Chapter 4
Convex analysis

The phrase *convex analysis* refers to a body of generalized calculus that can be developed for convex functions and sets. This topic, whose applications are widespread, is the subject of the chapter. The central element of the theory is the *subdifferential*, a construct which plays a role similar to that of the derivative. The operation of *conjugacy* as it applies to convex functions will also be important, as well as *polarity* of sets.

4.1 Subdifferential calculus

Let $f : X \to \mathbb{R}_\infty$ be a given function, where X is a normed space, and let x be a point in $\operatorname{dom} f$. An element ζ of X^* is called a **subgradient** of f at x (in the sense of convex analysis) if it satisfies the following *subgradient inequality*:

$$f(y) - f(x) \geqslant \langle \zeta, y - x \rangle, \quad y \in X.$$

A function is called *affine* when it differs by a constant from a linear functional. Thus, an affine function g has the form $g(y) = \langle \zeta, y \rangle + c$; the linear functional ζ is called the *slope* of g. When the subgradient inequality above holds, the affine function

$$y \mapsto f(x) + \langle \zeta, y - x \rangle$$

is said to *support* f at x; this means that it lies everywhere below f, and that equality holds at x. In geometric terms, we may formulate the situation as follows: the hyperplane $\{ (y, r) : r - f(x) - \langle \zeta, y - x \rangle = 0 \}$ in the product space $X \times \mathbb{R}$ passes through the point $(x, f(x))$, and the set $\operatorname{epi} f$ lies in the upper associated halfspace (see p. 41 for the terminology). We refer to this as a *supporting hyperplane*.

The set of all subgradients of f at x is denoted by $\partial f(x)$, and referred to as the **subdifferential** of f at x. It follows from the definition that the subdifferential $\partial f(x)$ is a convex set which is closed for the weak* topology, since, for each y, the

F. Clarke, *Functional Analysis, Calculus of Variations and Optimal Control*,
Graduate Texts in Mathematics 264, DOI 10.1007/978-1-4471-4820-3_4,

set of ζ satisfying the subgradient inequality is weak* closed and convex.[1] The map $x \mapsto \partial f(x)$ is set-valued: its values are subsets of X^*. We use the term *multifunction* in such a case: ∂f is a multifunction from X to X^*.

4.1 Example. We illustrate the geometry of subgradients with the help of Fig. 4.1, which we think of as depicting the epigraph of a convex function $f : \mathbb{R}^n \to \mathbb{R}_\infty$.

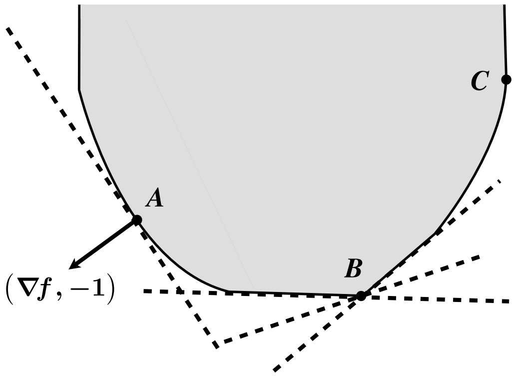

Fig. 4.1
The epigraph of a convex function, and some supporting hyperplanes.

The function is smooth near the point A on the boundary of its epigraph; let this point be $(x_1, f(x_1))$. There is a unique affine function $y = \langle \zeta, x \rangle + c$ that supports f at the point x_1; its slope ζ is given by $\nabla f(x_1)$. The vector $(\nabla f(x_1), -1)$ is orthogonal to the corresponding supporting hyperplane, and generates the normal cone to epi f at $(x_1, f(x_1))$ (here, a ray).

At the point B, which we take to be $(x_2, f(x_2))$, the function f has a corner, and there are infinitely many affine functions supporting f at x_2; the set of all their slopes constitutes $\partial f(x_2)$. There is a supporting hyperplane to epi f at the point C as well, but it is vertical, and therefore does not define a subgradient (it fails to correspond to the graph of an affine function of x). The subdifferential of f is empty at the corresponding value of x. □

4.2 Exercise. (subdifferential of the norm) Let f be the function $f(x) = \|x\|$.

(a) Prove that $\partial f(0)$ is the closed unit ball in X^*.

(b) Let $\zeta \in \partial f(x)$, where $x \neq 0$. Prove that $\langle \zeta, x \rangle = \|x\|$ and $\|\zeta\|_* = 1$. □

It is clear from the definition of subgradient that f attains a minimum at x if and only if $0 \in \partial f(x)$. This version of Fermat's rule is but the first of several ways in which the reader will detect a kinship between the subdifferential and the derivative. The following provides another example.

[1] In speaking, the subdifferential ∂f is often pronounced "dee eff" or "curly dee eff".

4.3 Proposition. *Let* $f : X \to \mathbb{R}_\infty$ *be a convex function, and* $x \in \operatorname{dom} f$. *Then*

$$\partial f(x) = \big\{ \zeta \in X^* : f'(x;v) \geqslant \langle \zeta, v \rangle \ \forall v \in X \big\}.$$

Proof. We recall that a convex function admits directional derivatives, as showed in Prop. 2.22. If $\zeta \in \partial f(x)$, then we have

$$f(x+tv) - f(x) \geqslant \langle \zeta, tv \rangle \ \forall v \in X, \ t > 0,$$

by the subgradient inequality. It follows that $f'(x;v) \geqslant \langle \zeta, v \rangle \ \forall v$. Conversely, if this last condition holds, then (by Prop. 2.22) we have

$$f(x+v) - f(x) \geqslant \inf_{t>0} \frac{f(x+tv) - f(x)}{t} \geqslant \langle \zeta, v \rangle \ \forall v \in X,$$

which implies $\zeta \in \partial f(x)$. □

It is a consequence of the proposition above that if f is differentiable at x, then $\partial f(x) = \{ f'(x) \}$. The reduction of $\partial f(x)$ to a singleton, however, is more closely linked to a weaker type of derivative, one that we proceed to introduce.

The Gâteaux derivative. Let $F : X \to Y$ be a function between two normed spaces. We say that F is *Gâteaux differentiable* at x if the directional derivative $F'(x;v)$ exists for all $v \in X$, and if there exists $\Lambda \in L_C(X,Y)$ such that

$$F'(x;v) = \langle \Lambda, v \rangle \ \forall v \in X.$$

It follows that the element Λ is unique; it is denoted $F'_G(x)$ and referred to as the Gâteaux derivative. It corresponds to a weaker concept of differentiability than the Fréchet derivative $F'(x)$ that we met in §1.4. In fact, Gâteaux differentiability at x does not even imply continuity of F at x. When F is Fréchet differentiable at x, then F is Gâteaux differentiable at x, and $F'_G(x) = F'(x)$. We stress that the unqualified word "differentiable" always refers to the usual (Fréchet) derivative.

The following is a direct consequence of Prop. 4.3.

4.4 Corollary. *Let* $f : X \to \mathbb{R}_\infty$ *be convex, with* $x \in \operatorname{dom} f$. *If* f *is Gâteaux differentiable at* x, *then* $\partial f(x) = \{ f'_G(x) \}$.

A characteristic of convex analysis that distinguishes it from classical differential analysis is the close link that it establishes between sets and functions. The following records an important (yet simple) example of this.

4.5 Exercise. Let $x \in S$, where S is a convex subset of X. Prove that $\partial I_S(x) = N_S(x)$; that is, the subdifferential of the indicator function is the normal cone. □

For the convex function f of Example 2.23, the reader may check that $\partial f(-1)$ and $\partial f(1)$ are empty, and that $\partial f(x)$ is a singleton when $-1 < x < +1$. It is possible for the subdifferential of a convex function to be empty at points in the interior of its effective domain: if Λ is a discontinuous linear functional, then $\partial \Lambda(x)$ is empty for each x, as the reader may care to show. This phenomenon does not occur at points of continuity, however, as we now see.

4.6 Proposition. *Let $f : X \to \mathbb{R}_\infty$ be a convex function, and let $x \in \operatorname{dom} f$ be a point of continuity of f. Then $\partial f(x)$ is nonempty and weak* compact. If f is Lipschitz of rank K in a neighborhood of x, then $\partial f(x) \subset K B_*$.*

Proof. The continuity at x implies that $\operatorname{int} \operatorname{epi} f \neq \emptyset$. We separate $\operatorname{int} \operatorname{epi} f$ and the point $(x, f(x))$ (using the first case of Theorem 2.37) to deduce the existence of $\zeta \in X^*$ and $\lambda \in \mathbb{R}$ such that

$$\langle \zeta, y \rangle + \lambda r < \langle \zeta, x \rangle + \lambda f(x) \;\; \forall (y, r) \in \operatorname{int} \operatorname{epi} f.$$

It follows that $\lambda < 0$; we may therefore normalize by taking $\lambda = -1$. Since

$$\operatorname{epi} f \subset \operatorname{cl} \operatorname{epi} f = \operatorname{cl}\big(\operatorname{int} \operatorname{epi} f\big)$$

(by Theorem 2.2), the separation inequality implies

$$\langle \zeta, y \rangle - r \leqslant \langle \zeta, x \rangle - f(x) \;\; \forall (y, r) \in \operatorname{epi} f.$$

This reveals that $\partial f(x)$ contains ζ, and is therefore nonempty.

Theorem 2.34 asserts that f is Lipschitz on some neighborhood of x. Let K be a Lipschitz constant for f on $B(x, r)$, and let ζ be any element of $\partial f(x)$. The subgradient inequality yields

$$\langle \zeta, y - x \rangle \leqslant f(y) - f(x) \leqslant |f(y) - f(x)| \leqslant K|y - x| \;\; \forall y \in B(x, r).$$

Putting $y = x + rv$, where v is a unit vector, leads to

$$\langle \zeta, v \rangle \leqslant K|v| \;\; \forall v \in B,$$

whence $|\zeta| \leqslant K$. Thus, $\partial f(x)$ is bounded, and weak* compact by Cor. 3.15. ☐

4.7 Corollary. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function. Then for any $x \in \mathbb{R}^n$, $\partial f(x)$ is a nonempty convex compact set.*

Proof. This follows from Cor. 2.35. ☐

Recall that we have agreed to identify the dual of $\mathbb{R}^n$ with $\mathbb{R}^n$ itself. Thus, for a function f defined on $\mathbb{R}^n$, the subdifferential $\partial f(x)$ is viewed as a subset of $\mathbb{R}^n$. The very existence of a subgradient is the key to the proof of the next result, a well-known inequality.

4.8 Corollary. (Jensen's inequality) *Let $\varphi : \mathbb{R}^k \to \mathbb{R}$ be convex. Then, for any summable function $g : (0,1) \to \mathbb{R}^k$, we have*

$$\varphi\Big(\int_0^1 g(t)\,dt \Big) \leqslant \int_0^1 \varphi\big(g(t)\big)\,dt.$$

Proof. Let us define a point in $\mathbb{R}^k$ by

$$\bar{g} := \int_0^1 g(t)\,dt,$$

the integral being understood in the vector sense. By Cor. 4.7, $\partial\varphi(\bar{g})$ contains an element ζ. Then, by definition, we have $\varphi(y) - \varphi(\bar{g}) \geqslant \langle \zeta, y - \bar{g} \rangle \;\; \forall y \in \mathbb{R}^k$. Substituting $y = g(t)$ and integrating over $[0,1]$, we obtain the stated inequality. □

4.9 Exercise.

(a) Modify the statement of Jensen's inequality appropriately when the underlying interval is $[a,b]$ rather than $[0,1]$.

(b) Formulate and prove Jensen's inequality in several dimensions, when the function g belongs to $L^1(\Omega, \mathbb{R}^k)$, Ω being a bounded open subset of $\mathbb{R}^n$. □

The appealing calculus formula $\partial(f+g)(x) = \partial f(x) + \partial g(x)$ turns out to be true under mild hypotheses, as we see below. Note that *some* hypothesis is certainly required in order to assert such a formula, since it fails when we take $f = \Lambda$ and $g = -\Lambda$, where Λ is a discontinuous linear functional.

4.10 Theorem. (Subdifferential of the sum) *Let $f, g : X \to \mathbb{R}_\infty$ be convex functions which admit a point in $\operatorname{dom} f \cap \operatorname{dom} g$ at which f is continuous. Then we have*

$$\partial\big(f+g\big)(x) = \partial f(x) + \partial g(x) \;\; \forall x \in \operatorname{dom} f \cap \operatorname{dom} g.$$

Proof. That the left side above contains the right follows directly from the definition of subdifferential. Now let ζ belong to $\partial(f+g)(x)$; we must show that ζ belongs to the right side. We may (and do) reduce to the case $x = 0$, $f(0) = g(0) = 0$. By hypothesis, there is a point $\bar{x}$ in $\operatorname{dom} f \cap \operatorname{dom} g$ at which f is continuous. Then the subsets of $X \times \mathbb{R}$ defined by

$$C = \operatorname{int} \operatorname{epi} f, \;\; D = \big\{ (w,t) : \langle \zeta, w \rangle - g(w) \geqslant t \big\}$$

are nonempty as a consequence of the existence of $\bar{x}$. They are also convex, and disjoint as a result of the subgradient inequality for ζ. By Theorem 2.37, C and D can be separated: there exist $\xi \in X^*$ and $\lambda \in \mathbb{R}$ such that

$$\langle \xi, w \rangle + \lambda t < \langle \xi, u \rangle + \lambda s \;\; \forall (w,t) \in D, \;\; \forall (u,s) \in \operatorname{int} \operatorname{epi} f.$$

It follows that $\lambda > 0$; we can normalize by taking $\lambda = 1$. Since (by Theorem 2.2) we have $\operatorname{epi} f \subset \operatorname{cl} \operatorname{epi} f = \operatorname{cl}(\operatorname{int} \operatorname{epi} f)$, we deduce

$$\langle \xi, w \rangle + t \leqslant \langle \xi, u \rangle + s \ \ \forall (w,t) \in D, \ \ \forall (u,s) \in \operatorname{epi} f.$$

Taking $(w,t) = (0,0)$, this implies $-\xi \in \partial f(0)$. Taking $(u,s) = (0,0)$ leads to the conclusion $\xi + \zeta \in \partial g(0)$. Thus, $\zeta \in \partial f(0) + \partial g(0)$. □

4.11 Exercise. Let C and D be convex subsets of X such that $(\operatorname{int} C) \cap D \neq \emptyset$. Let $x \in C \cap D$. Then $N_{C \cap D}(x) = N_C(x) + N_D(x)$. □

The following Fermat-type result is related to that of Prop. 2.25; f is now nondifferentiable (which means "not necessarily differentiable"), but convex, and the necessary condition turns out to be sufficient as well.

4.12 Proposition. *Let $f : X \to \mathbb{R}$ be a continuous convex function, A a convex subset of X, and x a point in A. Then the following are equivalent*:

(a) *x minimizes f over the set A.*

(b) *$-\partial f(x) \cap N_A(x) \neq \emptyset$; or equivalently, $0 \in \partial f(x) + N_A(x)$.*

Proof. If (a) holds, then x minimizes $u \mapsto f(u) + I_A(u)$. Thus $0 \in \partial(f + I_A)(x)$, by Fermat's rule. But we have

$$\partial\big(f + I_A\big)(x) = \partial f(x) + \partial I_A(x) = \partial f(x) + N_A(x),$$

by Theorem 4.10 and Exer. 4.5; thus, (b) holds. Conversely, if (b) holds, then we have $0 \in \partial(f + I_A)(x)$, which implies (a). □

Recall that the adjoint of a linear application T is denoted T^* (see p. 22). It plays a role in the next result, a calculus rule for a composition.

4.13 Theorem. *Let Y be a normed space, $T \in L_C(X,Y)$, and let $g : Y \to \mathbb{R}_\infty$ be a convex function. Let there be a point x_0 in X such that g is continuous at Tx_0. Then the function $f(x) = g(Tx)$ is convex, and we have*

$$\partial f(x) = T^* \partial g(Tx) \ \ \forall x \in \operatorname{dom} f.$$

Note: the meaning of this formula is that, given any $\zeta \in \partial f(x)$, there exists γ in $\partial g(Tx)$ such that

$$\langle \zeta, v \rangle = \langle T^* \gamma, v \rangle = \langle \gamma, Tv \rangle \ \ \forall v \in X,$$

and that, conversely, any element ζ of the form $T^*\gamma$, where $\gamma \in \partial g(Tx)$, belongs to $\partial f(x)$.

Proof. It is easily verified that f is convex, and that any element of $T^*\partial g(Tx)$ lies in $\partial f(x)$. We now prove the opposite inclusion. Let $\varphi : X \times Y \to \mathbb{R}_\infty$ be defined by

$$\varphi(x,y) = g(y) + I_{\operatorname{gr} T}(x,y),$$

where gr T is the graph of T: the set $\{(x,Tx) \in X \times Y : x \in X\}$. It follows from the definition of subgradient that

$$\zeta \in \partial f(x) \iff (\zeta, 0) \in \partial \varphi(x, Tx).$$

Because of the existence of x_0, we may apply Theorem 4.10 to φ, for any ζ as above. There results $(\alpha, \beta) \in N_{\operatorname{gr} T}(x,Tx)$ and $\gamma \in \partial g(Tx)$ such that

$$(\zeta, 0) = (0, \gamma) + (\alpha, \beta).$$

The normal vector (α, β) satisfies

$$\langle \alpha, u - x \rangle + \langle \beta, Tu - Tx \rangle \leqslant 0 \quad \forall u \in X,$$

which implies $\alpha = -T^*\beta$. We deduce $\zeta = T^*\gamma \in T^*\partial g(Tx)$, as required. □

Subdifferentials in Euclidean space. The basic theory of the subdifferential takes on a simpler form when we restrict attention to $\mathbb{R}^n$. We focus on this case in the remainder of this section.

4.14 Proposition. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex. Then*

(a) *The graph of ∂f is closed:* $\zeta_i \in \partial f(x_i),\ \zeta_i \to \zeta,\ x_i \to x \Longrightarrow \zeta \in \partial f(x)$;

(b) *For any compact subset S of $\mathbb{R}^n$, there exists M such that*

$$|\zeta| \leqslant M \quad \forall \zeta \in \partial f(x),\ x \in S;$$

(c) *For any x, for any $\varepsilon > 0$, there exists $\delta > 0$ such that*

$$|y - x| < \delta \implies \partial f(y) \subset \partial f(x) + \varepsilon B.$$

Proof. Consider the situation described in (a). For any $y \in \mathbb{R}^n$, we have

$$f(y) - f(x_i) \geqslant \langle \zeta_i, y - x_i \rangle \quad \forall i.$$

Taking limits, and bearing in mind that f is continuous (Cor. 2.35), we obtain in the limit $f(y) - f(x) \geqslant \langle \zeta, y - x \rangle$; thus, $\zeta \in \partial f(x)$.

We know that f is locally Lipschitz. An elementary argument using compactness shows that f is Lipschitz on bounded subsets of $\mathbb{R}^n$ (see Exer. 2.32). This, together with Prop. 4.6, implies (b). Part (c) follows from an argument by contradiction, using parts (a) and (b); we entrust this step to the reader. □

4.15 Exercise. (Mean value theorem) Let $f:\mathbb{R}^n \to \mathbb{R}$ be a convex function.

(a) We fix $x, v \in \mathbb{R}^n$ and we set $g(t) = f(x+tv)$ for $t \in \mathbb{R}$. Show that g is convex, and that

$$\partial g(t) = \big\langle \partial f(x+tv), v \big\rangle = \big\{ \langle \xi, v \rangle : \xi \in \partial f(x+tv) \big\}.$$

(b) Prove the following vaguely familiar-looking theorem: for all $x, y \in \mathbb{R}^n$, $x \neq y$, there exists $z \in (x,y)$ such that

$$f(y) - f(x) \in \big\langle \partial f(z), y - x \big\rangle.$$

(c) Let U be an open convex subset of $\mathbb{R}^n$. Use the above to prove the following subdifferential characterization of the Lipschitz property:

$$f \text{ is Lipschitz of rank } K \text{ on } U \iff |\zeta| \leqslant K \;\; \forall \zeta \in \partial f(x), \;\; \forall x \in U. \qquad \square$$

4.16 Proposition. *Let $f:\mathbb{R}^n \to \mathbb{R}$ be convex. Then f is differentiable at x if and only if $\partial f(x)$ is a singleton, and f is continuously differentiable in an open subset U if and only if $\partial f(x)$ reduces to a singleton for every $x \in U$.*

Proof. If f is differentiable at x, then Cor. 4.4 asserts that $\partial f(x)$ is the singleton $\{f'(x)\}$. Conversely, suppose that $\partial f(x)$ is a singleton $\{\zeta\}$. We proceed to prove that $f'(x) = \zeta$. Let x_i be any sequence converging to x ($x_i \neq x$). Then, by Exer. 4.15, there exist $z_i \in (x_i, x)$ and $\zeta_i \in \partial f(z_i)$ such that

$$f(x_i) - f(x) = \langle \zeta_i, x_i - x \rangle.$$

By part (c) of Prop. 4.14, the sequence ζ_i necessarily converges to ζ. Thus we have

$$\frac{|f(x_i) - f(x) - \langle \zeta, x_i - x \rangle|}{|x_i - x|} = \frac{|\langle \zeta_i - \zeta, x_i - x \rangle|}{|x_i - x|} \to 0,$$

whence $f'(x)$ exists and equals ζ. The final assertion is now easily verified with the help of part (c) of Prop. 4.14. $\square$

Strict convexity. Let U be a convex subset of X, and let $f: U \to \mathbb{R}$ be given. We say that f is *strictly convex* if the defining inequality of convexity is strict whenever it has any chance of being so:

$$x, y \in U, \; x \neq y, \; t \in (0,1) \implies f((1-t)x + ty) < (1-t)f(x) + tf(y).$$

It is easy to see that a C^2 function f of a single variable that satisfies $f''(t) > 0 \;\; \forall t$ is strictly convex (see Exer. 8.26 for an extension of this criterion to several dimensions). The role of strict convexity in optimization is partly to assure uniqueness of a minimum: the reader may check that a strictly convex function cannot attain its minimum at two different points.

4.17 Exercise. Let U be an open convex subset of $\mathbb{R}^n$, and let $f : U \to \mathbb{R}$ be convex.

(a) Prove that f is strictly convex if and only if

$$x, y \in U,\ x \neq y,\ \zeta \in \partial f(x) \implies f(y) - f(x) > \langle \zeta, y - x \rangle.$$

(b) Prove that f is strictly convex if and only if ∂f is injective, in the following sense:

$$x, y \in U,\ \partial f(x) \cap \partial f(y) \neq \emptyset \implies x = y. \qquad \square$$

4.2 Conjugate functions

Let X continue to designate a normed space, and let $f : X \to \mathbb{R}_\infty$ be a proper function (that is, $\operatorname{dom} f \neq \emptyset$). The *conjugate function* $f^* : X^* \to \mathbb{R}_\infty$ of f is defined by

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

(One also refers to f^* as the *Fenchel* conjugate.) Note that the properness of f rules out the possibility that f^* has the value $-\infty$; we say then that f^* is well defined.

If $g : X^* \to \mathbb{R}_\infty$ is a proper function, its conjugate $g^* : X \to \mathbb{R}_\infty$ is defined by

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

Note that g^* is defined on X, and not on the dual of X^*, which we wish to avoid here. A special case arises when we take g to be f^*; then we obtain the *biconjugate* of f, namely the function $f^{**} : X \to \mathbb{R}_\infty$ defined as follows (when f^* is proper):

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

Since taking upper envelopes preserves both convexity and lower semicontinuity, it follows from the definition that f^* is convex lsc on the normed space X^*, and that f^{**} is convex lsc on X. The reader will observe that $f \leqslant g \implies f^* \geqslant g^*$.

4.18 Exercise.

(a) Show that for any function $f : X \to \mathbb{R}_\infty$, we have $f^{**} \leqslant f$.

(b) If f is proper, prove *Fenchel's inequality* :

$$f(x) + f^*(\zeta) \geqslant \langle \zeta, x \rangle \ \forall x \in X \ \forall \zeta \in X^*,$$

with equality if and only if $\zeta \in \partial f(x)$.

(c) Let f be the function $|x|^p/p$ on $\mathbb{R}^n$ $(1 < p < \infty)$. Calculate f^*, and show that Fenchel's inequality reduces in this case to *Young's inequality*:

$$u \bullet v \leqslant \frac{1}{p}|u|^p + \frac{1}{p_*}|v|^{p_*}, \quad u, v \in \mathbb{R}^n.$$

When does equality hold? □

4.19 Proposition. *Let $f : X \to \mathbb{R}_\infty$ be proper, and $c \in \mathbb{R}$ and $\zeta \in X^*$ be given. Then*

$$f(x) \geqslant \langle \zeta, x \rangle - c \ \forall x \in X \iff f^*(\zeta) \leqslant c.$$

If f is bounded below by a continuous affine function, then f^* and f^{**} are proper.

Proof. The first assertion, an equivalence, follows directly from the definition of f^*; it evidently implies that f^* is proper whenever f is bounded below by (majorizes, some would say) a continuous affine function. Since $f^{**} \leqslant f$, and since f is proper, we also deduce that f^{**} is proper. □

4.20 Proposition. *Let f be convex and lsc. Then f is bounded below by a continuous affine function. More explicitly, let x_0 be any point in X. If $x_0 \in \operatorname{dom} f$, then for any $\varepsilon > 0$, there exists $\zeta \in \operatorname{dom} f^*$ such that*

$$f(x) > f(x_0) + \langle \zeta, x - x_0 \rangle - \varepsilon \ \forall x \in X. \tag{1}$$

If $f(x_0) = +\infty$, then for any $M \in \mathbb{R}$, there exists $\zeta \in \operatorname{dom} f^*$ such that

$$f(x) > M + \langle \zeta, x - x_0 \rangle \ \forall x \in X. \tag{2}$$

Proof. Consider first the case $x_0 \in \operatorname{dom} f$. We apply the separation theorem to the point $(x_0, f(x_0) - \varepsilon)$ (a compact set) and the (closed) set $\operatorname{epi} f$. There results $\zeta \in X^*$ and $\lambda \in \mathbb{R}$ such that

$$\lambda r + \langle \zeta, x \rangle < \lambda f(x_0) - \lambda \varepsilon + \langle \zeta, x_0 \rangle \ \forall (x, r) \in \operatorname{epi} f.$$

It follows that $\lambda < 0$; we normalize to take $\lambda = -1$. This yields

$$f(x) > \langle \zeta, x \rangle + f(x_0) - \varepsilon - \langle \zeta, x_0 \rangle \ \forall x \in X,$$

the required inequality (which implies $\zeta \in \operatorname{dom} f^*$). The case $f(x_0) = +\infty$ is handled similarly, by separating the point (x_0, M) from $\operatorname{epi} f$. □

Notation. We denote by $\Gamma(X)$ the collection of all functions $f : X \to \mathbb{R}_\infty$ that are convex, lower semicontinuous, and proper. (This is classical notation in convex analysis.)

It follows from the two propositions above that when $f \in \Gamma(X)$, then f^* and f^{**} are both proper, which is a propitious context for studying conjugacy.

4.21 Theorem. (Moreau) *Let $f : X \to \mathbb{R}_\infty$ be a proper function. Then*

$$f \in \Gamma(X) \iff f^* \text{ is proper and } f = f^{**}.$$

Proof. When f^* is proper, then f^{**} is well defined (not $-\infty$), and it is convex and lsc as a consequence of the way it is constructed. Thus, if in addition we have $f = f^{**}$, then f belongs to $\Gamma\big(X\big)$.

Now for the converse; let $f \in \Gamma\big(X\big)$. Then f^* is proper by Prop. 4.19, and f^{**} is well defined. Since $f^{**} \leqslant f$, it suffices to establish that, for any given $x_0 \in X$, we have $f(x_0) \leqslant f^{**}(x_0)$. We reason by the absurd, supposing therefore that $f(x_0)$ is strictly greater than $f^{**}(x_0)$.

Let $M \in \mathbb{R}$ and $\varepsilon > 0$ satisfy $f(x_0) > M > f^{**}(x_0) + 2\varepsilon$. Any $\zeta \in \operatorname{dom} f^*$ admits $x_\zeta \in \operatorname{dom} f$ such that

$$f^*(\zeta) \leqslant \langle \zeta, x_\zeta \rangle - f(x_\zeta) + \varepsilon,$$

whence

$$f^{**}(x_0) \geqslant \langle x_0, \zeta \rangle - f^*(\zeta) \geqslant \langle \zeta, x_0 - x_\zeta \rangle + f(x_\zeta) - \varepsilon.$$

These facts lead to

$$f(x_0) > M > \langle \zeta, x_0 - x_\zeta \rangle + f(x_\zeta) + \varepsilon. \tag{3}$$

Now consider the case $f(x_0) < +\infty$. Then we choose $\zeta \in \operatorname{dom} f^*$ so that (1) holds. Then, using (3), we deduce

$$\begin{aligned} f(x_\zeta) &> f(x_0) + \langle \zeta, x_\zeta - x_0 \rangle - \varepsilon \\ &> \langle \zeta, x_0 - x_\zeta \rangle + f(x_\zeta) + \varepsilon + \langle \zeta, x_\zeta - x_0 \rangle - \varepsilon = f(x_\zeta), \end{aligned}$$

a contradiction. In the other case, when $f(x_0) = +\infty$, choose ζ so that (2) holds. Then (3) yields

$$M > \langle \zeta, x_0 - x_\zeta \rangle + f(x_\zeta) > \langle \zeta, x_0 - x_\zeta \rangle + M + \langle \zeta, x_\zeta - x_0 \rangle = M,$$

a contradiction which completes the proof. □

4.22 Corollary. *Let $g : X \to \mathbb{R}_\infty$ be a proper function which is bounded below by a continuous affine function. Then g^{**}, which is well defined, is the largest lsc convex function on X which is bounded above by g.*

Proof. We know that g^* is proper and that g^{**} is well defined, by Prop. 4.19. Since $g^{**} \leqslant g$, it is clear that g^{**} is indeed a convex lsc function bounded above by g. Let f be any other such function. Then $f \in \Gamma(X)$ and, by the theorem,

$$f \leqslant g \Longrightarrow f^* \geqslant g^* \Longrightarrow f = f^{**} \leqslant g^{**}.$$

□

In the corollary above, one can show that $\operatorname{epi} g^{**} = \overline{\operatorname{co}} \operatorname{epi} g$, which explains why g^{**} is sometimes referred to as the closed convex hull of g.

4.23 Corollary. *A function $g : X \to \mathbb{R}_\infty$ is convex and lsc if and only if there exists a family $\{\varphi_\alpha\}_\alpha$ of continuous affine functions on X whose upper envelope is g:*

$$g(x) = \sup_{\alpha} \varphi_\alpha(x) \ \ \forall x \in X.$$

Proof. An envelope of the indicated type is always convex and lsc, so if g has that form, it shares these properties. For the converse, we may suppose that g is proper (in addition to being convex and lsc). According to the theorem, we then have

$$g(x) = g^{**}(x) = \sup_{\zeta \in X^*} \langle \zeta, x \rangle - g^*(\zeta),$$

which expresses g as an upper envelope of the desired type. □

4.24 Exercise. We study conjugacy as it applies to indicators and support functions.

(a) Let S be a nonempty subset of X. Prove that $I_S{}^* = H_S$. Deduce from this that if S is closed and convex, then $H_S{}^* = I_S$.

(b) Let $\Sigma \subset X^*$ be nonempty, convex, and weak* closed. Prove that $H_\Sigma{}^* = I_\Sigma$. □

The following result allows us to recognize support functions.

4.25 Theorem. *Let $g : X \to \mathbb{R}_\infty$ be lsc, subadditive, and positively homogeneous, with $g(0) = 0$. Then there exists a unique nonempty weak* closed convex set Σ in X^* such that $g = H_\Sigma$. The set Σ is characterized by*

$$\Sigma = \big\{ \zeta \in X^* : g(v) \geqslant \langle \zeta, v \rangle \ \ \forall v \in X \big\},$$

and is weak* compact if and only if the function g is bounded on the unit ball.

Proof. Observe that the set Σ defined in the theorem statement is convex and weak* closed. We have (by definition)

$$g^*(\zeta) = \sup_{v \in X} \langle \zeta, v \rangle - g(v).$$

It follows from this formula and the positive homogeneity of g that $g^*(\zeta) = \infty$ if $\zeta \notin \Sigma$, and otherwise $g^*(\zeta) = 0$; that is, we have $g^* = I_\Sigma$. But $g \in \Gamma(X)$, which allows us to write (by Theorem 4.21)

$$g(x) = g^{**}(x) = \sup_{\zeta \in X^*} \langle \zeta, x \rangle - I_\Sigma(\zeta) = H_\Sigma(x).$$

The uniqueness of Σ follows from Cor. 3.13. If g is bounded on B, then it is clear that Σ is bounded, from its very definition. Then, as a weak* closed subset of some ball, Σ is weak* compact (by Cor. 3.15). If, conversely, Σ is bounded, then $g = H_\Sigma$ is evidently bounded on the unit ball B. □

4.26 Corollary. *Let $f : X \to \mathbb{R}_\infty$ be a convex function which is Lipschitz near a point x. Then $f'(x;\cdot)$ is the support function of $\partial f(x)$:*

$$f'(x;v) = \max_{\zeta \in \partial f(x)} \langle \zeta, v \rangle \;\; \forall v \in X,$$

and f is Gâteaux differentiable at x if and only if $\partial f(x)$ is a singleton.

Proof. Consider the function $g(v) = f'(x;v)$. We invoke the convexity of f to write

$$f\big(x+\lambda[(1-t)v+tw]\big) \leqslant (1-t)f(x+\lambda v)+tf(x+\lambda w),$$

from which we deduce that $g((1-t)v+tw) \leqslant (1-t)g(v)+tg(w)$. Thus g is convex, and we see without difficulty that g is positively homogeneous and (from the Lipschitz condition) satisfies $|g(v)| \leqslant K\|v\| \;\; \forall v$. It follows that g is subadditive and continuous. Thus, g satisfies the hypotheses of Theorem 4.25. In light of Prop. 4.3, this implies the stated relation between $f'(x;\cdot)$ and $\partial f(x)$.

The last assertion of the corollary is now apparent, in view of Cor. 4.4. □

We remark that in more general circumstances than the above, the reduction of $\partial f(x)$ to a singleton does *not* imply Gâteaux differentiability; see Exer. 8.21.

4.27 Exercise. (Subdifferential inversion) Let $f \in \Gamma(X)$. For $\zeta \in \operatorname{dom} f^*$, the subdifferential $\partial f^*(\zeta)$ consists, by definition, of the points $x \in X$ such that

$$f^*(\xi) - f^*(\zeta) \geqslant \langle \xi - \zeta, x \rangle \;\; \forall \xi \in X^*.$$

Prove that

$$\zeta \in \partial f(x) \iff f(x)+f^*(\zeta) = \langle \zeta, x \rangle \iff x \in \partial f^*(\zeta). \qquad \square$$

4.28 Exercise. Prove that the points at which a function $f \in \Gamma(X)$ attains its minimum are those in $\partial f^*(0)$. □

4.3 Polarity

The geometric counterpart of conjugacy is *polarity*, an operation that plays an important role in the study of tangents and normals. That happens to be the context in which the reader has already made its acquaintance, in §1.4, in connection with defining the normal cone.

Let A be a subset of a normed space X. The **polar cone** of A (or, more simply, the polar), denoted $A^{\triangle}$, is defined by

$$A^{\triangle} = \{ \zeta \in X^* : \langle \zeta, x \rangle \leqslant 0 \ \forall x \in A \}.$$

It follows from the definition that $A^{\triangle}$ is a weak* closed convex cone. In the reverse direction, the polar of a subset Σ of X^* is defined by

$$\Sigma^{\triangle} = \{ x \in X : \langle \sigma, x \rangle \leqslant 0 \ \forall \sigma \in \Sigma \}.$$

The reader is asked to verify that we always have $A^{\triangle\triangle} \supset A$.

4.29 Exercise. Let A and Σ be nonempty cones in X and X^* respectively. Prove that $(I_A)^* = I_{A^{\triangle}}$ and $(I_\Sigma)^* = I_{\Sigma^{\triangle}}$. □

4.30 Proposition. *Let A be a nonempty subset of X. Then A is a closed convex cone if and only if $A^{\triangle\triangle} = A$.*

Proof. The set $A^{\triangle\triangle}$ is always a closed convex cone, by construction. Thus, if it equals A, then A has these properties. Conversely, let A be a closed convex cone. We have $I_A = (I_A)^{**}$ by Theorem 4.21. However, by Exer. 4.29, we also have

$$(I_A)^{**} = (I_{A^{\triangle}})^* = I_{A^{\triangle\triangle}}.$$

Thus, $I_A = I_{A^{\triangle\triangle}}$, whence $A^{\triangle\triangle} = A$. □

4.31 Corollary. *Let A be a nonempty subset of X. Then the bipolar $A^{\triangle\triangle}$ of A is the smallest closed convex cone containing A.*

Proof. If K is a closed convex cone such that $K \supset A$, then, by the proposition, we deduce $K = K^{\triangle\triangle} \supset A^{\triangle\triangle}$. □

4.32 Corollary. *Let $x \in S$, where S is a convex subset of a normed space X. Then the tangent and normal cones at x are mutually polar*:

$$T_S(x) = N_S(x)^{\triangle}, \quad N_S(x) = T_S(x)^{\triangle}.$$

Proof. The second formula holds by definition. The first one then follows from Prop. 4.30, since $T_S(x)$ is (closed and) convex when S is convex (Prop. 2.9). □

We proceed now to focus on cones in X^*.

4.33 Exercise. Let Σ be a nonempty cone in X^*. Prove that $H_\Sigma = I_{\Sigma^{\triangle}}$. □

For bipolarity relative to X^*, the weak* topology plays a role once again, in allowing us to refer back to X:

4.34 Proposition. *Let Σ be a nonempty subset of X^*. Then Σ is a weak*closed convex cone if and only if $\Sigma^{\triangle\triangle} = \Sigma$.*

Proof. If $\Sigma^{\triangle\triangle} = \Sigma$, then it follows that Σ is a weak*closed convex cone, since $\Sigma^{\triangle\triangle}$ always has these properties. Let us now suppose that Σ is a weak*closed convex cone; we proceed to prove that it coincides with its bipolar. We know that $\Sigma \subset \Sigma^{\triangle\triangle}$. For purposes of obtaining a contradiction, assume that the opposite inclusion fails, and let $\zeta \in \Sigma^{\triangle\triangle} \setminus \Sigma$. By Theorem 3.2, there exists $x \in X$ such that

$$\langle \zeta, x \rangle > \langle \sigma, x \rangle \ \forall \sigma \in \Sigma.$$

(This is where the weak*closedness of Σ is used.) Since Σ is a cone, we get

$$\langle \zeta, x \rangle > 0 \geqslant \langle \sigma, x \rangle \ \forall \sigma \in \Sigma.$$

It follows that $x \in \Sigma^{\triangle}$ and thus $\zeta \notin \Sigma^{\triangle\triangle}$, the required contradiction. □

4.35 Exercise. Find $A^{\triangle}$ and $A^{\triangle\triangle}$, when $A \subset \mathbb{R}^2$ consists of the two points $(1,0)$ and $(1,1)$. □

4.4 The minimax theorem

Given a function $f(u,v)$ of two variables, its infimum can be calculated either jointly or successively, in either order:

$$\inf_u \inf_v f(u,v) = \inf_v \inf_u f(u,v) = \inf_{u,v} f(u,v).$$

When a supremum with respect to one of the variables is involved, however, the inf sup and the sup inf will differ in general. The following theorem (said to be of *minimax* type) gives conditions under which equality does hold. The first of many such results, due to von Neumann (see Exer. 4.37 below), figured prominently in game theory; they have since become a useful tool in analysis.

4.36 Theorem. (Ky Fan) *Let U and V be nonempty convex sets in (possibly different) vector spaces. Let $f \colon U \times V \to \mathbb{R}$ be a function such that*

$$u \mapsto f(u,v) \text{ is convex on } U \ \ \forall v \in V, \text{ and } v \mapsto f(u,v) \text{ is concave on } V \ \ \forall u \in U.$$

Suppose in addition that there is a topology on U for which U is compact and $f(\cdot,v)$ is lsc for each $v \in V$. Then we have

$$\sup_{v \in V} \min_{u \in U} f(u,v) = \min_{u \in U} \sup_{v \in V} f(u,v), \text{ where the case } \infty = \infty \text{ is admitted.}$$

Proof. We set $\alpha =$ the sup min, $\beta =$ the min sup. Since taking upper envelopes preserves lower semicontinuity, the hypotheses authorize the use of "min" here. It is easy to see that the inequality $-\infty < \alpha \leqslant \beta$ always holds; we may therefore restrict attention to the case $\alpha < \infty$. We now suppose that $\alpha < \beta$, and we derive a contradiction.

We claim that the sets $U(v) = \{u \in U : f(u,v) \leqslant \alpha\}$ satisfy $\bigcap_{v\in V} U(v) = \emptyset$. If this were not the case, then there would be a point $\bar{u}$ common to them all, so that $f(\bar{u},v) \leqslant \alpha \;\forall\, v \in V$. But then $\beta \leqslant \sup_v f(\bar{u},v) \leqslant \alpha < \beta$, absurd. Since the sets $U(v)$ are closed subsets of the compact space U, the finite intersection property must fail. Thus, there exist $v_1,\dots,v_n \in V$ such that $\bigcap_1^n U(v_i) = \emptyset$. Consider now the set

$$E = \big\{ x \in \mathbb{R}^n : \exists\, u \in U,\ r_i \geqslant 0 \text{ such that } x_i = f(u,v_i) + r_i,\ i = 1,2\dots,n \big\}.$$

It is easy to see that E is convex. Using the compactness of U and the lower semicontinuity of each function $f(\cdot,v_i)$, it is an exercise to show that the complement of E is open (that is, E is closed). We claim that E does not contain the point $p := (\alpha,\alpha,\dots,\alpha)$. For if it did, there would exist $u \in U$ and $r_i \geqslant 0$ such that

$$\alpha = f(u,v_i) + r_i \;\;\forall\, i.$$

But then $u \in \bigcap_1^n U(v_i)$, a contradiction. This proves the claim, and allows us to invoke Theorem 2.37 to separate $\{p\}$ and E. There results a vector $\zeta \in \mathbb{R}^n$ and a scalar γ such that

$$\zeta \bullet p < \gamma < \sum_{i=1}^n \zeta_i \big(f(u,v_i) + r_i\big) \;\;\forall\, u \in U,\ r_i \geqslant 0.$$

It follows that ζ is nonzero and has nonnegative components. We may normalize to arrange $\sum_1^n \zeta_i = 1$. Then the point $\bar{v} = \sum_1^n \zeta_i v_i$ belongs to V, and the previous inequality, combined with the concavity of f with respect to v, implies

$$\alpha = \zeta \bullet p < \min_{u\in U} f(u,\bar{v}) \leqslant \alpha.$$

This contradiction completes the proof. □

4.37 Exercise. Let M be an $m\times n$ matrix, and let S, T be closed, convex, nonempty subsets of $\mathbb{R}^m$ and $\mathbb{R}^n$ respectively, at least one of which is bounded. Then

$$\inf_{x\in S} \sup_{y\in T} \langle x, My \rangle = \sup_{y\in T} \inf_{x\in S} \langle x, My \rangle .$$

□

4.38 Exercise. Let X be a normed space, and let $g : X \to \mathbb{R}_\infty$ be a convex function, $x_0 \in X$, and $k > 0$. Then

$$\inf_{x\in X} \max_{\zeta\in kB_*} g(x) + \langle \zeta, x - x_0 \rangle = \max_{\zeta\in kB_*} \inf_{x\in X} g(x) + \langle \zeta, x - x_0 \rangle .$$

□