It is also interesting to observe that the dual problem only depends on the inner product and we can generalize it easily by using the Kernel trick (Sect. 2.2.6).

2.4 Legendre–Fenchel Conjugate Duality

2.4.1 Closure of Convex Functions

We can extend the domain of a convex function $f : X \to \mathbb{R}$ to the whole space $\mathbb{R}^n$ by setting $f(x) = +\infty$ for any $x \notin X$. In view of the definition of a convex function in (2.2.8), and our discussion about the epigraphs in Sect. 2.2, a function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is convex if and only if its epigraph

$$\operatorname{epi}(f) = \{(x,t) \in \mathbb{R}^{n+1} : f(x) \leq t\}$$

is a nonempty convex set.

As we know, closed convex sets possess many nice topological properties. For example, a closed convex set is comprised of the limits of all converging sequences of elements. Moreover, by the Separation Theorem, a closed and nonempty convex set X is the intersection of all closed half-spaces containing X. Among these half-spaces, the most interesting ones are the supporting hyperplanes touching X on the relative boundary.

In functional Language, the "closedness" of epigraph corresponds to a special type of continuity, i.e., the lower semicontinuity. Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a given function (not necessarily convex). We say that f is lower semicontinuous (l.s.c.) at a point $\bar{x}$, if for every sequence of points $\{x_i\}$ converging to $\bar{x}$ one has

$$f(\bar{x}) \leq \liminf_{i\to\infty} f(x_i).$$

Of course, lim inf of a sequence with all terms equal to $+\infty$ is $+\infty$. f is called lower semicontinuous, if it is lower semicontinuous at every point.

A trivial example of a lower semicontinuous function is a continuous one. Note, however, that a semicontinuous function is not necessarily continuous. What it is obliged is to make only "jumps down." For example, the function

$$f(x) = \begin{cases} 0, & x \neq 0 \\ a, & x = 0 \end{cases}$$

is lower semicontinuous if $a \leq 0$ ("jump down at $x = 0$ or no jump at all"), and is not lower semicontinuous if $a > 0$ ("jump up").

The following statement links lower semicontinuity with the geometry of the epigraph.

Proposition 2.10. *A function f defined on $\mathbb{R}^n$ and taking values from $\mathbb{R} \cup \{+\infty\}$ is lower semicontinuous if and only if its epigraph is closed (e.g., due to its emptiness).*

Proof. We first prove the "only if" part (from lower semicontinuity to closed epigraph). Let (x,t) be the limit of the sequence $\{x_i, t_i\} \subset \operatorname{epi} f$. Then we have $f(x_i) \leq t_i$. Thus the following relation holds: $t = \lim_{i\to\infty} t_i \geq \lim_{i\to\infty} f(x_i) \geq f(x)$.

We now show the "if" part (from closed epigraph to lower semicontinuity). Suppose for contradiction that $f(x) > \gamma > \lim_{i\to\infty} f(x_i)$ for some constant γ, where x_i converges to x. Then there exists a subsequence $\{x_{i_k}\}$ such that $f(x_{i_k}) \leq \gamma$ for all i_k. Since the epigraph is closed, then x must belong to this set, which implies that $f(x) \leq \gamma$, which is a contradiction. ∎

As an immediate consequence of Proposition 2.10, the upper bound

$$f(x) = \sup_{\alpha \in \mathscr{A}} f_\alpha(x)$$

of arbitrary family of lower semicontinuous functions is lower semicontinuous. Indeed, the epigraph of the upper bound is the intersection of the epigraphs of the functions forming the bound, and the intersection of closed sets always is closed.

Now let us look at proper lower semicontinuous (l.s.c.) convex functions. According to our general convention, a convex function f is proper if $f(x) < +\infty$ for at least one x and $f(x) > -\infty$ for every x. This implies that proper convex functions have convex nonempty epigraphs. Also as we just have seen, "lower semicontinuous" means "with closed epigraph." Hence proper l.s.c. convex functions always have closed convex nonempty epigraphs.

Similar to the fact that a closed convex set is intersection of closed half-spaces, we can provide an outer description of a proper l.s.c. convex function. More specifically, we can show that a proper l.s.c. convex function f is the upper bound of all its affine minorants given in the form of $t \geq d^T x - a$. Moreover, at every point $\bar{x} \in \operatorname{ri} \operatorname{dom} f$ from the relative interior of the domain f, f is even not the upper bound, but simply the maximum of its minorants: there exists an affine function $f_{\bar{x}}(x)$ which underestimates $f(x)$ everywhere in $\mathbb{R}^n$ and is equal to f at $x = \bar{x}$. This is exactly the first-order approximation $f(\bar{x}) + \langle g(\bar{x}), x - \bar{x} \rangle$ given by the definition of subgradients.

Now, what if the convex function is not lower semicontinuous (see Fig. 2.5)? A similar question also arises about convex sets—what to do with a convex set which is not closed? To deal with these convex sets, we can pass from the set to its closure and thus get a "normal" object which is very "close" to the original one. Specifically, while the "main part" of the original set—its relative interior—remains unchanged, we add a relatively small "correction," i.e., the relative boundary, to the set. The same approach works for convex functions: if a proper convex function f is not l.s.c. (i.e., its epigraph, being convex and nonempty, is not closed), we can "correct" the function—replace it with a new function with the epigraph being the closure of $\operatorname{epi}(f)$. To justify this approach, we should make sure that the closure of the epigraph of a convex function is also an epigraph of such a function.

Thus, we conclude that the closure of the epigraph of a convex function f is the epigraph of certain function, referred to as *the closure* $\operatorname{cl} f$ *of* f. Of course, this

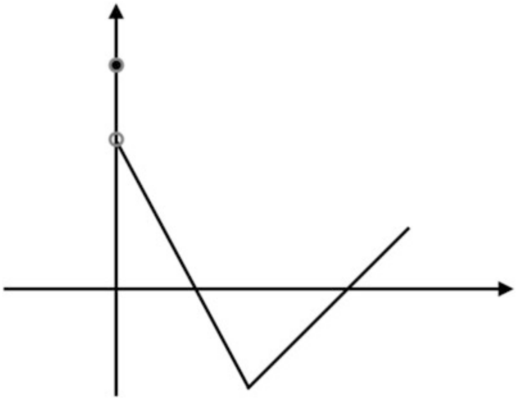

Fig. 2.5 Example for an upper semicontinuous function. The domain of this function is $[0,+\infty)$, and it "jumps up" at 0. However, the function is still convex

latter function is convex (its epigraph is convex—it is the closure of a convex set), and since its epigraph is closed, $\mathrm{cl} f$ is proper. The following statement gives direct description of $\mathrm{cl} f$ in terms of f:

(i) For every x one has $\mathrm{cl} f(x) = \lim\limits_{r\to+0} \inf\limits_{x':\|x'-x\|_2\le r} f(x')$. In particular,

$$f(x) \ge \mathrm{cl} f(x)$$

for all x, and

$$f(x) = \mathrm{cl} f(x)$$

whenever $x \in \mathrm{ri}\,\mathrm{dom} f$, or equivalently whenever $x \notin \mathrm{cl}\,\mathrm{dom} f$. Thus, the "correction" $f \mapsto \mathrm{cl} f$ may vary f only at the points from the relative boundary of $\mathrm{dom} f$,

$$\mathrm{dom} f \subset \mathrm{dom}\,\mathrm{cl} f \subset \mathrm{cl}\,\mathrm{dom} f,$$

hence

$$\mathrm{ri}\,\mathrm{dom} f = \mathrm{ri}\,\mathrm{dom}\,\mathrm{cl} f.$$

(ii) The family of affine minorants of $\mathrm{cl} f$ is exactly the family of affine minorants of f, so that

$$\mathrm{cl} f(x) = \sup\{\phi(x) : \phi \text{ is an affine minorant of } f\},$$

due to the fact that $\mathrm{cl} f$ is l.s.c. and is therefore the upper bound of its affine minorants, and the sup in the right-hand side can be replaced with max whenever $x \in \mathrm{ri}\,\mathrm{dom}\,\mathrm{cl} f = \mathrm{ri}\,\mathrm{dom} f$.

2.4.2 Conjugate Functions

Let f be a convex function. We know that f "basically" is the upper bound of all its affine minorants. This is exactly the case when f is proper, otherwise the corresponding equality takes place everywhere except, perhaps, some points from the relative boundary of $\mathrm{dom} f$. Now, when an affine function $d^T x - a$ is an affine minorant of f? It is the case if and only if

$$f(x) \geq d^T x - a$$

for all x or, which is the same, if and only if

$$a \geq d^T x - f(x)$$

for all x. We see that if the slope d of an affine function $d^T x - a$ is fixed, then in order for the function to be a minorant of f we should have

$$a \geq \sup_{x \in \mathbb{R}^n} [d^T x - f(x)].$$

The supremum in the right-hand side of the latter relation is certain function of d and this function, denoted by f^*, is called the *Legendre–Fenchel conjugate* of f:

$$f^*(d) = \sup_{x \in \mathbb{R}^n} [d^T x - f(x)].$$

Geometrically, the Legendre–Fenchel transformation answers the following question: given a slope d of an affine function, i.e., given the hyperplane $t = d^T x$ in $\mathbb{R}^{n+1}$, what is the minimal "shift down" of the hyperplane which places it below the graph of f?

From the definition of the conjugate it follows that this is a proper l.s.c. convex function. Indeed, we lose nothing when replacing $\sup_{x \in \mathbb{R}^n} [d^T x - f(x)]$ by $\sup_{x \in \mathrm{dom} f} [d^T x - f(x)]$, so that the conjugate function is the upper bound of a family of affine functions. This bound is finite at least at one point, namely, at every d coming from affine minorant of f, and we know that such a minorant exists. Therefore, f^* must be a proper l.s.c. convex function, as claimed.

The most elementary (and the most fundamental) fact about the conjugate function is its symmetry.

Proposition 2.11. *Let f be a convex function. Then $(f^*)^* = \mathrm{cl} f$. In particular, if f is l.s.c., then $(f^*)^* = f$.*

Proof. The conjugate function of f^* at the point x is, by definition,

$$\sup_{d \in \mathbb{R}^n} [x^T d - f^*(d)] = \sup_{d \in \mathbb{R}^n, a \geq f^*(d)} [d^T x - a].$$

The second sup here is exactly the supremum of all affine minorants of f due to the origin of the Legendre–Fenchel transformation: $a \geq f^*(d)$ if and only if the affine form $d^T x - a$ is a minorant of f. The result follows since we already know that the upper bound of all affine minorants of f is the closure of f. ■

The Legendre–Fenchel transformation is a very powerful tool—this is a "global" transformation, so that *local* properties of f^* correspond to *global* properties of f.

- $d = 0$ belongs to the domain of f^* if and only if f is below bounded, and if it is the case, then $f^*(0) = -\inf f$;
- if f is proper and l.s.c., then the subgradients of f^* at $d = 0$ are exactly the minimizers of f on $\mathbb{R}^n$;
- $\mathrm{dom} f^*$ is the entire $\mathbb{R}^n$ if and only if $f(x)$ grows, as $\|x\|_2 \to \infty$, faster than $\|x\|_2$: there exists a function $r(t) \to \infty$, as $t \to \infty$ such that

$$f(x) \geq r(\|x\|_2) \quad \forall x.$$

Thus, whenever we can compute explicitly the Legendre–Fenchel transformation of f, we get a lot of "global" information on f.

Unfortunately, the more detailed investigation of the properties of Legendre–Fenchel transformation is beyond our scope. Below we simply list some facts and examples.

- From the definition of Legendre transformation,

$$f(x) + f^*(d) \geq x^T d \quad \forall x, d.$$

  Specifying here f and f^*, we get certain inequality, e.g., the following one: [Young's Inequality] *if p and q are positive reals such that $\frac{1}{p} + \frac{1}{q} = 1$, then*

$$\frac{|x|^p}{p} + \frac{|d|^q}{q} \geq xd \quad \forall x, d \in \mathbb{R}$$

- The Legendre–Fenchel transformation of the function

$$f(x) \equiv -a$$

  is the function which is equal to a at the origin and is $+\infty$ outside the origin; similarly, the Legendre–Fenchel transformation of an affine function $\bar{d}^T x - a$ is equal to a at $d = \bar{d}$ and is $+\infty$ when $d \neq \bar{d}$;
- The Legendre–Fenchel transformation of the strictly convex quadratic form

$$f(x) = \frac{1}{2} x^T A x$$

  ($A \succeq 0$) is the quadratic form

$$f^*(d) = \frac{1}{2} d^T A^{-1} d$$

- The Legendre–Fenchel transformation of the Euclidean norm

$$f(x) = \|x\|_2$$

  is the function which is equal to 0 in the closed unit ball centered at the origin and is $+\infty$ outside the ball.