)$ on $X_0 \times Y_0$. Because the sets X_0 and Y_0 are closed and bounded, we know then from 11.52 that the saddle point set for $L'_+(0,\cdot,\cdot)$ on $X_0 \times Y_0$ is nonempty. Let $\mu = \operatorname{minimax}_{X_0,Y_0} L'_+(0,\cdot,\cdot)$. We have to demonstrate that

$$\lim_{t \searrow 0} \frac{1}{t}\big(\lambda(t) - \lambda(0)\big) = \mu.$$

Henceforth we can suppose for simplicity that $\varepsilon = T$ and write the set $\operatorname{argminimax}_{X,Y} L(t,\cdot,\cdot)$ as $X_t \times Y_t$; the mappings $t \mapsto X_t$ and $t \mapsto Y_t$ are nonempty-valued, and they are osc and locally bounded at $t = 0$. Consider any $(\bar{x},\bar{y}) \in X_0 \times Y_0$ and, for $t \in (0,T]$, pairs $(\bar{x}_t,\bar{y}_t) \in X_t \times Y_t$. As $t \searrow 0$, $(\bar{x}_t,\bar{y}_t)$

stays bounded, and any cluster point (x_0, y_0) belongs to $X_0 \times Y_0$. The saddle point conditions imply that

$$\begin{aligned} L(0, \bar{x}_t, \bar{y}) \geq L(0, \bar{x}, \bar{y}) \geq L(0, \bar{x}, \bar{y}_t), &\qquad L(0, \bar{x}, \bar{y}) = \lambda(0), \\ L(t, \bar{x}, \bar{y}_t) \geq L(t, \bar{x}_t, \bar{y}_t) \geq L(t, \bar{x}_t, \bar{y}), &\qquad L(t, \bar{x}_t, \bar{y}_t) = \lambda(t). \end{aligned}$$

Using these relations, we consider any sequence $t^\nu \searrow 0$ such that $\bar{y}_{t^\nu}$ converges to some y_0 and estimate that

$$\begin{aligned} \limsup_{\nu \to \infty} \frac{\lambda(t^\nu) - \lambda(0)}{t^\nu} &= \limsup_{\nu \to \infty} \frac{L(t^\nu, \bar{x}_{t^\nu}, \bar{y}_{t^\nu}) - L(0, \bar{x}, \bar{y})}{t^\nu} \\ &\leq \limsup_{\nu \to \infty} \frac{L(t^\nu, \bar{x}, \bar{y}_{t^\nu}) - L(0, \bar{x}, \bar{y}_{t^\nu})}{t^\nu} \\ &= \; L'_+(0, \bar{x}, y_0) \;\leq\; \sup_{y \in Y_0} L'_+(0, \bar{x}, y). \end{aligned}$$

Since $\bar{x}$ was an arbitrary point of X_0, this yields the bound

$$\limsup_{t \searrow 0} \frac{\lambda(t) - \lambda(0)}{t} \leq \inf_{x \in X_0} \left[\sup_{y \in Y_0} L'_+(0, x, y) \right] = \mu.$$

A parallel argument with the roles of x and y switched shows also that

$$\liminf_{t \searrow 0} \frac{\lambda(t) - \lambda(0)}{t} \geq \sup_{y \in Y_0} \left[\inf_{x \in X_0} L'_+(0, x, y) \right] = \mu.$$

Thus, $[\lambda(t) - \lambda(0)]/t$ does tend to μ as $t \searrow 0$. □

11.54 Example (perturbations in extended linear-quadratic programming)**.** *In the format and assumptions of 11.43, but with dependence additionally on a parameter t, consider for each $t \in [0, T]$ the primal and dual problems*

$$\begin{aligned} &\textit{minimize } \; \langle c(t), x \rangle + \tfrac{1}{2} \langle x, C(t) x \rangle + \theta_{Y, B(t)}\big(b(t) - A(t)x\big) \; \textit{ over } x \in X, \\ &\textit{maximize } \; \langle b(t), y \rangle - \tfrac{1}{2} \langle y, B(t) y \rangle - \theta_{X, C(t)}\big(A(t)^* y - c(t)\big) \; \textit{ over } y \in Y, \end{aligned}$$

denoting their optimal solution sets by X_t and Y_t. Suppose that X_0 and Y_0 are nonempty and bounded.

If $c(t)$, $C(t)$, $b(t)$, $B(t)$, and $A(t)$ depend continuously on t, there exists $\varepsilon > 0$ such that, relative to $t \in [0, \varepsilon]$, the mappings $t \mapsto X_t$ and $t \mapsto Y_t$ are nonempty-valued and, at $t = 0$, are osc and locally bounded. Then too, the common optimal value in the two problems, denoted by $\lambda(t)$, behaves continuously at $t = 0$. If in addition the right derivatives

$$c_0 := c'_+(0), \quad C_0 := C'_+(0), \quad b_0 := b'_+(0), \quad B_0 := B'_+(0), \quad A_0 := A'_+(0),$$

exist, then the right derivative $\lambda'_+(0)$ exists and is the common optimal value in the problems

$$\begin{aligned} &\textit{minimize}\ \ \langle c_0, x\rangle + \tfrac{1}{2}\langle x, C_0 x\rangle + \theta_{Y_0,B_0}(b_0 - A_0 x)\ \ \textit{over}\ x \in X_0,\\ &\textit{maximize}\ \ \langle b_0, y\rangle - \tfrac{1}{2}\langle y, B_0 y\rangle - \theta_{X_0,C_0}(A_0^* y - c_0)\ \ \textit{over}\ y \in Y_0. \end{aligned}$$

Detail. This specializes Theorem 11.53 to the Lagrangian setup for extended linear-quadratic programming in 11.47. □

Note that the matrices B_0 and C_0 in Example 11.54, although symmetric, might not be positive definite, so the subproblem giving the rate of change of the optimal value might not be one of extended linear-quadratic programming strictly as described in 11.43. But it's clear from the convexity-concavity of $L'_+(0,\cdot,\cdot)$ in Theorem 11.55 that the expression $\frac{1}{2}\langle x, C_0 x\rangle$ is convex *with respect to* $x \in X_0$, while $\frac{1}{2}\langle y, B_0 y\rangle$ is convex *with respect to* $y \in Y_0$. The primal and dual subproblems for $\lambda'_+(0)$ are thus of convex type nonetheless. (The notion of piecewise linear-quadratic programming can readily be refined in this direction. The special features of duality in Theorem 11.42 persist, and in 11.18 all that changes is the description of the effective domains of the θ functions.)

K*. Augmented Lagrangians and Nonconvex Duality

The question of the extent to which the duality relation $\inf \varphi = \sup \psi$ might hold for problems of nonconvex type, beyond the framework in 11.39, has its answer in a closer examination of the relationships in 11.38. The general situation is shown in Figure 11–6, where the notation is still that of $\varphi = f(\cdot, 0)$, $\psi = -f^*(0,\cdot)$, and $p(u) = \inf_x f(x,u)$. By the definition of p^{**}, the value $\sup \psi = p^{**}(0)$ is the supremum of the intercepts on the vertical axis that are achievable by affine functions majorized by p; the vectors $\bar{y} \in \operatorname{argmax} \psi$, if any, correspond to the affine functions that are 'highest' in this sense.

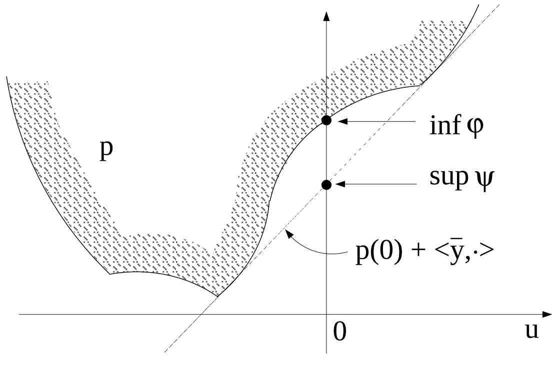

Fig. 11–6. Duality gap in minimization problems lacking adequate convexity.

A *duality gap* $\inf \varphi > \sup \psi$ arises precisely when the intercepts are prevented from getting as high as the value $\inf \varphi = p(0)$. A lack of convexity can evidently be the source of such a shortfall. (Of course, a duality gap can also occur even when p is convex if p fails to be lsc at 0 or if p takes on $-\infty$ but $0 \notin \mathrm{cl}(\mathrm{dom}\, p)$; cf. 11.2.) In particular, it's clear that

$$\left.\begin{array}{r}\inf \varphi = \sup \psi \\ \bar{y} \in \operatorname{argmax} \psi\end{array}\right\} \iff \left\{\begin{array}{l} p(u) \;\geq\; p(0) + \langle \bar{y}, u\rangle \;\text{ for all }\; u, \\ \text{with }\; p(0) \neq -\infty. \end{array}\right. \tag{11(27)}$$

This picture suggests that a duality gap might be avoided if the dual problem could be set up to correspond not to pushing *affine* functions up against $\mathrm{epi}\, p$, but some other class of functions capable of penetrating possible 'dents'. This turns out to be attainable with only a little extra effort.

11.55 Definition (augmented Lagrangian functions)**.** *For a primal problem of minimizing $\varphi(x)$ over $x \in \mathbb{R}^n$ and any dualizing parameterization $\varphi = f(\cdot, 0)$ for a choice of $f : \mathbb{R}^n \times \mathbb{R}^m \to \overline{\mathbb{R}}$, consider any augmenting function σ; by this is meant a proper, lsc, convex function*

$$\sigma : \mathbb{R}^m \to \overline{\mathbb{R}} \;\; \textit{with} \;\; \min \sigma = 0, \;\operatorname{argmin} \sigma = \{0\}.$$

The corresponding augmented Lagrangian with penalty parameter $r > 0$ is then the function $\bar{l} : \mathbb{R}^n \times \mathbb{R}^m \times (0, \infty) \to \overline{\mathbb{R}}$ defined by

$$\bar{l}(x, y, r) \;:=\; \inf_u \Big\{ f(x, u) + r\sigma(u) - \langle y, u\rangle \Big\}.$$

The corresponding augmented dual problem consists of maximizing over all $(y, r) \in \mathbb{R}^m \times (0, \infty)$ the function

$$\bar{\psi}(y, r) \;:=\; \inf_{x,u} \Big\{ f(x, u) + r\sigma(u) - \langle y, u\rangle \Big\}.$$

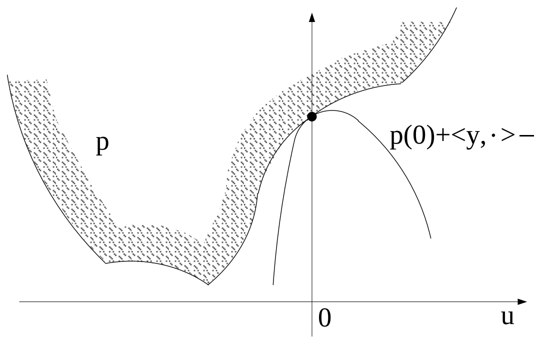

Fig. 11–7. Duality gap removed by an augmenting function.

The notion of the augmented Lagrangian grows from the idea of replacing the inequality in 11(27) by one of the form

$$p(u) \;\geq\; p(0) + \langle \bar{y}, u\rangle - r\sigma(u) \;\text{ for all }\; u$$

for some augmenting function σ (as just defined) and a parameter value $r > 0$ sufficiently high, as in Figure 11–7. What makes the approach successful in modifying the dual problem to get rid of the duality gap is that this inequality is identical to

$$p_{r\sigma}(u) \ \geq \ p_{r\sigma}(0) + \langle \bar{y}, u \rangle \ \text{ for all } \ u$$

where $p_{r\sigma}(u) := \inf_x f_{r\sigma}(x,u)$ for the function $f_{r\sigma}(x,u) := f(x,u) + r\sigma(u)$. Indeed, $p_{r\sigma}(u) = p(u) + r\sigma(u)$ and $p_{r\sigma}(0) = p(0)$, because $\sigma(0) = 0$.

We have $\varphi = f_{r\sigma}(\cdot, 0)$ as well as $\varphi = f(\cdot, 0)$. Moreover because σ is proper, lsc and convex, the representation of φ in terms of $f_{r\sigma}$, like that in terms of f, is a dualizing parameterization. The Lagrangian associated with $f_{r\sigma}$ is $l_{r\sigma}(x,y) = \bar{l}(x,y,r)$, as seen from the definition of $\bar{l}(x,y,r)$ above. The resulting dual problem, which consists of maximizing $\psi_{r\sigma} = -f^*_{r\sigma}(0,\cdot)$ over $y \in I\!R^m$, has $\psi_{r\sigma}(y) = \overline{\psi}(y,r)$. We can apply the theory already developed to this modified setting, where $f_{r\sigma}$ replaces f, and capture powerful new features.

Before translating this argument into a theorem about duality, we record some general consequences of Definition 11.55 for background, and we look at a couple of examples.

11.56 Exercise (properties of augmented Lagrangians). *For any dualizing parameterization f and augmenting function σ, the augmented Lagrangian $\bar{l}(x,y,r)$ is concave and usc in (y,r) and nondecreasing in r. It is convex in x if $f(x,u)$ is actually convex in (x,u). If σ is finite everywhere, the augmented Lagrangian is given in terms of the ordinary Lagrangian $l(x,y)$ by*

$$\bar{l}(x,y,r) \ = \ \sup_z \Big\{ l(x, y-z) - r\sigma^*(r^{-1}z) \Big\}. \qquad 11(28)$$

Likewise, the augmented dual expression $\overline{\psi}(y,r)$ is concave and usc in (y,r) and nondecreasing in r. In the case of $f(x,u)$ convex in (x,u) and σ finite everywhere, it is given in terms of the ordinary dual expression $\psi(y)$ by

$$\overline{\psi}(y,r) \ = \ \sup_z \Big\{ \psi(y-z) - r\sigma^*(r^{-1}z) \Big\}. \qquad 11(29)$$

Then in fact, $(\bar{y}, \bar{r})$ maximizes $\overline{\psi}$ if and only if $\bar{y}$ maximizes ψ; the value of $\bar{r} > 0$ can be chosen arbitrarily.

Guide. Derive the properties of $\bar{l}$ straight from the formula in 11.55, using 2.22(a) for the convexity in x. Develop 11(28) out of the fact that $-\bar{l}(x,\cdot,r)$ is conjugate to $f(x,\cdot) + r\sigma$, taking note of 11.23(a). Handle $\overline{\psi}$ similarly. The final assertion about maximizing pairs $(\bar{y}, \bar{r})$ falls out of 11(29). □

The monotonicity of $\bar{l}(x,y,r)$ and $\overline{\psi}(y,r)$ in r is consistent with r being a penalty parameter, an interpretation that will become clearer as we proceed. It lends special character to the augmented dual problem. In maximizing $\overline{\psi}(y,r)$ in y and r, the requirement that $r > 0$ doesn't act like a real constraint. There's no danger of having to move toward $r = 0$ to get higher values of $\overline{\psi}(y,r)$.

The fact at the end of 11.56, that, in the convex case, solutions to the augmented dual problem correspond simply to solutions to the ordinary dual problem, is reassuring. Augmentation doesn't destroy duality that may exist without it. This doesn't mean, though, that augmented Lagrangians have nothing to offer in the convex duality framework. Quite to the contrary, some of the main features of augmentation, for instance in achieving 'exact penalty representations' (as described below in 11.60), are most easily accessed in that special framework.

11.57 Example (proximal Lagrangians). *An augmented Lagrangian generated with the augmenting function $\sigma(u) = \frac{1}{2}|u|^2$ is a proximal Lagrangian. Then*

$$\begin{aligned}\bar{l}(x,y,r) &= \inf_u\Big\{f(x,u) + \frac{r}{2}|u|^2 - \langle y,u\rangle\Big\}\\ &= \sup_z\Big\{l(x,y-z) - \frac{1}{2r}|z|^2\Big\} = \sup_z\Big\{l(x,z) - \frac{1}{2r}|z-y|^2\Big\}.\end{aligned}$$

As an illustration, consider the case of Example 11.46 in which $\theta = \delta_D$ for a closed, convex, set $D \neq \emptyset$. This calculates out to

$$\bar{l}(x,y,r) = f_0(x) + \frac{r}{2}\Big[d_D\big(r^{-1}y + F(x)\big)^2 - |r^{-1}y|^2\Big],$$

which for $y = 0$ gives the standard quadratic penalty function for the constraint $F(x) \in D$. When $D = \{0\}$, one gets

$$\bar{l}(x,y,r) = f_0(x) + \big\langle y, F(x)\big\rangle + \frac{r}{2}\big|F(x)\big|^2.$$

Detail. These specializations are obtained from the first of the formulas for $\bar{l}(x,y,r)$ in writing $(r/2)|u|^2 - \langle y,u\rangle$ as $(r/2)\big(|u - r^{-1}y|^2 - |r^{-1}y|^2\big)$. □

11.58 Example (sharp Lagrangians). *An augmented Lagrangian generated with augmenting function $\sigma(u) = \|u\|$ (any norm $\|\cdot\|$) is a sharp Lagrangian. Then*

$$\begin{aligned}\bar{l}(x,y,r) &= \inf_u\Big\{f(x,u) + r\|u\| - \langle y,u\rangle\Big\}\\ &= \sup_z\Big\{l(x,y-z) - \delta_{rB^\circ}(z)\Big\} = \sup_{\|z-y\|^\circ \le r} l(x,z),\end{aligned}$$

where $\|\cdot\|^\circ$ is the polar norm and B° its unit ball. For Example 11.46 in the case of $\theta = \delta_D$ for a closed, convex set D, one gets

$$\bar{l}(x,y,r) = f_0(x) + \sup_{\|z-y\|^\circ \le r}\Big\{\big\langle z, F(x)\big\rangle - \sigma_D(z)\Big\},$$

which can be calculated out completely for instance when D is a box and $\|\cdot\| = \|\cdot\|_1$, so that $\|\cdot\|^\circ = \|\cdot\|_\infty$. Anyway, for $y = 0$ one has the standard linear penalty representation of the constraint $F(x) \in D$:

$$\bar{l}(x,0,r) = f_0(x) + r d_D\big(F(x)\big) \qquad \text{(distance in } \|\cdot\|\text{)}.$$

For general y but $D = \{0\}$ one obtains

$$\overline{l}(x,y,r) \;=\; f_0(x) + \big\langle y, F(x)\big\rangle + r\|F(x)\|.$$

Outfitted with these background facts and examples, we return now to the derivation of a duality theory in terms of augmented Lagrangians that is able even to cover certain nonconvex problems.

11.59 Theorem (duality without convexity)**.** *For a problem of minimizing φ on $I\!R^n$, consider the augmented Lagrangian $\overline{l}(x,y,r)$ associated with a dualizing parameterization $\varphi = f(\cdot,0)$, $f : I\!R^n \times I\!R^m \to \overline{I\!R}$, and some augmenting function $\sigma : I\!R^m \to \overline{I\!R}$. Suppose that $f(x,u)$ is level-bounded in x locally uniformly in u, and let $p(u) := \inf_x f(x,u)$. Suppose further that $\inf_x \overline{l}(x,y,r) > -\infty$ for at least one $(y,r) \in I\!R^m \times (0,\infty)$. Then*

$$\varphi(x) = \sup\nolimits_{y,r} \overline{l}(x,y,r), \qquad \overline{\psi}(y,r) = \inf\nolimits_x \overline{l}(x,y,r),$$

where actually $\varphi(x) = \sup_y \overline{l}(x,y,r)$ for every $r > 0$, and in fact

$$\begin{aligned} \inf\nolimits_x \varphi(x) \;&=\; \inf\nolimits_x\big[\sup\nolimits_{y,r} \overline{l}(x,y,r)\big] \\ &=\; \sup\nolimits_{y,r}\big[\inf\nolimits_x \overline{l}(x,y,r)\big] \;=\; \sup\nolimits_{y,r} \overline{\psi}(y,r). \end{aligned} \tag{11(30)}$$

Moreover, optimal solutions to the primal and augmented dual problems are characterized as saddle points of the augmented Lagrangian:

$$\begin{aligned} &\bar{x} \in \operatorname{argmin}_x \varphi(x) \;\; \text{and} \;\; (\bar{y},\bar{r}) \in \operatorname{argmax}_{y,r} \overline{\psi}(y,r) \\ &\iff \;\; \inf\nolimits_x \overline{l}(x,\bar{y},\bar{r}) \;=\; \overline{l}(\bar{x},\bar{y},\bar{r}) \;=\; \sup\nolimits_{y,r} \overline{l}(\bar{x},y,r), \end{aligned} \tag{11(31)}$$

the elements of $\operatorname{argmax}_{y,r} \overline{\psi}(y,r)$ being the pairs $(\bar{y},\bar{r})$ with the property that

$$p(u) \;\geq\; p(0) + \langle \bar{y}, u\rangle - \bar{r}\,\sigma(u) \;\; \text{for all} \;\; u. \tag{11(32)}$$

Proof. Most of this is evident from the monotonicity with respect to r in 11.56 and the explanation after Definition 11.55 of how 11(27) and Theorem 11.50 can be applied. The crucial new fact needing justification is the equality in the middle of 11(30), in place of merely the automatic '$\geq$'.

By hypothesis there's at least one pair $(\tilde{y},\tilde{r})$ such that $\overline{\psi}(\tilde{y},\tilde{r})$ is finite. To get the equality in question, it will suffice to demonstrate that $\overline{\psi}(\tilde{y},r) \to p(0)$ as $r \to \infty$, since $p(0) = \inf_x \varphi(x)$. We can utilize along the way the fact that p is proper and lsc by virtue of the level boundedness assumption placed on f (cf. 1.17). From the definition of $\overline{\psi}$ in 11.55 we have for all $r > 0$ that

$$\overline{\psi}(\tilde{y},r) \;=\; \inf\nolimits_u\big\{p(u) + r\sigma(u) - \langle \tilde{y}, u\rangle\big\}.$$

Let $\tilde{p}(u) := p(u) + \tilde{r}\sigma(u) - \langle \tilde{y}, u\rangle$, noting that $\tilde{p}(0) = p(0)$ because $\sigma(0) = 0$. This function is bounded below by $\overline{\psi}(\tilde{y},\tilde{r})$, and like p it is proper and lsc because σ is proper and lsc. We can write

$$\overline{\psi}(\tilde{y},\tilde{r}+s) \;=\; \inf\nolimits_u\big\{\tilde{p}(u) + s\sigma(u)\big\} \;\; \text{for} \;\; s > 0$$

and concentrate on proving that the limit of this as $s \to \infty$ is $\tilde{p}(0)$. Because σ is convex with $\operatorname{argmin} \sigma = \{0\}$, it's level-coercive (by 3.27), and so too then is $\tilde{p} + s\sigma$, due to $\tilde{p}$ being bounded below. The positivity of σ away from 0 guarantees that $\tilde{p} + s\sigma$ increases pointwise as $s \to \infty$ to the function $\delta_{\{0\}} + \gamma$ for the constant $\gamma = \tilde{p}(0)$. In particular then $\tilde{p} + s\sigma$ epi-converges to $\delta_{\{0\}} + \gamma$ in the setting of Theorem 7.33 (see also 7.4(f)), and we are able to conclude that $\inf(\tilde{p} + s\sigma) \to \inf(\delta_{\{0\}} + \gamma) = \gamma$. This was what we needed. □

The importance of the solutions to the augmented dual problem is found in the following idea.

11.60 Definition (exact penalty representations). *In the augmented Lagrangian framework of 11.55, a vector $\bar{y}$ is said to support an exact penalty representation for the problem of minimizing φ on $I\!R^n$ if, for all $r > 0$ sufficiently large, this problem is equivalent to minimizing $\bar{l}(\cdot, \bar{y}, r)$ on $I\!R^n$ in the sense that*

$$\inf_x \varphi(x) = \inf_x \bar{l}(x, \bar{y}, r), \qquad \operatorname{argmin}_x \varphi(x) = \operatorname{argmin}_x \bar{l}(x, \bar{y}, r).$$

Specifically, a value $\bar{r} > 0$ is said to serve as an adequate penalty threshold in this respect if the property holds for all $r \in (\bar{r}, \infty)$.

11.61 Theorem (criterion for exact penalty representations). *In the notation and assumptions of Theorem 11.59, a vector $\bar{y}$ supports an exact penalty representation for the primal problem if and only if there exist $W \in \mathcal{N}(0)$ and $\hat{r} > 0$ such that*

$$p(u) \;\geq\; p(0) + \langle \bar{y}, u \rangle - \hat{r}\sigma(u) \;\; \text{for all} \;\; u \in W. \qquad 11(33)$$

This criterion is equivalent in turn to the existence of an $\bar{r} > 0$ with $(\bar{y}, \bar{r}) \in \operatorname{argmax}_{y,r} \overline{\psi}(y, r)$, and moreover such values $\bar{r}$ are the ones serving as adequate penalty thresholds for the exact penalty property with respect to $\bar{y}$.

Proof. As a starter, note that the assumptions of Theorem 11.59 guarantee that p is lsc and proper and hence that the condition in 11(33), and for that matter the one in 11(32), can't hold unless the value $p(0) = \inf \varphi$ is finite.

First we argue that the condition $(\bar{y}, \bar{r}) \in \operatorname{argmax}_{y,r} \overline{\psi}(y, r)$ is both necessary and sufficient for $\bar{y}$ to support an exact penalty representation with $\bar{r}$ as an adequate penalty threshold. For the necessity, note that the latter conditions imply through Definition 11.60 that $\overline{\psi}(\bar{y}, r) = \inf \varphi$ for all $r \in (\bar{r}, \infty)$ and hence, because $\overline{\psi}$ is usc (by 11.56), that $\overline{\psi}(\bar{y}, \bar{r}) \geq \inf \varphi$. Since $\sup \overline{\psi} = \inf \varphi$ by 11(31), it follows that $(\bar{y}, \bar{r})$ maximizes $\overline{\psi}$.

For the sufficiency, recall from the end of Theorem 11.59 that the condition $(\bar{y}, \bar{r}) \in \operatorname{argmax}_{y,r} \overline{\psi}(y, r)$ corresponds to the inequality 11(32) holding. When $\bar{r}$ is replaced in 11(32) by any higher value r, this inequality becomes strict for $u \neq 0$, because $\sigma(0) = 0$ but $\sigma(u) > 0$ for $u \neq 0$. Then

$$\operatorname{argmin}_u \Big\{ p(u) + r\sigma(u) - \langle \bar{y}, u \rangle \Big\} \;=\; \{0\}.$$

Fixing such $r > \bar{r}$, consider the function $g(x,u) := f(x,u) + r\sigma(u) - \langle \bar{y}, u\rangle$ and its associated inf-projections $h(u) := \inf_x g(x,u)$ and $k(x) := \inf_u g(x,u)$, noting that $h(u) = p(u) + r\sigma(u) - \langle \bar{y}, u\rangle$ whereas $k(x) = \overline{l}(x, \bar{y}, r)$. According to the interchange rule in 1.35, one has

$$\left.\begin{aligned} \bar{u} \in \operatorname{argmin}_u h(u) \\ \bar{x} \in \operatorname{argmin}_x g(x,\bar{u}) \end{aligned}\right\} \iff \left\{\begin{aligned} &\bar{x} \in \operatorname{argmin}_x k(x) \\ &\bar{u} \in \operatorname{argmin}_u g(\bar{x}, u) \end{aligned}\right.$$

We've just seen that $\operatorname{argmin}_u h(u) = \{0\}$. The pairs $(\bar{x}, \bar{u})$ on the left of this rule are thus the ones with $\bar{u} = 0$ and $\bar{x} \in \operatorname{argmin}_x g(x, 0) = \operatorname{argmin}_x \varphi(x)$. These are then also the pairs on the right; in other words, in minimizing $\overline{l}(x, \bar{y}, r)$ in x one obtains exactly the points $\bar{x} \in \operatorname{argmin}_x \varphi(x)$, and then in maximizing for any such $\bar{x}$ the expression $f(\bar{x}, u) + r\sigma(u) - \langle \bar{y}, u\rangle$ in u one finds that the maximum is attained uniquely at 0. In particular, the sets $\operatorname{argmin}_x \varphi(x)$ and $\operatorname{argmin}_x \overline{l}(x, \bar{y}, r)$ are the same.

To complete the proof of the theorem we need only show now that when 11(33) holds there must exist $\bar{r} \in (\hat{r}, \infty)$ such that the stronger condition 11(32) holds. In assuming 11(33) there's no loss of generality in taking W to be a ball $\varepsilon I\!B$, $\varepsilon > 0$. Obviously 11(33) continues to hold when $\hat{r}$ is replaced by a higher value $\bar{r}$, so the question is whether, once $\bar{r}$ is high enough, we will have, in addition, the inequality in 11(32) holding for all $|u| > \varepsilon$. By hypothesis (inherited from Theorem 11.59) there exists $(\tilde{y}, \tilde{r}) \in I\!R^m \times (0, \infty)$ such that $\overline{\psi}(\tilde{y}, \tilde{r})$ is finite; then for $\alpha := \overline{\psi}(\tilde{y}, \tilde{r})$ we have

$$p(u) \;\geq\; \alpha + \langle \tilde{y}, u\rangle - \tilde{r}\sigma(u) \;\text{ for all }\; u.$$

It will suffice to show that, when $\bar{r}$ is chosen high enough, one will have

$$\alpha + \langle \tilde{y}, u\rangle - \tilde{r}\sigma(u) > p(0) + \langle \bar{y}, u\rangle - \bar{r}\sigma(u) \;\text{ when }\; |u| > \varepsilon.$$

This amounts to showing that for high enough values of $\bar{r}$ one will have

$$\big\{ u \,\big|\, (\bar{r} - \tilde{r})\sigma(u) \leq \langle \bar{y} - \tilde{y}, u\rangle + p(0) - \alpha \big\} \;\subset\; \varepsilon I\!B.$$

The issue can be simplified by working with $s = \bar{r} - \tilde{r} > 0$ and letting $\lambda := |\bar{y} - \tilde{y}|$ and $\mu := p(0) - \alpha$. Then we only need to check that the set

$$C(s) \;:=\; \big\{ u \,\big|\, s\sigma(u) \leq \lambda |u| + \mu \big\}$$

lies in $\varepsilon I\!B$ when s is chosen high enough.

We know σ is level-coercive, because $\operatorname{argmin} \sigma = \{0\}$ (cf. 3.23, 3.27), hence there exist $\gamma > 0$ and β such that $\sigma(u) \geq \gamma|u| + \beta$ for all u (cf. 3.26). Let $s_0 > 0$ be high enough that $s_0\gamma - \lambda > 0$. For $u \in C(s_0)$ we have $s_0(\gamma|u| + \beta) \leq \lambda|u| + \mu$, hence $|u| \leq \rho := (\mu - s_0\beta)/(s_0\gamma - \lambda)$. But $C(s) \subset C(s_0)$ when $s > s_0$, inasmuch as $\sigma(u) \geq 0$ for all u. Therefore

$$C(s) \subset \big\{ u \,\big|\, \sigma(u) \leq (\lambda\rho + \mu)/s \big\}$$

when $s > s_0$. On the other hand, because $\sigma(u) = 0$ only for $u = 0$, the level set $\{u \mid \sigma(u) \leq \delta\}$ must lie in $\varepsilon\mathbb{B}$ for small $\delta > 0$. Taking such a δ and noting that $(\lambda\rho + \mu)/s \leq \delta$ when s exceeds a certain s_1, we conclude that $C(s) \subset \varepsilon\mathbb{B}$ when $s > s_1$. □

11.62 Example (exactness of linear or quadratic penalties).

(a) *Consider the proximal Lagrangian of 11.57 under the assumption that* $\inf_x l(x, y, r) > -\infty$ *for at least one choice of* (y, r). *Then a necessary and sufficient condition for a vector* $\bar{y}$ *to support an exact penalty representation is that* $\bar{y}$ *be a proximal subgradient of the function* $p(u) = \inf_x f(x, u)$ *at* $u = 0$.

(b) *Consider the sharp Lagrangian of 11.58 under the assumption that* $\inf_x l(x, 0, r) > -\infty$ *for some* r. *Then a necessary and sufficient condition for the vector* $\bar{y} = 0$ *to support an exact penalty representation is that the function* $p(u) = \inf_x f(x, u)$ *be calm from below at* $u = 0$.

Detail. These results specialize Theorem 11.61. Relation 11(33) corresponds in (a) to the definition of a proximal subgradient (see 8.45), whereas in (b) it means calmness from below (see the material around 8.32). □

L.* Generalized Conjugacy

The notion of conjugate functions can be generalized in a number of ways, although none achieves the full power of the Legendre-Fenchel transform. Consider *any* nonempty sets X and Y and *any* function $\Phi : X \times Y \to \overline{\mathbb{R}}$. (The 'ordinary case', corresponding to what we have been involved with until now, has $X = \mathbb{R}^n$, $Y = \mathbb{R}^n$, and $\Phi(x, y) = \langle x, y\rangle$.) For any function $f : X \to \overline{\mathbb{R}}$, the *$\Phi$-conjugate* of f on Y is the function

$$f^\Phi(y) := \sup_{x \in X} \{\Phi(x, y) - f(x)\},$$

while the *Φ-biconjugate* of f back on X is the function

$$f^{\Phi\Phi}(x) := \sup_{y \in Y} \{\Phi(x, y) - f^\Phi(y)\}.$$

Likewise, for any function $g : Y \to \overline{\mathbb{R}}$, the Φ-conjugate of g on X is the function

$$g^\Phi(x) := \sup_{y \in Y} \{\Phi(x, y) - g(y)\},$$

while the Φ-biconjugate of g back on Y is the function

$$g^{\Phi\Phi}(y) := \sup_{x \in X} \{\Phi(x, y) - g^\Phi(x)\}.$$

Define the *Φ-envelope* functions on X to be the functions expressible as the pointwise supremum of some collection of functions of the form $\Phi(\cdot, y)$+constant for various choices of $y \in Y$, and define the Φ-envelope functions on Y analogously. (In the 'ordinary case', the proper Φ-envelope functions are the proper, lsc, convex functions.) Note that in circumstances where $X = Y$ but $\Phi(x, y)$ isn't symmetric in the arguments x and y, two different kinds of Φ-envelope functions might have to be distinguished on the same space.

11.63 Exercise (generalized conjugate functions). *In the scheme of Φ-conjugacy, for any function $f : X \to \overline{I\!R}$, f^{Φ} is a Φ-envelope function on Y, while $f^{\Phi\Phi}$ is the greatest Φ-envelope function on X majorized by f. Similarly, for any function $g : Y \to \overline{I\!R}$, g^{Φ} is a Φ-envelope function on X, while $g^{\Phi\Phi}$ is the greatest Φ-envelope function on Y majorized by g.*

Thus, Φ-conjugacy sets up a one-to-one correspondence between all the Φ-envelope functions f on X and all the Φ-envelope functions g on Y, with

$$g(y) = \sup_{x \in X} \{\Phi(x, y) - f(x)\}, \qquad f(x) = \sup_{y \in Y} \{\Phi(x, y) - g(y)\}.$$

Guide. Derive all this right from the definitions by elementary reasoning. □

11.64 Example (proximal transform). *For fixed $\lambda > 0$, pair $I\!R^n$ with itself under*

$$\Phi(x, y) = -\frac{1}{2\lambda}|x - y|^2 = \frac{1}{\lambda}\langle x, y\rangle - \frac{1}{2\lambda}|x|^2 - \frac{1}{2\lambda}|y|^2.$$

Then for any function $f : I\!R^n \to \overline{I\!R}$ and its Moreau envelope $e_\lambda f$ and λ-proximal hull $h_\lambda f$ one has

$$f^{\Phi} = -e_\lambda f, \qquad f^{\Phi\Phi}(x) = -e_\lambda[-e_\lambda f] = h_\lambda f.$$

In this case the Φ-envelope functions are the λ-proximal functions, defined as agreeing with their λ-proximal hull.

A one-to-one correspondence $f \leftrightarrow g$ in the collection of all proper functions of such type is obtained in which

$$g = -e_\lambda f, \qquad f = -e_\lambda g.$$

Detail. This follows from the definitions of f^{Φ} and $f^{\Phi\Phi}$ along with those of $e_\lambda f$ in 1.22 and $h_\lambda f$; see 1.44 and also 11.26(c). The symmetry of Φ in its two arguments yields the symmetry of the correspondence that is obtained. □

For the next example we recall from 2(13) the notation $I\!R^{n\times n}_{\rm sym}$ for the space of all symmetric real matrices of order n.

11.65 Example (full quadratic transform). *Pair $I\!R^n$ with $I\!R^n \times I\!R^{n\times n}_{\rm sym}$ under*

$$\Phi(x, y) = \langle v, x\rangle - j_Q(x) \;\text{ for }\; y = (v, Q), \;\text{ with }\; j_Q(x) = \frac{1}{2}\langle x, Qx\rangle.$$

In this case one has for any function $f : I\!R^n \to \overline{I\!R}$ that

$$f^{\Phi}(v,Q) = (f + j_Q)^*(v),$$
$$f^{\Phi\Phi}(x) = \begin{cases} (\mathrm{cl}\, f)(x) & \text{if } f \text{ is prox-bounded,} \\ -\infty & \text{otherwise.} \end{cases}$$

Here $f^{\Phi} \equiv -\infty$ if $f \equiv \infty$, whereas $f^{\Phi} \equiv \infty$ if f fails to be prox-bounded; otherwise f^{Φ} is a proper, lsc, convex function on the space $\mathbb{R}^n \times \mathbb{R}^{n\times n}_{\mathrm{sym}}$.

Thus, Φ-conjugacy of this kind sets up a one-to-one correspondence $f \leftrightarrow g$ between the proper, lsc, prox-bounded functions f on $\mathbb{R}^n$ and certain proper, lsc, convex functions g on $\mathbb{R}^n \times \mathbb{R}^{n\times n}_{\mathrm{sym}}$.

Detail. The formula for f^{Φ} is immediate from the definitions, and the one for $f^{\Phi\Phi}$ follows then from 1.44. The convexity of $f^{\Phi\Phi}$ comes from the linearity of $\Phi(x,y)$ with respect to the y argument for fixed x. □

11.66 Example (basic quadratic transform)**.** *Pair $\mathbb{R}^n$ with $\mathbb{R}^n \times \mathbb{R}$ under*

$$\Phi(x,y) \;=\; \langle v, x\rangle - rj(x) \;\; \text{for} \;\; y = (v,r), \;\; \text{with} \;\; j(x) = \tfrac{1}{2}|x|^2.$$

In this case one has for any function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ that

$$f^{\Phi}(v,r) = (f + rj)^*(v), \quad f^{\Phi\Phi}(x) = \begin{cases} (\mathrm{cl}\, f)(x) & \text{if } f \text{ is prox-bounded,} \\ -\infty & \text{otherwise.} \end{cases}$$

Here $f^{\Phi} \equiv -\infty$ if $f \equiv \infty$, whereas $f^{\Phi} \equiv \infty$ if f fails to be prox-bounded; aside from those extreme cases, f^{Φ} is a proper, lsc, convex function on $\mathbb{R}^n \times \mathbb{R}$. Thus, Φ-conjugacy of this kind sets up a one-to-one correspondence $f \leftrightarrow g$ between the proper, lsc, prox-bounded functions f on $\mathbb{R}^n$ and certain proper, lsc, convex functions g on $\mathbb{R}^n \times \mathbb{R}$.

Detail. This restricts the conjugate function in the preceding example to the subspace of $\mathbb{R}^{n\times n}_{\mathrm{sym}}$ consisting of the matrices of form $Q = rI$. The biconjugate function is unaffected by this restriction because the derivation of its formula only relied on such special matrices, through 1.44. □

Yet another way that problems of optimization can be meaningfully be paired with one another is through the interchange rule for minimization in 1.35. In this framework the duality relations take the form of 'min = min' instead of 'min = max'. The idea is very simple. For comparison with the duality schemes in this chapter, it can be formulated as follows.

Given arbitrary nonempty sets X and Y and any function $l : X \times Y \mapsto \overline{\mathbb{R}}$, consider the problems,

$$\text{minimize } \varphi(x) \text{ over } x \in X, \text{ where } \varphi(x) := \inf\nolimits_y l(x,y),$$

$$\text{minimize } \psi(y) \text{ over } y \in Y, \text{ where } \psi(y) := \inf\nolimits_x l(x,y).$$

It's obvious that $\inf_X \varphi = \inf_Y \psi$; both values coincide with $\inf_{X\times Y} l$.

In contrast to the duality theory of convex optimization in 11.39, the two problems in this scheme aren't related to each other through perturbations, and the solutions to one problem can't be interpreted as 'multipliers' for the other. Rather, the two problems represent intermediate stages in an overall problem, namely that of minimizing $l(x,y)$ with respect to $(x,y) \in X \times Y$. Nonetheless, the scheme can be interesting in situations where a way exists for passing directly from one problem to the other without writing down the function l, especially if a correspondence between more than optimal values emerges. This is seen in the case of a Fenchel-type format resembling 11.41, where however the functions φ and ψ being minimized generally aren't convex.

11.67 Theorem (double-min duality in optimization). *Consider any primal problem of the form*

$$\textit{minimize } \varphi(x) \textit{ over } x \in \mathbb{R}^n, \textit{ with } \varphi(x) = \langle c, x\rangle + k(x) - h(Ax - b),$$

for convex functions k on $\mathbb{R}^n$ and h on $\mathbb{R}^m$ such that k is proper, lsc, fully coercive and almost strictly convex, while h is finite and differentiable. Pair this with the dual problem

$$\textit{minimize } \psi(y) \textit{ over } y \in \mathbb{R}^m, \textit{ with } \psi(y) = \langle b, y\rangle + h^*(y) - k^*(A^*y - c);$$

this similarly has h^* lsc, proper, fully coercive and almost strictly convex, while k^* is finite and differentiable. The objective functions φ and ψ in these two problems are amenable, and the two optimal values always agree: $\inf \varphi = \inf \psi$. *Furthermore, in terms of*

$$\begin{aligned} S := &\big\{(x,y) \,\big|\, y = \nabla h(Ax - b),\ A^*y - c \in \partial k(x)\big\} \\ = &\big\{(x,y) \,\big|\, x = \nabla k^*(A^*y - c),\ Ax - b \in \partial h^*(y)\big\}, \end{aligned}$$

one has the following relations, which tie not only optimal solutions but also generalized stationary points of either problem to those of the other problem:

$$\begin{aligned} 0 \in \partial\varphi(\bar{x}) \quad &\Longleftrightarrow \quad \exists\, \bar{y} \textit{ with } (\bar{x},\bar{y}) \in S, \\ 0 \in \partial\psi(\bar{y}) \quad &\Longleftrightarrow \quad \exists\, \bar{x} \textit{ with } (\bar{x},\bar{y}) \in S, \\ (\bar{x},\bar{y}) \in S \quad &\Longrightarrow \quad \varphi(\bar{x}) = \psi(\bar{y}). \end{aligned}$$

Proof. In the general format described before the theorem, these problems correspond to $l(x,y) = \langle c, x\rangle + k(x) + \langle b, y\rangle + h^*(y) - \langle y, Ax\rangle$.

The claim that h^* is fully coercive and almost strictly convex is justified by 11.5 and 3.27 for the coercivity and 11.13 for the strict convexity. The same results, in reverse implication, support the claim that k^* is finite and differentiable. Because convex functions that are finite and differentiable actually are $\mathcal{C}^1$ (by 9.20), the functions $h_0(x) = -h(Ax - b)$ and $k_0(y) = -k^*(A^*y - c)$ are $\mathcal{C}^1$, and consequently φ and ψ are amenable by 10.24(g). Then also

$$\begin{aligned} \partial\varphi(x) = \partial k(x) + \nabla h_0(x) \ \text{ with } \ \nabla h_0(x) = -A^*\nabla h(Ax - b), \\ \partial\psi(y) = \partial h^*(y) + \nabla k_0(x) \ \text{ with } \ \nabla k_0(x) = -A\nabla k^*(A^*y - c), \end{aligned}$$

from which the characterizations of $0 \in \partial\varphi(\bar{x})$ and $0 \in \partial\psi(\bar{y})$ are evident, the equivalence between the two expressions for S being a consequence of the subgradient inversion rule in 11.3. Pairs $(\bar{x}, \bar{y}) \in S$ satisfy $h(A\bar{x} - b) + h^*(\bar{y}) = \langle A\bar{x} - b, \bar{y}\rangle$ and $k(\bar{x}) + k^*(A^*\bar{y} - c) = \langle \bar{x}, A^*\bar{y} - c\rangle$ (again on the basis of 11.3), and these equations give $\varphi(\bar{x}) = \psi(\bar{y})$. $\square$

A case worth noting in 11.67 is the one in which $k(x) = \delta_X(x) + \frac{1}{2}\langle x, Cx\rangle$ and $h(u) = \theta_{Y,B}(u)$ (cf. 11.18) for polyhedral sets X and Y and symmetric *positive-definite* matrices C and B. Then in minimizing φ one is minimizing the smooth piecewise linear-quadratic function

$$\varphi_0(x) = \langle c, x\rangle + \tfrac{1}{2}\langle x, Cx\rangle - \theta_{Y,B}(Ax - b)$$

subject to the linear constraints represented by $x \in X$, whereas in minimizing ψ one is minimizing the smooth piecewise linear-quadratic function

$$\psi_0(y) = \langle b, y\rangle + \tfrac{1}{2}\langle y, By\rangle - \theta_{X,C}(A^*y - c)$$

subject to the linear constraints represented by $y \in Y$. The functions φ_0 and ψ_0 needn't be convex, because the θ-expressions are subtracted rather than added as they were in 11.43, so this is a form of *nonconvex* extended linear-quadratic programming duality.

Commentary

A description of the 'Legendre transform' can be found in any text on the calculus of variations, since it's the means of generating the Hamiltonian functions and Hamiltonian equations that are crucial to that subject. The treatments in such texts often fall short in rigor, however. They revolve around inverting a gradient mapping ∇f as in 11.9, but typically without serious attention being paid to the mapping's domain and range, or to the conditions needed to ensure its *global* single-valued invertibility, such as (for most cases in practice) the strict convexity of f. The beauty of the Legendre-Fenchel transform, devised by Fenchel [1949], [1951], is that gradient inversion is replaced by an operation of maximization. Moreover, convexity properties are embraced from the start. In this way, a much more powerful tool is created which, interestingly, is perhaps the first in mathematical analysis to have relied on minimization/maximization in its very definition.

Mandelbrojt [1939] had earlier developed a limited case of conjugacy for functions of a single real variable. In the still more special context of nondecreasing convex functions on $\mathbb{R}_+$, similar notions were explored by Young [1912] and utilized in the theory of Banach spaces and beyond; cf. Birnbaum and Orlicz [1931] and Krasnosel'skii and Rutitskii [1961]. None of this captured the n-dimensional character of the Legendre transform, however, or allowed for the kind of focus on domains that's essential to handling *constraints* in the process of dualization.

Fenchel formulated the basic result in Theorem 11.1 in terms of pairs (C, f) consisting of a finite convex function f on a nonempty convex set C in $\mathbb{R}^n$. His transform was extended to infinite-dimensional spaces by Moreau [1962] and Brøndsted [1964] (publication of a dissertation written under Fenchel's supervision), with Moreau adopting the pattern of extended-real-valued functions f defined on all the whole space. The details of the Legendre case in 11.9 were worked out by Rockafellar [1967d]. All these authors were, in some way, aware of the facts in 11.2–11.4.

Rockafellar [1966b] discovered the support function meaning of horizon functions in 11.5 as well as the formula in 11.6 for the support function of a level set and the dualizations of coercivity and level-coercivity in 11.8(c)(d). The dualizations of differentiability in 11.8(b) and 11.13 were developed in Rockafellar [1970a], which is also the source for the connection in 11.7 between conjugate functions and cone polarity (probably known to Fenchel), and for the linear-quadratic examples in 11.10 and 11.11 and the log-exponential example in 11.12.

In the important case of convex functions on $I\!R^1$, special techniques can be used for determining the conjugate functions, for instance by 'integrating' right or left derivatives; for a thorough treatment see Chapter 8 of Rockafellar [1984b]. The one-dimensional version of Theorem 11.14, that f^* inherits from f the property of being piecewise linear, or piecewise linear-quadratic, can be found there as well.

In the full, n-dimensional version of Theorem 11.14, part (a) is new in its statement about the preservation of piecewise linearity when passing from a convex function to its conjugate, but this property corresponds through 2.49 to the fact that if $\operatorname{epi} f$ is polyhedral, the same is true also of $\operatorname{epi} f^*$. In that form, the observation goes back to Rockafellar [1963]. The assertion in 11.17(a) about the support functions of polyhedral sets being piecewise linear has similar status. The fact in 11.17(b), that the polar of a polyhedral cone is polyhedral, was recognized by Weyl [1935].

As for part (b) of Theorem 11.14, concerning the self-duality of the class of piecewise linear-quadratic functions, this was proved by Sun [1986], but with machinery from the theory of monotone mappings. Without the availability of that machinery in this chapter (it won't be set up until Chapter 12), we had to invent a different, more direct proof. The line segment test in 11.15, on which it rests, doesn't seem to have been formulated previously.

The consequent fact in 11.16, about piecewise linear-quadratic convex functions $f : I\!R^n \to \overline{I\!R}$ attaining their minimum value when that value is finite, can be compared to the result of Frank and Wolfe [1956] that a linear-quadratic convex function f_0 achieves its minimum relative to any polyhedral convex set C on which it's bounded from below. The latter can be identified with the case of 11.16 where $f = f_0 + \delta_C$.

The general polarity correspondence in 11.19, for sets that contain the origin but aren't necessarily cones, was developed first for bounded sets by Minkowski [1911], whose insights extended to the dualizations in 11.20. The formula for conjugate composite functions in 11.21 comes from Rockafellar [1970a].

The dual operations in 11.22 and 11.23 were investigated in various degrees by Fenchel [1951], Rockafellar [1963] and Moreau [1967]. The special cases in 11.24 for support functions and in 11.25 for cones were known much earlier. Moreau [1965] established the dual envelope formula in 11.26(b) (for $\lambda = 1$) and also the gradient interpretation in 11.27 for the proximal mappings associated with convex functions.

The results on the norms of sublinear mappings in 11.29 and 11.30 are new, but the adjoint duality for operations on such mappings in 11.31 was already disclosed in Rockafellar [1970a]. The special results in 11.32 and 11.33 on duality for operations on piecewise linear-quadratic functions are original as well.

The epi-continuity of the Legendre-Fenchel transform in Theorem 11.34 was discovered by Wijsman [1964], [1966], who also reported the corresponding behavior of support functions and polar cones in 11.35. The 'epi-semicontinuity' results in

Theorem 11.34, which provide inequalities for dualizing e-liminf and e-limsup, are actually new, and likewise for the support function version in 11.35(a). The inner and outer limit relations for cone polarity in 11.35(b) were noted by Walkup and Wets [1967], however. Those authors, in the same paper, established the polar cone isometry in Theorem 11.36, moreover not just for $\mathbb{R}^n$ but any reflexive Banach space. The proof of this isometry that we give here in terms of the duality between addition and epi-addition of convex functions is new. It readily generalizes to any norm and its polar, not only the Euclidean norm and unit ball, in setting up cone distances, and it too could therefore be followed in the Banach space setting.

The application in 11.37, showing that the Legendre-Fenchel transform is an isometry with respect to cosmic epi-distances, is new, but an isometry result with respect to a different metric, based instead on uniform convergence of Moreau envelopes on bounded sets, was obtained by Attouch and Wets [1986]. The technique of passing through cones and the continuity of the polar operation in 11.36 for purposes of investigating duality in the epi-convergence of convex functions (apart from questions of isometry), and thereby developing extensions of Wijsman's theorem, began with Wets [1980]. That approach has been followed more recently in infinite dimensions by Penot [1991] and Beer [1993]. Other epi-continuity results for the Legendre-Fenchel transform in spaces beyond $\mathbb{R}^n$ can be found in those works and in earlier papers of Mosco [1971], Joly [1973], Back [1986] and Beer [1990].

Many researchers have been fascinated by dual problems of optimization. The most significant early example was linear programming duality (cf. 11.43), which was laid out by Gale, Kuhn and Tucker [1951]. This duality grew out of the theory of minimax problems and two-person, zero-sum games that was initiated by von Neumann [1928]. Fenchel [1951], while visiting at Princeton where Gale and Kuhn were Ph.D. students of Tucker, sought to set up a parallel theory of dual problems in which the primal problem consisted of minimizing $h(x) + k(x)$, for what we now call lsc, proper, convex functions h and k on $\mathbb{R}^n$, while the dual problem consisted of maximizing $-h^*(y) - k^*(-y)$. Fenchel's duality theorem suffered from an error, however. Rockafellar [1963], [1964a], [1966c], [1967a], fixed the error and incorporated a linear mapping into the problem statement so as to obtain the scheme in 11.41. Perturbations played a role in that duality, but the scheme in 11.39 and 11.40, explicitly built around primal and dual perturbations, didn't emerge until Rockafellar [1970a].

Extended linear-quadratic programming was developed by Rockafellar and Wets [1986] along with the duality results in 11.43; further details were added by Rockafellar [1987]. It was in those papers that the dualizing penalty expressions $\theta_{Y,B}$ in 11.18 made their debut. The application to dualized composition in 11.44 is new.

The Lagrangian perturbation format in 11.45 came out in Rockafellar [1970a], [1974a], for the case of $f(x, u)$ convex in (x, u). The associated minimax theory in 11.50–11.52 was developed there also. The extension of Lagrangian dualization to the case of $f(x, u)$ convex merely in u started with Rockafellar [1993a], and that's where the multiplier rule in 11.46 was proved.

The formula in 11.48, relating the subgradients of $f(x, u)$ to those of $l(x, y)$ for convex f, can be interpreted as a generalization of the inversion rule in 11.3 through the fact that the functions $f(x, \cdot)$ and $-l(x, \cdot)$ are conjugate to each other. Instead of full inversion one has a partial inversion along with a change of sign in the residual argument. A similar partial inversion rule is known classically for smooth f with respect to taking the Legendre transform in the u argument, this being essential to the theory of Hamiltonian equations in the 'calculus of variations'. One could ask whether a broader formula of such type can be established for nonsmooth functions f that might only be convex in u, not necessarily in x and u jointly. Rockafellar [1993b], [1996], has answered this in the affirmative for particular function structures and more generally in terms of a partial convexification operation on subgradients.

The results on perturbing saddle values and saddle points in Theorem 11.53 are largely due to Golshtein [1972]. The linear programming case in Example 11.54 was treated earlier by Williams [1970]. The application here to linear-quadratic programming is new. For some nonconvex extensions of such perturbation theory that instead utilize augmented Lagrangians, see Rockafellar [1984a].

Augmented Lagrangians were introduced in connection with numerical methods for solving problems in nonlinear programming, and they have mainly been viewed in that special context; see Rockafellar [1974b], [1993a], Bertsekas [1982], and Golshtein and Tretyakov [1996]. They haven't been considered before with the degree of generality in 11.55–11.61. Exact penalty representations of the linear type in 11.62(b) have a separate history; see Burke [1991] for this background.

The generalized conjugacy in 11.64 was brought to light by Moreau [1967], [1970]. Something similar was noted by Weiss [1969], [1974], and also by Elster and Nehse [1974]. Such ideas were utilized by Balder [1977] in work related to augmented Lagrangians; this expanded on the strategy in Rockafellar [1974b], where the basic quadratic transform in 11.66 was implicitly utilized for this purpose. For related work see also Dolecki and Kurcyusz [1978]. The basic quadratic transform was fleshed out by Poliquin [1990], who put it to work in nonsmooth analysis; he demonstrated by this means, for instance, that any proper, lsc function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ that's bounded from below can be expressed as a composite function $g \circ F$ with g lsc convex and F of class $\mathcal{C}^\infty$. The full quadratic transform in 11.65 was set up earlier by Janin [1973]. He observed that its main properties could be derived as consequences of known features of the Legendre-Fenchel transform.

The earliest duality theory of the double-min variety as in 11.67 is due to Toland [1978], [1979]. He concentrated on the difference of two convex functions; there was no linear mapping, as associated here with the matrix A. The particular content of Theorem 11.67, in adding a linear transformation and drawing on facts in 11.8 (coercivity versus finiteness) and in 11.13 (strict convexity versus differentiability) hasn't been furnished before. The idea of relating 'extremal points' of one problem to those of another was carried forward on a broader front by Ekeland [1977], who, like Toland, was motivated by applications in the calculus of variations.

Also in the line of duality for nonconvex problems of optimization, the work of Aubin and Ekeland [1976] deserves special mentions. They constructed quantitative estimates of the 'lack of convexity' of a function and showed how these estimates can be utilized to get bounds on the size of the duality gap (between primal and dual optimal values) in a Fenchel-like format.

Although we haven't taken it up here, there is also a concept of conjugacy for convex-concave functions; see Rockafellar [1964b], [1970a]. A theory of dual minimax problems has been developed in such terms by McLinden [1973], [1974]. Epi-convergence isn't the right convergence for such functions and must be replaced by *epi-hypo-convergence*; see Attouch and Wets [1983a], Attouch, Azé and Wets [1988].