Chapter 9

Convex duality and optimization

In this chapter, unless otherwise specified, $(V, \|\cdot\|_V)$ is a general normed linear space with topological dual V^*. For any $v \in V$ and $v^* \in V^*$, we write $v^*(v) = \langle v^*, v\rangle_{(V^*,V)}$. Recall that V^* is a Banach space when equipped with the dual norm

$$\|v^*\|_{V^*} = \sup\{\langle v^*, v\rangle \; : \; \|v\|_V \leq 1\}.$$

Without ambiguity, for simplicity of notation, we write $\|\cdot\|$ instead of $\|\cdot\|_V$, $\|\cdot\|_*$ instead of $\|\cdot\|_{V^*}$ and $\langle v^*, v\rangle$ instead of $\langle v^*, v\rangle_{(V^*,V)}$.

9.1 ▪ Dual representation of convex sets

We know that several basic geometrical objects in a normed linear space V can be described by using continuous linear forms, i.e., elements of the topological dual space. For example, a closed hyperplane H can be written

$$H = \{v \in V \; : \; \langle v^*, v\rangle = \alpha\}$$

for some $v^* \in V^*$, $v^* \neq 0$, and $\alpha \in \mathbf{R}$. Similarly, a closed half-space $\mathscr{H}$ can be written

$$\mathscr{H} = \{v \in V \; : \; \langle v^*, v\rangle \leq \alpha\}.$$

Intersections of finite collections of closed half-spaces yield convex polyhedra. Indeed, we are going to show that arbitrary closed convex sets in V can be described by using only linear continuous forms. This is what we call a dual representation. This theory is based on the Hahn–Banach theorem (which we stated in Theorem 3.3.1), which is formulated below.

Theorem 9.1.1. *Let C be a nonempty closed convex subset of a normed linear space V. Then, each point $u \notin C$ can be strongly separated from C by a closed hyperplane, which means*

$$\exists u^* \in V^*,\ u^* \neq 0,\ \exists \alpha \in \mathbf{R} \text{ such that } \langle u^*, u\rangle > \alpha \text{ and } \langle u^*, v\rangle \leq \alpha \ \forall\, v \in C.$$

From a geometrical point of view, this means that C is contained in the closed half-space

$$\mathscr{H}_{\{u^* \leq \alpha\}} := \Big\{v \in V \; : \; \langle u^*, v\rangle \leq \alpha\Big\},$$

whereas u is in the complement: $\mathscr{H}_{\{u^* > \alpha\}} := \{v \in V \; : \; \langle u^*, v\rangle > \alpha\}$.

PROOF. Let us give the proof of Theorem 9.1.1 when V is a Hilbert space. In that case, one can give a constructive proof relying on the projection theorem on a closed convex subset. (In the general case of a normed linear space V, one can use the analytic version of the Hahn–Banach theorem, which itself is a consequence of the Zorn lemma.)

Let us denote by $P_C(u)$ the projection of u on C. It is characterized by the angle condition (optimality condition)

$$\begin{cases} \langle u - P_C(u), v - P_C(u)\rangle \leq 0 \qquad \forall\, v \in C, \\ P_C(u) \in C. \end{cases}$$

Set $z := u - P_C(u)$. Since $u \notin C$, we have $z \neq 0$ and we can rewrite the above inequality in the following form:

$$\sup_{v\in C}\langle z, v\rangle \leq \langle z, P_C(u)\rangle. \tag{9.1}$$

On the other hand, by definition of z and since $z \neq 0$

$$\begin{aligned} 0 < |z|^2 &= \langle z, z\rangle \\ &= \langle z, u\rangle - \langle z, P_C(u)\rangle, \end{aligned}$$

which implies

$$\langle z, P_C(u)\rangle < \langle z, u\rangle. \tag{9.2}$$

Take $\alpha := \langle z, P_C(u)\rangle$. Combining (9.1) and (9.2) we obtain

$$\sup_{v\in C}\langle z, v\rangle \leq \alpha < \langle z, u\rangle,$$

i.e., $C \subset \mathcal{H}_{\{\langle z,.\rangle \leq \alpha\}}$ and $u \in \mathcal{H}_{\{\langle z,.\rangle > \alpha\}}$. $\square$

As a direct consequence of Theorem 9.1.1 we obtain the following corollary.

Corollary 9.1.1. *Let C be a nonempty closed convex subset of a normed linear space V. Then C is equal to the intersection of all closed half-spaces that contain it:*

$$C = \bigcap_{C\subset \mathcal{H}_{\{v^*\leq\alpha\}}} \mathcal{H}_{\{v^*\leq\alpha\}}.$$

PROOF. Let us denote by $\mathcal{F}$ the set

$$\mathcal{F} = \Big\{(v^*, \alpha) \in V^* \times \mathbf{R} \,:\, C \subset \mathcal{H}_{\{v^*\leq\alpha\}}\Big\}.$$

Clearly $C \subset \bigcap_{(v^*,\alpha)\in\mathcal{F}} \mathcal{H}_{\{v^*\leq\alpha\}}$. Let us prove the converse inclusion $\bigcap_{(v^*,\alpha)\in\mathcal{F}} \mathcal{H}_{\{v^*\leq\alpha\}} \subset C$. By taking the complement, this is equivalent to proving

$$V \setminus C \subset \bigcup_{(v^*,\alpha)\in\mathcal{F}} \Big(V \setminus \mathcal{H}_{\{v^*\leq\alpha\}}\Big),$$

which is precisely the conclusion of the Hahn–Banach separation theorem, Theorem 9.1.1. $\square$

Among closed convex sets, an important subclass is obtained by taking the intersection of a finite number of closed half-spaces.

Definition 9.1.1. *A closed convex polyhedron P is an intersection of finitely many closed half-spaces: in other words, there exist $v_1^*, \dots, v_k^* \in V^*$ with $v_i^* \neq 0$ and $\alpha_1, \dots, \alpha_k \in \mathbf{R}$ such that*

$$P = \left\{ v \in V : \langle v_i^*, v \rangle \leq \alpha_i \text{ for } i = 1, \dots, k \right\}.$$

In the representation of closed convex sets as the intersection of closed half-spaces, it is natural to look for the simplest representation. To that end, let us observe the following elementary facts:

(a) $\alpha' \geq \alpha$ and $C \subset \mathcal{H}_{\{v^* \leq \alpha\}} \Longrightarrow C \subset \mathcal{H}_{\{v^* \leq \alpha'\}}$;

(b) fixing $v^* \neq 0$ and making α vary provides parallel hyperplanes.

From Corollary 9.1.1 and the above observations, we deduce

$$C = \bigcap_{v^* \in V^*,\ v^* \neq 0} \ \bigcap_{\{\exists \alpha \in \mathbf{R}\,:\, C \subset \mathcal{H}_{\{v^* \leq \alpha\}}\}} \mathcal{H}_{\{v^* \leq \alpha\}}. \tag{9.3}$$

The question we have to examine is to describe, for a given $v^* \in V^*$, $v^* \neq 0$, such that there exists some $\alpha \in \mathbf{R}$ with $C \subset \mathcal{H}_{\{v^* \leq \alpha\}}$, what is the intersection of all the parallel half-spaces $\mathcal{H}_{\{v^* \leq \alpha\}}$ which contain C. The answer to this question gives rise to the notion of support function.

Proposition 9.1.1. *For any $v^* \in V^*$, $v^* \neq 0$, such that $C \subset \mathcal{H}_{\{v^* \leq \alpha\}}$ for some $\alpha \in \mathbf{R}$ we have*

$$\bigcap_{\{\alpha\,:\, C \subset \mathcal{H}_{\{v^* \leq \alpha\}}\}} \mathcal{H}_{\{v^* \leq \alpha\}} = \mathcal{H}_{\{v^* \leq \sigma_C(v^*)\}},$$

where $\sigma_C(v^*) := \sup\left\{ \langle v^*, v \rangle \ :\ v \in C \right\}$. In other words, for any given $v^* \in V^*$, $v^* \neq 0$, such that $C \subset \mathcal{H}_{\{v^* \leq \alpha\}}$ for some $\alpha \in \mathbf{R}$, the intersection of all the "parallel" closed half-spaces $\mathcal{H}_{\{v^* \leq \alpha\}}$ containing C is the closed half-space $\mathcal{H}_{\{v^* \leq \sigma_C(v^*)\}}$, where $\sigma_C(v^*)$ is defined as above.

It is convenient to extend the definition of σ_C to an arbitrary $v^* \in V^*$ by allowing it to take the value $+\infty$.

Definition 9.1.2. *For any subset C of V, the function $\sigma_C : V^* \to \mathbf{R} \cup \{+\infty\}$ defined by*

$$\sigma_C(v^*) = \sup\left\{ \langle v^*, v \rangle \ :\ v \in C \right\}$$

is called the support function of the set C.

Proof of Proposition 9.1.1. (a) For any $v \in C$, by definition of σ_C, we have

$$\langle v^*, v \rangle \leq \sigma_C(v^*).$$

Hence, $C \subset \mathcal{H}_{\{v^* \leq \sigma_C(v^*)\}}$, which clearly implies

$$\bigcap_{\{\alpha\,:\, C \subset \mathcal{H}_{\{v^* \leq \alpha\}}\}} \mathcal{H}_{\{v^* \leq \alpha\}} \subset \mathcal{H}_{\{v^* \leq \sigma_C(v^*)\}}.$$

(b) For any $\alpha \in \mathbf{R}$ such that $C \subset \mathcal{H}_{\{v^* \leq \alpha\}}$, we have

$$\alpha \geq \sup\left\{ \langle v^*, v \rangle \ :\ v \in C \right\} = \sigma_C(v^*).$$

Hence, $\mathscr{H}_{\{v^* \le \sigma_C(v^*)\}} \subset \mathscr{H}_{\{v^* \le \alpha\}}$ and

$$\mathscr{H}_{\{v^* \le \sigma_C(v^*)\}} \subset \bigcap_{\{\alpha \,:\, C \subset \mathscr{H}_{\{v^* \le \alpha\}}\}} \mathscr{H}_{\{v^* \le \alpha\}},$$

which completes the proof. □

As a direct consequence of formula (9.3) and Proposition 9.1.1 we obtain the following important result.

Theorem 9.1.2. *Let C be a nonempty closed convex subset of a normed linear space V. Then*

$$C = \bigcap_{v^* \in V^*,\ v^* \neq 0} \mathscr{H}_{\{v^* \le \sigma_C(v^*)\}},$$

where σ_C is the support function of C. Equivalently,

$$C = \left\{ v \in V \ :\ \langle v^*, v\rangle \le \sigma_C(v^*)\ \forall\ v^* \in V^* \right\}.$$

Remark 9.1.1. The dual representation of a closed convex set C has been obtained with the help of the support function $\sigma_C : V^* \to \mathbf{R} \cup \{+\infty\}$. As we will see in this chapter, the mapping $C \mapsto \sigma_C$ can be viewed as a particular case of the general duality correspondence, namely, the Legendre–Fenchel transform $f \mapsto f^*$. More precisely, by taking $f = \delta_C$ the indicator of C, we have $f^* = \sigma_C$. We examine below the properties of σ_C which are direct consequences of its definition.

Proposition 9.1.2. *The support function $\sigma_C : V^* \to \mathbf{R} \cup \{+\infty\}$ of a closed convex nonempty subset C is a function which is closed, convex, proper, and positively homogeneous of degree* 1.

PROOF. For any $v \in C$, the mapping

$$v^* \in V^* \mapsto \langle v^*, v\rangle$$

is a linear continuous form on V^*, hence convex and continuous. The function σ_C as a supremum of convex functions is still convex and, as a supremum of continuous functions, it is closed (lower semicontinuous); see Proposition 3.2.3. Moreover, $\sigma_C(0) = 0$ and σ_C is proper. Finally, for any $v^* \in V^*$ and $t > 0$ we have

$$\begin{aligned} \sigma_C(tv^*) &= \sup \left\{ t\langle v^*, v\rangle \ :\ v \in C \right\} \\ &= t \sup \left\{ \langle v^*, v\rangle \ :\ v \in C \right\} \\ &= t\,\sigma_C(v^*), \end{aligned}$$

which expresses that σ_c is positively homogeneous of degree 1. □

To have a sharper view of the dual generation of closed convex sets, it is interesting to introduce the notion of supporting hyperplane. This notion is closely related to the question, In the definition of $\sigma_C(v^*) = \sup \left\{ \langle v^*, v\rangle \ :\ v \in C \right\}$, is the supremum attained?

Definition 9.1.3. *An element $v^* \in V^*$, $v^* \neq 0$, is said to support C at a point $u \in C$ if*

$$\begin{aligned} \sigma_C(v^*) &= \langle v^*, u\rangle \\ &= \sup \left\{ \langle v^*, v\rangle \ :\ v \in C \right\}. \end{aligned}$$

An equivalent terminology consists in saying that v^* is a supporting functional of C at $u \in C$.

The geometric terminology above comes from the fact that when v^* supports C at $u \in C$ we have that the closed half-space

$$\mathscr{H}_{\{v^* \le \sigma_C(v^*)\}}$$

contains C and that the corresponding hyperplane

$$H = \left\{ v \in V \ : \ \langle v^*, v \rangle = \sigma_C(v^*) \right\}$$

intersects C at u. (Note that the intersection of H with C may contain some other points.)

An interesting question is to know whether it is possible to obtain a dual representation of closed convex sets by supporting functionals. As we will see, this is a quite involved question which is intimately connected with the properties of the subdifferential of a closed convex function and the Bishop–Phelps theorem (density properties of the domain of the subdifferential).

Let us end this section with some elementary examples illustrating the concept of support function.

Example 9.1.1. (1) Take $C = \mathbf{B}(0,1)$ the unit ball of V. Then, for any $v^* \in V^*$

$$\begin{aligned} \sigma_C(v^*) &= \sup \left\{ \langle v^*, v \rangle \ : \ \|v\|_V \le 1 \right\} \\ &= \|v^*\|_{V^*}, \end{aligned}$$

i.e., σ_C is the dual norm $\|\cdot\|_{V^*}$ of $\|\cdot\|_V$.

(2) Take C as a cone, i.e., $\lambda v \in C$ for all $v \in C$ and $\lambda \ge 0$ (note that necessarily $0 \in C$). Let us assume moreover that C is closed and convex. Then

$$\begin{aligned} \sigma_C(v^*) &= \sup \left\{ \langle v^*, v \rangle \ : \ v \in C \right\} \\ &= \begin{cases} 0 & \text{whenever } \langle v^*, v \rangle \le 0 \ \forall v \in C, \\ +\infty & \text{otherwise.} \end{cases} \end{aligned}$$

Let us notice that the set

$$C^* = \left\{ v^* \in V^* \ : \ \langle v^*, v \rangle \le 0 \text{ for all } v \in C \right\}$$

is a closed convex cone; it is called the polar cone of C. We have that σ_C is equal to the indicator function of this polar cone

$$\sigma_C = \delta_{C^*}.$$

9.2 ▪ Passing from sets to functions: Elements of epigraphical calculus

Our next goal is to apply the dual representation Theorem 9.1.2 to the set $C = \operatorname{epi} f$, where $f : V \to \mathbf{R} \cup \{+\infty\}$ is a closed convex proper function. So doing, we will obtain by a pure geometrical approach the Legendre–Fenchel duality theory for closed convex functions.

To that end, it will be useful to develop some tools of epigraphical calculus, which consists of viewing functions as sets, via their epigraphs. As stressed in Section 3.2.2, the epigraph of an extended real-valued function is a geometrical object that carries most of the properties of the corresponding variational problems. In our context, given $f : V \to \mathbf{R}\cup\{+\infty\}$, recall that

$$f \text{ is closed (lsc)} \iff \text{epi} f \text{ is closed},$$

$$f \text{ is convex} \iff \text{epi} f \text{ is convex},$$

and that the basic operation in convex analysis and duality which consists in taking the supremum of a family of convex (affine) functions has an immediate epigraphical interpretation

$$\text{epi}\left(\sup_{k\in I} f_k\right) = \bigcap_{k\in I} \text{epi} f_k.$$

Beyond the classical operations on extended real-valued functions (sum and multiplication by a positive scalar) let us introduce the epi-addition, also called inf-convolution.

Definition 9.2.1. *Let V be a linear space and $f, g : V \to \mathbf{R}\cup\{+\infty\}$ two extended real-valued functions. The epi-sum of f and g (also called inf-convolution) is the function*

$$f \#_e g : V \to \overline{\mathbf{R}}$$

defined by

$$\begin{aligned}(f\#_e g)(v) &= \inf\left\{f(v_1)+g(v_2) \,:\, v_1+v_2=v,\ v_1,v_2\in V\right\}\\ &= \inf\left\{f(v-w)+g(w) \,:\, w\in V\right\}\\ &= \inf\left\{f(w)+g(v-w) \,:\, w\in V\right\}.\end{aligned}$$

We often briefly write $f\#g$.

Note that $f\#_e g$ may take the value $-\infty$ (for example, take $g=0$ and f not minorized). The term *epi-sum* comes from the following geometrical interpretation of this operation.

Proposition 9.2.1. *For any $f, g : V \to \mathbf{R}\cup\{+\infty\}$*

$$\text{epi}_S(f\#_e g) = \text{epi}_S f + \text{epi}_S g,$$

where $\text{epi}_S f$ stands for the strict epigraph of f, i.e.,

$$\text{epi}_S f = \left\{(v,\lambda)\in V\times\mathbf{R} \,:\, \lambda > f(v)\right\},$$

and the sum $\text{epi}_S f + \text{epi}_S g$ is the vectorial sum (also called Minkowski sum) of the two sets $\text{epi}_S f$ and $\text{epi}_S g$.

PROOF. We have

$$\lambda > (f\#_e g)(v)$$

iff there exists $v_1, v_2 \in V$, with $v = v_1+v_2$ such that

$$\lambda > f(v_1)+g(v_2).$$

This is clearly equivalent to the existence of $v_1, v_2 \in V$ and $\lambda_1, \lambda_2 \in \mathbf{R}$ such that $\lambda_1 > f(v_1)$, $\lambda_2 > g(v_2)$ and $v = v_1+v_2$, $\lambda = \lambda_1+\lambda_2$. Equivalently, $(\lambda, v) = (\lambda_1, v_1)+(\lambda_2, v_2)$ with $(\lambda_1, v_1)\in \text{epi}_S f$ and $(\lambda_2, v_2)\in \text{epi}_S g$. $\square$

Remark 9.2.1. The term *inf-convolution* refers to the (formal) similarities of this operation with the usual convolution of functions on $\mathbf{R}^N$

$$(f * g)(x) = \int_{\mathbf{R}^N} f(x-y)g(y)\,dy,$$

where one has to replace $\int_{\mathbf{R}^N}$ by inf and product by addition. As we will see, there are many striking similarities between these two operations.

Proposition 9.2.2. *Let $f, g : V \to \mathbf{R} \cup \{+\infty\}$ be two convex functions. Then, their epi-sum $f \#_e g$ is still a convex function.*

PROOF. This property is a clear consequence of the geometrical interpretation of the epi-sum via epigraphs (Proposition 9.2.1)

$$\mathrm{epi}_S(f \#_e g) = \mathrm{epi}_S f + \mathrm{epi}_S g$$

and of the fact that the Minkowski (vectorial) sum of two convex sets is still convex: indeed, given C and D two convex subsets of a vector space E, consider two points of $C+D$, let $v_1 = c_1 + d_1$, $v_2 = c_2 + d_2$ with $c_i \in C$ and $d_i \in D$ $(i = 1, 2)$. For any $0 \le \lambda \le 1$, one has

$$\begin{aligned}\lambda v_1 + (1-\lambda)v_2 &= \lambda(c_1 + d_1) + (1-\lambda)(c_2 + d_2)\\ &= \big(\lambda c_1 + (1-\lambda)c_2\big) + \big(\lambda d_1 + (1-\lambda)d_2\big),\end{aligned}$$

which still belongs to $C+D$. One can then easily verify that the convexity of an extended real-valued function is equivalent to the convexity of its strict epigraph. □

The fact that the epi-sum preserves the convexity, as we have just observed, follows clearly from its geometrical interpretation. On the other hand, it is somewhat surprising from the analytical point of view, since $(f \#_e g)(v)$ is expressed as an infimum of convex functions, namely, $v \mapsto f(v-w) + g(w)$, and the class of convex functions is not stable by infimal operations. This calls for some explanation.

Indeed, the convexity of $f \#_e g$ when f and g are convex functions is a consequence of the observation "the function of two variables $(v, w) \mapsto h(v, w) := f(v-w) + g(w)$ is convex with respect to the pair (v, w)" and of the following proposition.

Proposition 9.2.3. *Let V and W be two linear spaces and $h : V \times W \to \mathbf{R} \cup \{+\infty\}$ be a convex function. Then, the function $p : V \to \overline{\mathbf{R}}$ defined by*

$$p(v) = \inf_{w \in W} h(v, w)$$

is still convex.

PROOF. Let us prove that for any $u, v \in V$ and $\lambda \in\,]0, 1[$

$$p(\lambda u + (1-\lambda)v) \le \lambda p(u) + (1-\lambda)p(v).$$

Without any restriction, we can assume $p(u) < +\infty$ and $p(v) < +\infty$; otherwise the inequality is trivially satisfied. Take arbitrary $s > p(u)$ and $t > p(v)$. By definition of p, one can find elements $w_{u,s}$ and $w_{v,t}$ in W such that

$$s > h(u, w_{u,s}) \text{ and } t > h(v, w_{v,t}).$$

By convexity of $h(\cdot,\cdot)$ with respect to the couple of variables (v,w)

$$\begin{aligned} h(\lambda u+(1-\lambda)v, \lambda w_{u,s}+(1-\lambda)w_{v,t}) &\le \lambda h(u,w_{u,s})+(1-\lambda)h(v,w_{v,t}) \\ &\le \lambda s+(1-\lambda)t. \end{aligned}$$

By definition of p

$$p(\lambda u+(1-\lambda)v) \le h(\lambda u+(1-\lambda)v, \lambda w_{u,s}+(1-\lambda)w_{v,t}).$$

We combine the two above inequalities to obtain

$$p(\lambda u+(1-\lambda)v) \le \lambda s+(1-\lambda)t.$$

This being true for any $s > p(u)$ and $t > p(v)$, by letting s tend to $p(u)$ and t tend to $p(v)$, we obtain the required convexity inequality. □

We will see that the epi-sum (inf-convolution) is the dual operation of the usual sum. The epi-sum plays also an important role in the regularization of lower semicontinuous extended real-valued functions. The following theorem has a long history (it goes back to Hausdorff, Pasch, and Baire and has been revisited by many authors).

Theorem 9.2.1 (Lipschitz regularization via epi-sum). *Let $(V,\|\cdot\|)$ be a normed space and let $f : V \to \mathbf{R}\cup\{+\infty\}$ be a proper and lower semicontinuous function. Suppose moreover that f is conically minorized, i.e., there exists some $k_0 \ge 0$ such that for all $v \in V$*

$$f(v) \ge -k_0(1+\|v\|).$$

Let us define for all $k \in \mathbf{R}^+$ the function $f_k := f\#_e k\|\cdot\|$, i.e.,

$$f_k(v) = \inf_{w\in V}\left\{f(w)+k\|v-w\|\right\}.$$

Then we have

(a) *for all $k \ge k_0$, f_k is Lipschitz continuous on V with constant k, that is, for all $u,v \in V$*

$$|f_k(u)-f_k(v)| \le k\|u-v\|;$$

(b) *for all $v \in V$ one has*

$$f(v) = \lim_{k\to+\infty} f_k(v).$$

More precisely, the sequence $(f_k)_k$ monotonically increases to f as $k \uparrow \infty$.

(c) *When f is convex, so is f_k for all $k \ge k_0$.*

PROOF. (a) For all $k \ge k_0$, we have

$$\begin{aligned} f_k(v) &\ge \inf_{w\in V}\left\{-k_0-k_0\|w\|+k\|v-w\|\right\} \\ &\ge \inf_{w\in V}\left\{-k_0-k_0\|w\|+k\|w\|-k\|v\|\right\} \\ &\ge -k_0-k\|v\| > -\infty. \end{aligned}$$

On the other hand, taking some $w_0 \in \operatorname{dom} f \ne \emptyset$ (f is proper)

$$f_k(v) \le f(w_0)+k\|v-w_0\| < +\infty.$$

Hence, for all $k \geq k_0$ and all $v \in V$, $f_k(v)$ is a real number. Take now $u, v \in V$. The triangle inequality yields for any $w \in V$

$$\|v - w\| \leq \|u - w\| + \|v - u\|.$$

Hence, for all $w \in V$, for all $k \in \mathbf{R}^+$

$$f(w) + k\|v - w\| \leq f(w) + k\|u - w\| + k\|v - u\|.$$

Taking the infimum with respect to $w \in V$ yields

$$f_k(v) \leq f_k(u) + k\|v - u\|.$$

Exchanging the role of v and u and noticing that for $k \geq k_0$ both $f_k(v)$ and $f_k(u)$ are finitely valued yields

$$|f_k(v) - f_k(u)| \leq k\|v - u\|.$$

(b) By taking $w = v$ in the definition of $f_k(v)$, one has

$$f_k(v) \leq f(v).$$

Clearly, the sequence $(f_k)_k$ is increasing with respect to k. Hence

$$\lim_{k \to +\infty} f_k(v) \leq f(v).$$

Let us prove the reverse inequality

$$f(v) \leq \lim_{k \to +\infty} f_k(v).$$

If $\lim_{k \to +\infty} f_k(v) = +\infty$, there is nothing to prove. So let us assume that $\lim_{k \to +\infty} f_k(v) < +\infty$. For each $k \geq k_0$, let us introduce some $w_k \in V$ such that

$$f_k(v) \geq f(w_k) + k\|v - w_k\| - \varepsilon_k$$

for some $\varepsilon_k > 0$ with $\varepsilon_k \to 0$ as $k \to +\infty$. Using the growth condition on f we obtain

$$+\infty > \sup_{k > 0} f_k(v) \geq -k_0(1 + \|w_k\|) + k\|v - w_k\| - \varepsilon_k,$$

which clearly implies $w_k \to v$ in $(V, \|\cdot\|)$ as $k \to +\infty$. Let us now pass to the limit on the inequality

$$f_k(v) \geq f(w_k) - \varepsilon_k$$

and use the lower semicontinuity of f to obtain

$$\begin{aligned} \lim_{k \to +\infty} f_k(v) &\geq \liminf_k f(w_k) \\ &\geq f(v). \end{aligned}$$

(c) Noticing that f and $k\|\cdot\|$ are both convex functions, the convexity of $f_k = f \#_e k\|\cdot\|$ is a straightforward consequence of Proposition 9.2.2. $\square$

Let us end this section with the following striking property of closed convex functions. We know that a linear operator from a normed space into another normed space is continuous iff it is bounded on bounded sets. This is an important property since it reduces the study of continuity of a linear operator $A : E \to F$ to the establishment of majorizations of the following type: there exists some $M \in \mathbf{R}^+$ such that

$$||v||_E \leq 1 \Longrightarrow ||Av||_F \leq M.$$

We will prove in Theorem 9.3.1 that any closed convex function is the supremum of all its continuous affine minorants.

Thus, it is not surprising that also for closed convex functions local boundedness implies continuity. Let us make this precise in the following statement.

Theorem 9.2.2. *Let $(V, ||\cdot||)$ be a normed space and let $f : V \to \mathbf{R} \cup \{+\infty\}$ be a convex function which is majorized on a neighborhood of a point $v_0 \in \operatorname{dom} f$, i.e.,*

$$\exists r > 0 \text{ such that } \sup_{||v-v_0||<r} f(v) := M < +\infty.$$

Then f is continuous at the point v_0. More precisely, f is Lipschitz continuous on all balls $\mathbf{B}(v_0, r')$ with $r' < r$ and

$$|f(w) - f(v)| \leq \frac{2(M + |f(v_0)|)}{r - r'} ||w - v|| \quad \forall v, w \in \mathbf{B}(v_0, r').$$

PROOF. (a) Let us first prove that f is continuous at v_0. By translation (consider the function $f(v + v_0) - f(v_0)$), one can reduce the problem to the case $v_0 = 0$ and $f(0) = 0$. Take an arbitrary ε such that $1 \geq \varepsilon > 0$ and observe that for any $v \in \mathbf{B}(0, r\varepsilon)$, the following convex inequalities hold:

writing $v = (1-\varepsilon)0 + \varepsilon\left(\frac{1}{\varepsilon}v\right)$ we have

$$f(v) \leq (1-\varepsilon)f(0) + \varepsilon f\left(\frac{1}{\varepsilon}v\right) \leq \varepsilon M\,;$$

writing $0 = \frac{1}{1+\varepsilon}v + \frac{\varepsilon}{1+\varepsilon}\left(\frac{-1}{\varepsilon}v\right)$ we have

$$0 = f(0) \leq \frac{1}{1+\varepsilon}f(v) + \frac{\varepsilon}{1+\varepsilon}f\left(\frac{-1}{\varepsilon}v\right) \leq \frac{1}{1+\varepsilon}f(v) + \frac{\varepsilon M}{1+\varepsilon},$$

which yields

$$f(v) \geq -\varepsilon M.$$

Combining the two above inequalities, we obtain

$$|f(v)| \leq \varepsilon M \text{ for } v \in \mathbf{B}(0, r\varepsilon),$$

which yields the continuity of f at the origin.

(b) First observe that in the above argument, when taking $\varepsilon = 1$ we have the existence of some positive constant, which we still denote by M, such that

$$|f(v + v_0) - f(v_0)| \leq M \qquad \forall\ v \in \mathbf{B}(0, r).$$

Take arbitrary $v, w \in \mathbf{B}(v_0, r')$ with $v \neq w$. Set $\varepsilon = r - r' > 0$ and

$$u = w + \frac{\varepsilon}{||w - v||}(w - v), \qquad \lambda = \frac{||w - v||}{\varepsilon + ||w - v||}.$$

We have $u \in \mathbf{B}(v_0, r)$ and $\|w-v\|u = (\varepsilon + \|w-v\|)w - \varepsilon v$. Equivalently,

$$w = \lambda u + \frac{\varepsilon}{\varepsilon + \|w-v\|} v,$$

$$w = \lambda u + (1-\lambda)v \text{ with } \lambda \in]0,1[.$$

By convexity of f

$$f(w) \le \lambda f(u) + (1-\lambda) f(v) = f(v) + \lambda(f(u) - f(v)),$$

which yields (observe that $|f(u)-f(v)| \le |f(u)-f(v_0)| + |f(v)-f(v_0)| \le 2M$)

$$f(w) - f(v) \le \frac{\|w-v\|}{\varepsilon + \|w-v\|} 2M \le \frac{2M}{\varepsilon}\|w-v\|.$$

Exchanging the role of w and v, we obtain

$$|f(w) - f(v)| \le \frac{2M}{\varepsilon}\|w-v\|,$$

which completes the proof. □

9.3 ▪ Legendre–Fenchel transform

Given $f : V \to \mathbf{R} \cup \{+\infty\}$ a closed convex proper function, we are going to introduce f^*, the Legendre–Fenchel transform of f, by considering the set $C = \operatorname{epi} f \subset V \times \mathbf{R}$ and its dual representation, as given by Theorem 9.1.2. To that end, we need to exploit the particular structure of the set $C = \operatorname{epi} f$ in $V \times \mathbf{R}$ and describe the family of the closed half-spaces in $V \times \mathbf{R}$ containing it.

Let us start with the following elementary result, which describes the closed half-spaces in $V \times \mathbf{R}$.

Lemma 9.3.1. *Let $l \in (V \times \mathbf{R})^*$ be a linear continuous form on $V \times \mathbf{R}$, $l \neq 0$. Then there exist $u^* \in V^*$ and $\gamma \in \mathbf{R}$, $(u^*, \gamma) \neq 0$ such that*

$$l(v,t) = \langle u^*, v\rangle + \gamma t \quad \forall (v,t) \in V \times \mathbf{R}.$$

A closed half-space $\mathscr{H}$ in $V \times \mathbf{R}$ is of the following form:

$$\mathscr{H} = \mathscr{H}_{\{(u^*,\gamma)\le\alpha\}} := \left\{(v,t) \in V \times \mathbf{R} : \langle u^*, v\rangle + \gamma t \le \alpha\right\}.$$

Depending on the value of γ ($\gamma = 0$ or $\gamma \neq 0$), we have two distinct situations:

(a) $\gamma = 0$. *Then $\mathscr{H} = \mathscr{H}_{\{(u^*,\gamma)\le\alpha\}} = \left\{(v,t) \in V \times \mathbf{R} : \langle u^*, v\rangle \le \alpha\right\} = \left\{u^* \le \alpha\right\} \times \mathbf{R}$ is invariant by all the translations parallel to $\left\{0\right\} \times \mathbf{R}$. In that case, we say that the half-space is "vertical."*

(b) $\gamma \neq 0$. *By normalization (divide by $-\gamma$) one can rewrite the closed half-space $\mathscr{H}$ in the form*

$$\mathscr{H} = \left\{(v,t) \in V \times \mathbf{R} : t \ge \langle u^*, v\rangle - \alpha\right\}$$

or

$$\mathscr{H} = \left\{(v,t) \in V \times \mathbf{R} : t \le \langle u^*, v\rangle - \alpha\right\}.$$

It is the epigraph or the hypograph of the affine continuous function

$$v \mapsto \langle u^*, v\rangle - \alpha.$$

Because of its particular structure, the epigraph of a proper function cannot be contained in a half-space of the form $\{(v,t) \in V \times \mathbf{R} \,:\, t \le \langle u^*, v\rangle - \alpha\}$. We can summarize the previous results in the following lemma.

Lemma 9.3.2. *Let $f : V \longrightarrow \mathbf{R} \cup \{+\infty\}$ be a proper function. Then, a closed half-space $\mathscr{H}$ in $V \times \mathbf{R}$ containing the set $C = \operatorname{epi} f$ is either vertical or equal to the epigraph of an affine continuous function, i.e., there exist some $u^* \in V$ and $\alpha \in \mathbf{R}$ such that*

$$\mathscr{H} = \{(v,t) \in V \times \mathbf{R} \,:\, t \ge \langle u^*, v\rangle - \alpha\}.$$

Let us keep in mind that we are looking for the simplest dual representation of convex functions f. In this perspective, it is a striking and important property that one can get rid of the vertical half-spaces in their dual representations. Indeed, this is not a surprising result since one can approach them, and get arbitrarily "close" to closed vertical half-spaces, by epigraphs of continuous affine functions.

Let us make this precise in the following statement.

Theorem 9.3.1. *Let $(V, \|\cdot\|)$ be a normed space and let $f : V \to \mathbf{R} \cup \{+\infty\}$ be a closed convex proper function. Then f is equal to the supremum of all the continuous affine functions which minorize f.*

PROOF. (a) Let us first notice that among all the closed half-spaces containing $C = \operatorname{epi} f$, at least one of them is the epigraph of an affine continuous function. Otherwise, there would be only vertical half-spaces in this family, and C, as an intersection of such sets, would be vertically invariant. This is impossible, because f is proper. Indeed, for any $v_0 \in \operatorname{dom} f$, $C \cap (\{v_0\} \times \mathbf{R}) = \{v_0\} \times [f(v_0), +\infty[$ with $[f(v_0), +\infty[$ strictly included in $\mathbf{R}$. Then notice that

$$\operatorname{epi} f \subset \operatorname{epi}\{v \mapsto \langle u^*, v\rangle - \alpha\}$$

is equivalent to

$$f(v) \ge \langle u^*, v\rangle - \alpha \quad \forall v \in V,$$

i.e., f admits at least an affine continuous minorant.

(b) Let us recall the general property: $f = \sup f_i \iff \operatorname{epi} f = \bigcap_{i \in I} \operatorname{epi} f_i$. Thus, to establish the assertion of the theorem, it suffices to show that each point $(v_0, t_0) \notin \operatorname{epi} f$ is outside the epigraph of an affine continuous function that is majorized by f.

We know that $C = \operatorname{epi} f$ is equal to the intersection of all the closed half-spaces that contain it (Corollary 9.1.1) and that any such half-space either is vertical or is the epigraph of an affine continuous function (Lemma 9.3.2). To eliminate the vertical half-spaces in their dual representation, we use the Lipschitz regularization theorem, Theorem 9.2.1. Since f admits an affine continuous minorant, it is conically minorized and $f = \sup_k f_k$ with f_k convex and Lipschitz continuous.

Since $t_0 < f(v_0)$, for some k_0 sufficiently large $t_0 < f_{k_0}(v_0) \le f(v_0)$ and $(v_0, t_0) \notin \operatorname{epi} f_{k_0}$. Let us now use the dual representation of the closed convex set $\operatorname{epi} f_{k_0}$ as the intersection of all the closed half-spaces that contain it. Since f_{k_0} is everywhere defined, there is no vertical half-space containing $\operatorname{epi} f_{k_0}$. Thus, f_{k_0} is the supremum of all its affine continuous minorants. As a consequence, one can find an affine continuous function

$v \mapsto l(v) = \langle u^*, v\rangle - \alpha$ with

$$(v_0, t_0) \notin \operatorname{epi} l \quad \text{and} \quad l \leq f_{k_0}.$$

Since $f_{k_0} \leq f$ we have $l \leq f$ and $(v_0, t_0) \notin \operatorname{epi} l$, that is, l satisfies all required properties. $\square$

Just like for convex sets, we are going to look for the simplest dual description of closed convex functions, i.e., using the simplest continuous affine minorants. Theorem 9.3.1 tells us that

$$f(v) = \sup\left\{\langle v^*, v\rangle - \alpha \ : \ \langle v^*, v\rangle - \alpha \leq f(v) \ \forall v \in V\right\}.$$

Let us now observe that for $v^* \in V^*$ being fixed, making α vary provides parallel minorizing affine continuous functions. Clearly, the best α is obtained by taking

$$\alpha = \sup\left\{\langle v^*, v\rangle - f(v) \ : \ v \in V\right\},$$

which, for $v^* \in V^*$ being fixed, is a real number iff f admits a continuous affine minorant with slope v^*. This is precisely the quantity which is classically denoted by

$$f^*(v^*) = \sup\left\{\langle v^*, v\rangle - f(v) \ : \ v \in V\right\}$$

and which makes sense for an arbitrary $v^* \in V^*$, with possibly $+\infty$ values.

The above geometrical considerations allow us to reformulate Theorem 9.3.1 in the following form:

$$\forall v \in V \quad f(v) = \sup\left\{\langle v^*, v\rangle - f^*(v^*) \ : \ v^* \in V^*\right\}. \tag{9.4}$$

We are now ready to introduce classical notation, terminology, and basic facts concerning the Legendre–Fenchel transform which is defined below for arbitrary proper function f.

Definition 9.3.1. *Let V be a normed linear space and let $f : V \to \mathbf{R} \cup \{+\infty\}$ be a proper function. The Legendre–Fenchel conjugate of f is the function*

$$f^* : V^* \to \mathbf{R} \cup \{+\infty\}$$

defined by

$$f^*(v^*) = \sup\left\{\langle v^*, v\rangle - f(v) \ : \ v \in V\right\}.$$

Let us notice that since f is proper, by taking some $v_0 \in \operatorname{dom} f$

$$f^*(v^*) \geq \langle v^*, v_0\rangle - f(v_0),$$

i.e., f^* admits an affine continuous minorant and $f^* : V^* \to \mathbf{R} \cup \{+\infty\}$.

Moreover, $v^* \in \operatorname{dom} f^*$ iff there exists $\alpha \in \mathbf{R}$ such that for all $v \in V$ one has $\langle v^*, v\rangle - f(v) \leq \alpha$, i.e.,

$$f(v) \geq \langle v^*, v\rangle - \alpha.$$

Let us now return to the case when f is closed convex and proper. We know that f admits at least one such affine continuous minorant. This implies that f^* is proper. Since f^* is a supremum of continuous affine functions it is a closed convex proper function from V^* into $\mathbf{R}\cup\{+\infty\}$.

Let us examine the two formulas

$$f^*(v^*) = \sup\left\{\langle v^*, v\rangle - f(v) \ : \ v \in V\right\} \quad \text{(definition of } f^*\text{)},$$

$$f(v) = \sup\left\{\langle v^*, v\rangle - f^*(v^*) \ : \ v^* \in V^*\right\} \quad \text{(Theorem 9.3.1)}.$$

They are essentially the same. Let us make this precise.

Since f^* is closed convex and proper we can compute its conjugate $f^{**} : V^{**} \to \mathbf{R}\cup\{+\infty\}$. By using the canonical embedding of V into V^{**}, we can restrict f^{**} to V to obtain

$$\forall v \in V \quad f^{**}(v) = \sup\left\{\langle v^*, v\rangle - f^*(v^*) \ : \ v^* \in V^*\right\}.$$

(Recall that $i : V \to V^{**}$ is defined by $i(v)(v^*) = v^*(v)$.)

The dual representation theorem, Theorem 9.3.1, for closed convex functions then can be reformulated in the following form: $f = f^{**}$. This is the Fenchel–Moreau–Rockafellar theorem that we now state.

Theorem 9.3.2. *Let V be a normed space and let $f : V \to \mathbf{R}\cup\{+\infty\}$ be a closed convex proper function. Then*

$$f = f^{**},$$

i.e., f is equal to its biconjugate. Equivalently,

$$\forall v \in V \quad f(v) = \sup\left\{\langle v^*, v\rangle - f^*(v^*) \ : \ v^* \in V^*\right\}.$$

It is worth noticing that the dual representation of closed convex proper functions has been given, in the above theorem, a quite simple formulation. On the counterpart, it is a precise analytic formulation which may hide the geometrical features of this duality theory. It is good to keep in mind both aspects.

At this point, it is interesting to observe that the duality for functions has been derived from duality for sets (via the representation of $f : V \to \mathbf{R}\cup\{+\infty\}$ by its epigraph $C = \operatorname{epi} f$). Conversely, the duality for sets can be obtained as a particular case of the duality for functions. Let us associate to a set C its indicator function δ_C and first observe that whenever C is a nonempty closed convex set, then δ_C is a closed convex proper function. Then notice that

$$\begin{aligned}(\delta_C)^*(v^*) &= \sup\left\{\langle v^*, v\rangle - \delta_C(v) \ : \ v \in V\right\}\\ &= \sup\left\{\langle v^*, v\rangle \ : \ v \in C\right\}\\ &= \sigma_C(v^*),\end{aligned}$$

i.e., $(\delta_C)^*$ is the support function of C.

The Fenchel–Moreau–Rockafellar theorem, Theorem 9.3.2, says that $(\delta_C)^{**} = \delta_C$, which is equivalent to

$$\delta_C(v) = \sup\left\{\langle v^*, v\rangle - \sigma_C(v^*) \ : \ v^* \in V^*\right\}.$$

Noticing that $v \in C$ iff $\delta_C(v) = 0$ one gets

$$C = \left\{v \in V \ : \ \langle v^*, v\rangle \le \sigma_C(v^*) \ \forall\, v^* \in V^*\right\}.$$

Let us summarize the previous results.

Proposition 9.3.1. *Let C be a nonempty closed convex subset of a normed linear space V; then*

$$(\delta_C)^* = \sigma_C \text{ and } (\sigma_C)^* = \delta_C,$$

that is,

$$C = \left\{ v \in V \ : \ \langle v^*, v \rangle \le \sigma_C(v^*)\, \forall v^* \in V^* \right\}.$$

It is worth noticing that the biconjugate operation $f \mapsto f^{**}$ enjoys nice properties for convex functions which are not necessarily closed.

We recall (see Section 3.2.4) that given (X,τ) a general topological space and $f : X \to \mathbf{R} \cup \{+\infty\}$, $cl_\tau f$ is the largest τ-lsc function that minorizes f. We have

$$\text{epi}(cl_\tau f) = cl(\text{epi}\, f);$$

$cl_\tau f$ is called the lower semicontinuous regularization of f. Moreover (see Proposition 3.2.5(d)), f is τ-lsc at x iff $f(x) = (cl_\tau f)(x)$.

For convex functions $f : V :\to \mathbf{R} \cup \{+\infty\}$, with V a normed space, we have the following elegant characterization of clf (for the topology of the norm of the space V) in terms of the biconjugate f^{**}.

Proposition 9.3.2. *Let V be a normed linear space and let $f : V \to \mathbf{R} \cup \{+\infty\}$ be a convex proper function. Let us assume that f admits a continuous affine minorant. Then the following equality holds:*

$$f^{**} = clf.$$

As a consequence,

$$f \text{ is lower semicontinuous at } u \in V \iff f(u) = f^{**}(u).$$

PROOF. By definition, for any $v \in V$

$$f^{**}(v) = \sup \left\{ \langle v^*, v \rangle - f^*(v^*) \ : \ v^* \in V^* \right\},$$

which implies that f^{**} is the upper envelope of the continuous affine minorants of f. It is a closed (convex) proper function, hence

$$f^{**} \le clf \le f.$$

Let us now observe that clf is still convex, because the epigraph of clf is the closure of epi f which is a convex set. Hence, clf is a closed convex proper function.

Since f^{**} and clf are both closed convex proper functions, we apply Theorem 9.3.2 to obtain

$$(f^{**})^{**} = f^{**},$$

$$(clf)^{**} = clf.$$

The inequality $f^{**} \le clf \le f$ implies, by taking the biconjugate of each term, that

$$(f^{**})^{**} \le (clf)^{**} \le f^{**},$$

i.e.,

$$f^{**} \le clf \le f^{**},$$

and the equality $f^{**} = clf$ follows.

By Proposition 3.2.5(d), for general functions $f : V \to \mathbf{R} \cup \{+\infty\}$ we have the equivalence

$$f \text{ is lower semicontinuous at } u \in V \iff f(u) = cl f(u).$$

As a consequence, when f is proper and convex, we have

$$f \text{ is lower semicontinuous at } u \in V \iff f(u) = f^{**}(u),$$

which completes the proof. □

Example 9.3.1. (1) Take $C = \mathbf{B}(0,1)$. By definition of the dual norm, for any $v^* \in V^*$

$$\sigma_{\mathbf{B}(0,1)}(v^*) = \sup_{v \in \mathbf{B}(0,1)} \langle v^*, v\rangle = \|v^*\|_{V^*}.$$

Thus, $\sigma_{\mathbf{B}(0,1)} = \|\cdot\|_{V^*}$. Conversely the convex duality theorem yields

$$(\|\cdot\|_*)^* = \delta_{\mathbf{B}(0,1)}. \tag{9.5}$$

(2) As suggested by the result above we have

$$(\|\cdot\|)^* = \delta_{\mathbf{B}^*(0,1)}. \tag{9.6}$$

Let us prove (9.6). Indeed by contrast with (9.5), (9.6) is an elementary result which does not use the Hahn–Banach theorem. Set $f(v) = \|v\|_V$. Thus, for any $v^* \in V^*$

$$f^*(v^*) = \sup\left\{\langle v^*, v\rangle - \|v\| \ : \ v \in V\right\}.$$

If $\|v^*\|_* \leq 1$, then $\langle v^*, v\rangle - \|v\| \leq 0$ for all $v \in V$ and $\langle v^*, v\rangle - \|v\| = 0$ for $v = 0$. Thus $f^*(v^*) = 0$.

If $\|v^*\|_* > 1$, by definition of $\|\cdot\|_*$ there exists some $v_0 \in V$ such that $\langle v^*, v_0\rangle > \|v_0\|$. It follows that for all $t > 0$

$$\langle v^*, t v_0\rangle - \|t v_0\| = t(\langle v^*, v_0\rangle - \|v_0\|).$$

Hence $\lim_{t\to+\infty}\langle v^*, t v_0\rangle - \|t v_0\| = +\infty$, which implies

$$f^*(v^*) = \sup_{v \in V}\left\{\langle v^*, v\rangle - f(v)\right\} = +\infty.$$

By taking the conjugate in (9.6) and applying the duality Theorem 9.3.2 we obtain for any $v \in V$

$$\|v\| = \sup\left\{\langle v^*, v\rangle : \|v^*\|_* \leq 1\right\};$$

this is the isometrical embedding theorem from V into its bidual V^{**}.

We give in the following proposition an important example of a dual convex function which indeed is an extension of Example 9.3.1, case (2).

Proposition 9.3.3. *Let $(V, \|\cdot\|)$ be a normed space with topological dual space $(V^*, \|\cdot\|_*)$. Let $\varphi : \mathbf{R} \to \mathbf{R} \cup \{+\infty\}$ be a closed convex function which is even (i.e., $\varphi(-t) = \varphi(t)$). Then the function*

$$f : V \to \mathbf{R} \cup \{+\infty\}, \quad f(v) = \varphi(\|v\|),$$

is a closed convex proper function and

$$f^*(v^*) = \varphi^*(\|v^*\|_*).$$

PROOF. The assumptions on φ imply that $\varphi : \mathbf{R}^+ \to \mathbf{R} \cup \{+\infty\}$ is increasing. Thus f is still convex and clearly closed. Moreover,

$$\begin{aligned} f^*(v^*) &= \sup_{v \in V} \left\{ \langle v^*, v \rangle - \varphi(\|v\|) \right\} \\ &= \sup_{t \geq 0} \sup_{v \in V, \|v\| = t} \left\{ \langle v^*, v \rangle - \varphi(\|v\|) \right\} \\ &= \sup_{t \geq 0} \left\{ t\|v^*\|_* - \varphi(t) \right\} \\ &= \sup_{t \in \mathbf{R}} \left\{ t\|v^*\|_* - \varphi(t) \right\} \quad \text{(because } \varphi \text{ is even)} \\ &= \varphi^*(\|v^*\|_*), \end{aligned}$$

which completes the proof. □

As a straightforward consequence we obtain the following useful result.

Corollary 9.3.1. *Set $f(v) = \frac{1}{p}\|v\|^p$ with $1 < p < +\infty$. Then $f^*(v^*) = \frac{1}{p'}\|v^*\|_*^{p'}$ where p' is the Hölder conjugate exponent of p, i.e., $1/p + 1/p' = 1$. In particular, taking $V = L^p(\Omega, \mathscr{A}, \mu)$, $1 < p < +\infty$, we have $V^* = L^{p'}(\Omega, \mathscr{A}, \mu)$ and the conjugate function of*

$$f(v) = \frac{1}{p} \int_\Omega \|v(x)\|^p \, d\mu(x)$$

is equal to

$$f^*(v^*) = \frac{1}{p'} \int_\Omega \|v^*(x)\|^{p'} \, d\mu(x).$$

This makes the transition with the next important example, which is concerned with integral functionals and which will be studied in detail in Chapter 13.

Theorem 9.3.3. *Let $V = L^p(\Omega, \mathscr{A}, \mu)$, $1 < p < +\infty$, and*

$$f(v) = \int_\Omega j(x, v(x)) \, d\mu(x)$$

a convex integral functional associated to a convex normal integrand j. Then

$$f^* : V^* = L^{p'}(\Omega, \mathscr{A}, \mu) \to \mathbf{R} \cup \{+\infty\}$$

is given by

$$f^*(v^*) = \int_\Omega j^*(x, v^*(x)) \, d\mu(x),$$

where $j^*(x, \cdot)$ is the convex conjugate of $j(x, \cdot)$.

Remark 9.3.1. When $V = H$ is a Hilbert space the Legendre–Fenchel transform $f \mapsto f^*$ is an involution from $\Gamma_0(H)$ into itself, where $\Gamma_0(H)$ is the set of closed convex proper functions on H,

$$\Gamma_0(H) \xrightarrow{*} \Gamma_0(H),$$
$$f \mapsto f^*,$$

i.e., $f^{**} = f$. This transform has some analogy with the Fourier–Plancherel transform,

$$\mathscr{F} : L^2(\mathbf{R}^N) \to L^2(\mathbf{R}^N),\ f \mapsto \mathscr{F}(f),$$

where $\overline{\mathscr{F}}\mathscr{F} f = f$, which is indeed an isometry. Let us notice that $(\|\cdot\|^2/2)^* = \|\cdot\|^2/2$ i.e., $\|\cdot\|^2/2$ is invariant for the Legendre–Fenchel transform while $f(x) = \frac{1}{2}e^{-\|x\|^2}$ is invariant for the Fourier–Plancherel transform. A basic property of the Fourier–Plancherel transform is

$$\mathscr{F}(f * g) = \mathscr{F}(f)\mathscr{F}(g),$$

where $f * g$ is the convolution of functions. This property can be seen as the (formal) analogue of the property

$$(f \#_e g)^* = f^* + g^*,$$

which is studied in Section 9.4.

Let us complete this section by studying the natural setting in which the Legendre–Fenchel transform acts as an operator. We will pay particular attention to the description of its range. Our final result (Theorem 9.3.5) shows that the Legendre–Fenchel transform is a one-to-one mapping from $\Gamma_0(V)$ onto $\Gamma_0(V^*)$ (cf. Definition 9.3.2). Given a general normed space $(V, \|.\|)$, let us recall that the Legendre–Fenchel transform is the mapping which associates to a closed convex proper function $f : V \longrightarrow \mathbf{R} \cup \{+\infty\}$ its conjugate $f^* : V^* \longrightarrow \mathbf{R} \cup \{+\infty\}$ which is defined by

$$\forall v^* \in V^* \qquad f^*(v^*) = \sup\left\{\langle v^*, v\rangle - f(v) : v \in V\right\}.$$

Let us start with some simple observations.

(a) The Legendre–Fenchel transform is one-to-one. This results from the implication

$$f^* = g^* \Longrightarrow f^{**} = g^{**}$$

and $f^{**} = f,\ g^{**} = g$, which hold true for arbitrary closed convex proper functions f and g (Theorem 9.3.2).

(b) The description of the range of the Legendre–Fenchel transform requires some attention. Let us return to the (above) definition of f^*. In this supremum, one just needs to consider $v \in \mathrm{dom} f$. For such $v \in \mathrm{dom} f$, the mapping $v^* \mapsto \langle v^*, v\rangle - f(v)$ is continuous on V^* for the topology $\sigma(V^*, V)$, which is the weak star topology of the dual. This follows directly from the definition of this topology. Hence, f^*, as a supremum of such affine, $\sigma(V^*, V)$ continuous functions, is a convex, proper function which is $\sigma(V^*, V)$ lower semicontinuous.

Indeed, when V is not reflexive, for a convex function $g : V^* \longrightarrow \mathbf{R} \cup \{+\infty\}$, to be $\sigma(V^*, V)$ lower semicontinuous is a strictly stronger property than just to be lower semicontinuous for the topology of the norm. One can exhibit a closed convex function $g : V^* \longrightarrow \mathbf{R}$ which is not $\sigma(V^*, V)$ lower semicontinuous. Take any $\xi \in V^{**} \setminus I(V)$, where I is the canonical embedding of V into its bidual V^{**} (recall that $\langle I(v), v^*\rangle_{(V^{**}, V^*)} = \langle v^*, v\rangle_{(V^*, V)}$). Then define $g(v^*) := \langle \xi, v^*\rangle_{(V^{**}, V^*)}$. Clearly g is continuous on V^*, but it is not $\sigma(V^*,V)$ lower semicontinuous. Otherwise, by linearity, it would be continuous for the topology $\sigma(V^*,V)$, which would imply $\xi \in I(V)$.

It turns out that this $\sigma(V^*,V)$ lower semicontinuity property allows us to characterize the range of the Legendre–Fenchel transform. Let us make this precise in the following statement.

Theorem 9.3.4. *Let $(V,||.||)$ be a normed space.*

(a) *For any $f : V \longrightarrow \mathbf{R}\cup\{+\infty\}$ which is closed, convex, and proper, its Legendre–Fenchel conjugate $f^* : V^* \longrightarrow \mathbf{R}\cup\{+\infty\}$ is a convex proper function which is $\sigma(V^*,V)$ lower semicontinuous.*

(b) *Conversely, let $g : V^* \longrightarrow \mathbf{R}\cup\{+\infty\}$ be a convex, proper, and $\sigma(V^*,V)$ lower semicontinuous function. Then,*

$$g = g^{**}$$

and g belongs to the range of the Legendre–Fenchel transform. More precisely, $g = (g^*)^*$ is equal to the Legendre–Fenchel transform of the closed convex proper function $g^* : V \longrightarrow \mathbf{R}\cup\{+\infty\}$ which is defined by

$$\forall v \in V \qquad g^*(v) = \sup\left\{\langle v^*, v\rangle - g(v^*) : v^* \in V^*\right\}.$$

PROOF. Part (a) has already been proved. Proof of part (b) requires some further topological tools. When equipped with the topology $\sigma(V^*,V)$, the space V^* is a locally convex topological vector space, whose dual can be identified with V. The Hahn–Banach theorem still holds in locally convex topological vector spaces, and from that point, the proof is essentially the same as in Theorem 9.3.2. □

To give a unified formulation of Theorem 9.3.2 and Theorem 9.3.4 where V and V^*, f and f^* play symmetrical roles, it is convenient to introduce the following notions and notation.

Definition 9.3.2. *Let $(V,||.||)$ be a normed space with topological dual V^*. We set*

$$\Gamma_{0,V^*}(V) = \big\{f : V \longrightarrow \mathbf{R}\cup\{+\infty\}, f \textit{ is a pointwise supremum of a nonvoid family of affine functions with slopes in } V^*,\ f \not\equiv +\infty\big\};$$

$$\Gamma_{0,V}(V^*) = \big\{g : V^* \longrightarrow \mathbf{R}\cup\{+\infty\},\ g \textit{ is a pointwise supremum of a nonvoid family of affine functions with slopes in } V,\ g \not\equiv +\infty\big\}.$$

These definitions make explicit references to the pairing between the two spaces V and V^*, that is, $(v,v^*) \in V\times V^* \mapsto \langle v^*, v\rangle_{(V^*,V)} = v^*(v)$. Without ambiguity, one often omits the subscript referring to the coupled space and writes briefly $\Gamma_0(V)$ and $\Gamma_0(V^*)$. To be more precise, one has

$$f \in \Gamma_0(V) \iff f = \sup_{i\in I} f_i$$

with $f_i(v) = \langle v_i^*, v\rangle - \alpha_i$ for some index set I, $v_i^* \in V^*$ (slope), and $\alpha_i \in \mathbf{R}$;

$$g \in \Gamma_0(V^*) \iff g = \sup_{j\in J} g_j$$

with $g_j(v^*) = \langle v^*, v_j\rangle - \beta_j$ for some index set J, $v_j \in V$ (slope), and $\beta_j \in \mathbf{R}$.

We can now reformulate Theorem 9.3.1 and its corresponding version when considering the locally convex topological vector space $(V^*, \sigma(V^*, V))$, together with Theorems 9.3.2 and 9.3.4 in the following final statement.

Theorem 9.3.5. *Let $(V, ||.||)$ be a normed space with topological dual V^*. Then,*
(a) *one has*

$$\begin{aligned}\Gamma_0(V) &= \{f : V \longrightarrow \mathbf{R} \cup \{+\infty\},\ f \text{ closed, convex, proper}\}\\ &= \{f : V \longrightarrow \mathbf{R} \cup \{+\infty\},\ f\ \sigma(V, V^*) \text{ closed, convex, proper}\},\end{aligned}$$

while

$$\Gamma_0(V^*) = \{g : V^* \longrightarrow \mathbf{R} \cup \{+\infty\},\ g\ \sigma(V^*, V) \text{ closed, convex, proper}\}.$$

(b) *The Legendre–Fenchel transform is a one-to-one mapping from $\Gamma_0(V)$ onto $\Gamma_0(V^*)$:*

$$\begin{aligned}\Gamma_0(V) &\xrightarrow{\ *\ } \Gamma_0(V^*)\\ f &\longrightarrow f^*.\end{aligned}$$

For any $f \in \Gamma_0(V)$ one has $f = f^{**}$ and for any $g \in \Gamma_0(V^*)$ one has $g = g^{**}$.

Remark 9.3.2. The preceding theory can be developed in the general setting of two vector spaces V and W in separate duality.

Let us denote by $\langle v, w\rangle_{(V,W)}$ a given pairing between elements $v \in V$ and $w \in W$. It is a bilinear form with separating properties, namely,

$$\begin{cases} \forall v \in V,\ v \neq 0,\ \exists w \in W \text{ with } \langle v, w\rangle \neq 0,\\ \forall w \in W,\ w \neq 0,\ \exists v \in V \text{ with } \langle v, w\rangle \neq 0.\end{cases}$$

Then W is the dual of $(V, \sigma(V, W))$ and conversely. The set $\Gamma_0(V)$ (respectively, $\Gamma_0(W)$) is defined by taking suprema of affine functions with slopes in W (respectively, V), and the Legendre–Fenchel transform is a one-to-one mapping from $\Gamma_0(V)$ onto $\Gamma_0(W)$; see Moreau [296] for further details.

9.4 ▪ Legendre–Fenchel calculus

As we have already stressed, most optimization problems can be written as

$$\inf\{f(v) : v \in V\},$$

where $f = f_0 + \delta_C$ is the sum of the objective function f_0 and the indicator function of the constraint C. This explains the importance of getting a formula for the Legendre–Fenchel conjugate of a sum of functions. At this point, the epi-sum plays a central role, because of the following general property.

Proposition 9.4.1. *Let $\varphi, \psi : V \to \mathbf{R} \cup \{+\infty\}$ be two proper functions. Then*

$$(\varphi \# \psi)^* = \varphi^* + \psi^*.$$

PROOF. It is enough to take $v^* \in V^*$ and compute

$$\begin{aligned}(\varphi\#\psi)^*(v^*) &= \sup_{v\in V}\left\{\langle v^*, v\rangle - (\varphi\#\psi)(v)\right\}\\ &= \sup_{v\in V}\left\{\langle v^*, v\rangle - \inf_{v_1+v_2=v}\big(\varphi(v_1)+\psi(v_2)\big)\right\}\\ &= \sup_{v\in V}\left\{\langle v^*, v\rangle + \sup_{v_1+v_2=v}\big(-\varphi(v_1)-\psi(v_2)\big)\right\}\\ &= \sup_{v\in V,\ v_1+v_2=v}\left\{\big(\langle v^*, v_1\rangle - \varphi(v_1)\big)+\big(\langle v^*, v_2\rangle - \psi(v_2)\big)\right\}\\ &= \varphi^*(v^*)+\psi^*(v^*),\end{aligned}$$

which completes the proof. □

Corollary 9.4.1. *Let $f, g : V \to \mathbf{R}\cup\{+\infty\}$ be two closed convex proper functions. Then*

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

As a consequence, when the convex function $f^*\#g^*$ is a $\sigma(V^*, V)$ closed proper function, we have

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

PROOF. By Proposition 9.4.1 we have

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

When f and g are assumed to be closed convex and proper, one gets

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

Taking again the Legendre–Fenchel conjugate, we obtain

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

The function $f^*\#g^*$, as the epi-sum of two convex functions, is still convex (Proposition 9.2.2). When it is $\sigma(V^*, V)$ closed and proper, Theorem 9.3.4 yields

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

which completes the proof. □

We can now state the following theorem from Rockafellar [325] and Moreau [296], which, under a so-called qualification assumption on f and g, asserts that $f^*\#g^*$ is $\sigma(V^*, V)$ closed and hence $(f+g)^* = f^*\#g^*$.

Theorem 9.4.1. *Let V be a normed linear space and let $f, g : V \to \mathbf{R}\cup\{+\infty\}$ be two closed convex and proper functions which satisfy the following qualification assumption:*

$$\text{there is a point } u_0 \in \operatorname{dom} f \cap \operatorname{dom} g \text{ where } f \text{ is continuous.} \qquad (Q)$$

Then $f^*\#g^*$ is a $\sigma(V^*, V)$ closed convex proper function and the following equality holds:

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

Moreover, for any $v^* \in V^*$, the infimum in the definition of $f^*\#g^*$ is achieved.

PROOF. Corollary 9.4.1 tells us that the only point we need to verify is that $f^*\#g^*$ is $\sigma(V^*,V)$ closed. Equivalently, we have to prove that for $\lambda \in \mathbf{R}$, the sublevel set of $f^*\#g^*$

$$C = \left\{ v^* \in V^* \ : \ (f^*\#g^*)(v^*) \leq \lambda \right\}$$

is $\sigma(V^*,V)$ closed. Indeed, we are going to establish that for each $\rho > 0$, $C \cap \rho \mathbf{B}_{V^*}$ is $\sigma(V^*,V)$ closed, i.e., the traces of C on all closed balls of V^* are $\sigma(V^*,V)$ closed. It will follow from the Banach–Dieudonné–Krein–Smulian theorem (see, e.g., [203, Theorem V 5.7]) that C is $\sigma(V^*,V)$ closed. Let $(v_n^*)_{n\in\mathbf{N}}$ be a bounded sequence of elements of C with $v_n^* \to v^*$, $\sigma(V^*,V)$. When V is separable, it is not restrictive to consider sequences. For general V the argument can be readily extended by considering generalized sequences. By definition of $f^*\#g^*$, for each $n \in \mathbf{N}$, there exists some $w_n \in V^*$ such that

$$f^*(v_n^* - w_n^*) + g^*(w_n^*) \leq \lambda + \frac{1}{n}. \tag{9.7}$$

The key point of the proof is to prove that the sequence $(w_n^*)_{n\in\mathbf{N}}$ is bounded in V^*. To that end, we use as an essential fact the qualification assumption (Q): there exist some $r > 0$ and some $M \in \mathbf{R}$ such that

$$\sup_{\|v\|_V \leq 1} f(u_0 + rv) \leq M. \tag{9.8}$$

For any $v \in \mathbf{B}(0,1)$ let us majorize $\langle w_n^*, v\rangle$. To that end, let us write

$$\begin{aligned} r\langle w_n^*, v\rangle_{(V^*,V)} &= \langle w_n^*, rv\rangle \\ &= \langle w_n^*, u_0\rangle + \langle w_n^*, rv - u_0\rangle \\ &= \langle w_n^*, u_0\rangle + \langle v_n^* - w_n^*, u_0 - rv\rangle - \langle v_n^*, u_0 - rv\rangle \\ &\leq g(u_0) + g^*(w_n^*) + f(u_0 - rv) + f^*(v_n^* - w_n^*) - \langle v_n^*, u_0 - rv\rangle. \end{aligned}$$

We rewrite the above inequality in the form

$$r\langle w_n^*, v\rangle_{(V^*,V)} \leq \Big(f^*(v_n^* - w_n^*) + g^*(w_n^*) \Big) + f(u_0 - rv) + g(u_0) + \|u_0 - rv\| \|v_n^*\|_*$$

and use (9.7), (9.8) to obtain

$$r\langle w_n^*, v\rangle_{(V,V^*)} \leq \lambda + \frac{1}{n} + M + g(u_0) + \|v_n^*\|_*(\|u_0\| + r).$$

Using that the sequence (v_n^*) is bounded, we immediately obtain from the above inequality (which is valid for any $v \in B(0,1)$) that $\sup_n \|w_n^*\|_* < +\infty$.

We now use the Banach–Alaoglu–Bourbaki theorem, Theorem 1.4.7, and Corollary 1.4.2: when V is separable (for a general V one can use a device of Attouch and Brezis [43]), the unit ball of V^* is $\sigma(V^*,V)$ sequentially compact. As a consequence, one can find a subsequence $(w_{n_k})_{k\in\mathbf{N}}$ and some $w^* \in V^*$ such that $w_{n_k} \rightharpoonup w^*$ in $\sigma(V^*,V)$.

Let us now use the lower semicontinuity of f^* and g^* for the topology $\sigma(V^*,V)$ and pass to the limit in (9.7) to obtain

$$f^*(v^* - w^*) + g^*(w^*) \leq \lambda.$$

As a consequence,

$$(f^*\#g^*)(v^*) \leq f^*(v^* - w^*) + g^*(w^*) \leq \lambda$$

and $v^* \in C$.

The same argument with $\lambda = f^*\#g^*$ and $v_n^* = v^*$ gives that the infimum in the definition of $f^*\#g^*$ is achieved. □

The qualification assumption (Q), because of its importance, has been intensively studied and many weakened versions of it have been established. Let us quote the following result (see Aubin [64]).

Theorem 9.4.2. *Let V be a Banach space and let $f, g : V \to \mathbf{R}\cup\{+\infty\}$ be two closed convex proper functions such that*

$$\operatorname{dom} f - \operatorname{dom} g \text{ is a neighborhood of the origin.}$$

Then, the same conclusions as Theorem 9.4.1 hold and

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

In the same spirit, the same result was established by Attouch and Brezis in [43] under the even weaker assumption

$$\bigcup_{\lambda>0} \lambda(\operatorname{dom} f - \operatorname{dom} g) \text{ is a closed subspace of } V.$$

Note that by contrast with the Rockafellar theorem, which holds in general normed spaces, the Aubin and Attouch–Brezis theorems require that the space V is a Banach space. Indeed, an essential ingredient in the proof of these theorems is the Banach–Steinhaus theorem. Otherwise, the proof is essentially the same as in Theorem 9.4.1.

9.5 ▪ Subdifferential calculus for convex functions

To obtain the simplest possible dual representation of a closed convex set C of a normed linear space $(V, \|\cdot\|)$, we introduce the notion of supporting hyperplane. When taking $C = \operatorname{epi} f$, the epigraph of a closed convex proper function, the corresponding notion is the exact minorization: a continuous affine function $l : V \to \mathbf{R}$ is an exact minorant of f at u if $l \le f$ and $l(u) = f(u)$.

Equivalently, when setting $l(v) = \langle u^*, v\rangle + \alpha$, this becomes

$$\begin{cases} f(v) \ge \langle u^*, v\rangle + \alpha & \forall\, v \in V, \\ f(u) = \langle u^*, u\rangle + \alpha, \end{cases}$$

i.e., $\alpha = f(u) - \langle u^*, u\rangle$, $l(v) = f(u) + \langle u^*, v-u\rangle$, which is equivalent to

$$\forall v \in V \quad f(v) \ge f(u) + \langle u^*, v-u\rangle.$$

This leads to the following definition.

Definition 9.5.1. *Let $(V, \|\cdot\|)$ be a normed space and $f : V \to \mathbf{R}\cup\{+\infty\}$ be a closed convex proper function. We say that an element $u^* \in V^*$ belongs to the subdifferential of f at $u \in V$ if*

$$\forall v \in V \quad f(v) \ge f(u) + \langle u^*, v-u\rangle_{(V^*,V)}.$$

We then write $u^* \in \partial f(u)$.

The terminology reflects the fact that when f is continuously differentiable and convex, the following inequality holds:

$$\forall v \in V \quad f(v) \geq f(u) + \langle \nabla f(u), v-u \rangle.$$

Moreover, this inequality characterizes $\nabla f(u)$. For this reason, when $u^* \in \partial f(u)$, we say either that u^* belongs to the subdifferential of f at u or that u^* is a subgradient of f at u.

Note that if $u^* \in \partial f(u)$, then necessarily $u \in \operatorname{dom} f$ (take $v_0 \in \operatorname{dom} f \neq \emptyset$; we have $f(v_0) - \langle u^*, v_0 - u \rangle \geq f(u)$ and $f(u) < +\infty$). It is also important to notice that given $u \in \operatorname{dom} f$, the set $\partial f(u)$ may be empty; see Phelps [320, Example 3.8].

Proposition 9.5.1. *Let $(V, \|\cdot\|)$ be a normed space and $f : V \to \mathbf{R} \cup \{+\infty\}$ a closed convex proper function. Then the two following conditions are equivalent:*

(i) $u^* \in \partial f(u)$,

(ii) $f(u) + f^*(u^*) - \langle u^*, u \rangle = 0$.

PROOF. (a) Let us first give a geometrical proof: to say that $u^* \in \partial f(u)$ means that $v \mapsto \langle u^*, v \rangle + f(u) - \langle u^*, u \rangle$ is an exact minorant of f at u. This implies that it is a maximal minorant with slope u^*, i.e., $f(u) - \langle u^*, u \rangle = -f^*(u^*)$.

(b) The analytic proof is also immediate: the inequality

$$f(u) + f^*(u^*) - \langle u^*, u \rangle \geq 0$$

is always true. Thus the equality $f(u) + f^*(u^*) - \langle u^*, u \rangle = 0$ is equivalent to the inequality

$$f(u) + f^*(u^*) - \langle u^*, u \rangle \leq 0.$$

By definition of f^* this is equivalent to saying

$$\langle u^*, u \rangle - f(u) \geq \langle u^*, v \rangle - f(v) \;\forall\, v \in V,$$

i.e., $u^* \in \partial f(u)$. □

Remark 9.5.1. As we have already stressed, for any $v \in V$ and $v^* \in V^*$ the inequality

$$f(v) + f^*(v^*) - \langle v^*, v \rangle \geq 0$$

is always true. Thus, when writing the characterization of $u^* \in \partial f(u)$,

$$f(u) + f^*(u^*) - \langle u^*, u \rangle = 0,$$

we express that for the pair $(u, u^*) \in V \times V^*$, the function $(v, v^*) \mapsto f(v) + f^*(v^*) - \langle v^*, v \rangle$ takes its minimal value. For this reason, relation (ii) in Proposition 9.5.1 is called the *Fenchel extremality relation*.

A major interest of the Fenchel extremality characterization of subdifferentials is that f and f^* play a symmetric role in its formulation. This together with the Fenchel–Moreau–Rockafellar duality theorem, Theorem 9.3.2 (which expresses that $f = f^{**}$), yields the following result.

Theorem 9.5.1. *Let $(V, \|\cdot\|)$ be a normed space and let $f : V \to \mathbf{R} \cup \{+\infty\}$ be a closed convex and proper function. Then, for $u \in V$ and $u^* \in V^*$ we have*

$$u^* \in \partial f(u) \iff u \in \partial f^*(u^*).$$

PROOF. In fact we obtain

$$\begin{aligned} u^* \in \partial f(u) &\iff f(u)+f^*(u^*)-\langle u^*,u\rangle = 0 \\ &\iff f^{**}(u)+f^*(u^*)-\langle u^*,u\rangle = 0 \\ &\iff u \in \partial f^*(u^*), \end{aligned}$$

where we use the Fenchel extremality characterization of the subdifferential and the Fenchel–Moreau–Rockafellar duality theorem, Theorem 9.3.2 ($f = f^{**}$). ☐

Remark 9.5.2. When using the notation of set-valued analysis we can write

$$(\partial f)^{-1} = \partial f^*.$$

This is indeed the formulation, in terms of subdifferentials, of the convex duality theory.

For theoretical reasons it is important to know if a closed convex proper function can be uniquely determined (up to a constant) by its subdifferential. From a geometrical point of view, this can be formulated as follows: *Is a closed convex proper function the upper envelope of its exact continuous affine minorants?* Indeed the answer is yes when V is a Banach space. The proof of this result relies on the Ekeland's ε-variational principle. (See Section 3.4. We state it without proof, referring, for instance, to Phelps [320, Corollary 3.1.9].)

Theorem 9.5.2. *Suppose* $(V,\|\cdot\|)$ *is a Banach space and* $f : V \to \mathbf{R}\cup\{+\infty\}$ *is a closed convex proper function. Then, for any* $u \in \operatorname{dom} f$

$$\begin{aligned} f(u) &= \sup\left\{ f(v)+\langle v^*, u-v\rangle \;:\; v\in V,\ v^*\in V^* \textit{ with } v^*\in\partial f(v)\right\} \\ &= \sup\left\{ \langle v^*,u\rangle - f^*(v^*) \;:\; \exists v\in V \textit{ such that } v^*\in\partial f(v)\right\}. \end{aligned}$$

Note that this theorem, when specialized to convex sets, says that any closed convex nonempty set in a Banach space is the intersection of the closed half-spaces defined by its supporting hyperplanes (Phelps [320, Proposition 3.2.1]).

When proving the above theorem via Ekeland's variational principle one obtains in the process the following density result.

Theorem 9.5.3. *Suppose* $(V,\|\cdot\|)$ *is a Banach space and* $f : V \to \mathbf{R}\cup\{+\infty\}$ *is a closed convex proper function. Then,* $\operatorname{dom}\partial f$ *is dense in* $\operatorname{dom} f$*. More precisely, for any* $v\in\operatorname{dom} f$*, there exists a sequence* $(v_n)_{n\in\mathbf{N}}$ *with* $v_n\in\operatorname{dom}\partial f$ *for all* $n\in\mathbf{N}$ *such that*

$$v_n \to v \quad \textit{and} \quad f(v_n)\to f(v).$$

PROOF. For the proof, see Azé [70, Theorem 3.2.4] and Aubin and Ekeland [67, Theorem 3]. For a proof in the case when $(V,\|\cdot\|)$ is a reflexive Banach space, see Proposition 17.4.3. ☐

To develop a calculus for subdifferentials it is convenient to consider ∂f as a multivalued operator,

$$\partial f : V \rightrightarrows V^*,$$

and to identify ∂f with its graph

$$\partial f = \left\{ (v,v^*)\in V\times V^* \;:\; v^*\in\partial f(v)\right\}.$$

We recall the basic definitions for calculus of set-valued mappings: given $A, B : V \overset{\longrightarrow}{\longrightarrow} V^*$ we have

$$\begin{aligned}&\mathrm{dom}\, A = \{v \in V \ : \ \exists v^* \in V^* \text{ with } (v, v^*) \in A\},\\ &A^{-1} = \{(v^*, v) \in V^* \times V \ : \ (v, v^*) \in A\},\end{aligned}$$

$$\begin{cases} \mathrm{dom}(A+B) = \mathrm{dom}\, A \cap \mathrm{dom}\, B,\\ (A+B)(v) = Av + Bv \quad \text{in the sense of vectorial sum.}\end{cases}$$

Moreover, we say that $A \subset B$ if graph $A \subset$ graph B.

As we have already stressed, the convex duality theory can be expressed as

$$\partial f^* = (\partial f)^{-1}.$$

Theorem 9.5.4. *Let $(V, \|\cdot\|)$ be a normed space and let $f, g : V \to \mathbf{R} \cup \{+\infty\}$ be two closed convex proper functions.*

(a) *The following inclusion is always true:*

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

(b) *If moreover the qualification assumption* (Q) *holds,*

$$f \text{ is finite and continuous at a point of } \mathrm{dom}\, g, \tag{Q}$$

then we have

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

PROOF. (a) Take $u \in \mathrm{dom}\, \partial f \cap \mathrm{dom}\, \partial g$, $u^* \in \partial f(u)$, and $w^* \in \partial g(u)$. By the definition of ∂f and ∂g, for any $v \in V$,

$$f(v) \geq f(u) + \langle u^*, v-u\rangle,$$

$$g(v) \geq g(u) + \langle w^*, v-u\rangle.$$

By adding these two inequalities, we obtain for any $v \in V$

$$(f+g)(v) \geq (f+g)(u) + \langle u^* + w^*, v-u\rangle,$$

i.e., $u^* + w^* \in \partial(f+g)(u)$.

(b) Take $u^* \in \partial(f+g)(u)$. Equivalently, by using the Fenchel extremality relation, we obtain

$$(f+g)(u) + (f+g)^*(u^*) - \langle u^*, u\rangle = 0.$$

By Theorem 9.4.1, we have

$$(f+g)^*(u^*) = (f^* \# g^*)(u^*)$$

and the infimum in the definition of $(f^* \# g^*)(u^*)$ is achieved. Consequently, there exists some $w^* \in V^*$ such that

$$(f+g)(u) + f^*(u^* - w^*) + g^*(w^*) - \langle u^*, u\rangle = 0.$$

Equivalently,

$$(f(u)+f^*(u^*-w^*)-\langle u^*-w^*,u\rangle)+(g(u)+g^*(w^*)-\langle w^*,u\rangle)=0.$$

By the Fenchel inequality,

$$f(u)+f^*(u^*-w^*)-\langle u^*-w^*,u\rangle \geq 0,$$

$$g(u)+g^*(w^*)-\langle w^*,u\rangle \geq 0.$$

Since the sum of these two quantities is equal to zero, we obtain

$$f(u)+f^*(u^*-w^*)-\langle u^*-w^*,u\rangle = 0,$$

$$g(u)+g^*(w^*)-\langle w^*,u\rangle = 0.$$

These are the Fenchel extremality relations (Proposition 9.5.1) and they are equivalent to

$$u^*-w^* \in \partial f(u) \quad \text{and} \quad w^* \in \partial g(u).$$

Finally, we obtain

$$u^*=(u^*-w^*)+w^* \in \partial f(u)+\partial g(u),$$

i.e., $u^* \in (\partial f + \partial g)(u)$. □

We already stressed the fact that for $u \in \operatorname{dom} f$, the set $\partial f(u)$ may be empty. The following result, which can be viewed as a corollary of Theorem 9.5.4, gives a sufficient condition for the set $\partial f(u)$ to be nonempty. This result, as we will see in Sections 9.6 and 9.8, is quite useful for applications.

Proposition 9.5.2. *Let $(V,\|\cdot\|)$ be a normed space and let $f : V \to \mathbf{R}\cup\{+\infty\}$ be a closed convex and proper function. Let us assume that f is continuous at $u \in \operatorname{dom} f$. Then $\partial f(u) \neq \emptyset$ and $\partial f(u)$ is a closed convex and bounded subset of V^*.*

PROOF. Let us apply Theorem 9.4.1 to the sum of the two closed convex and proper functions f and $g=\delta_{\{u\}}$ (g is the indicator function of the singleton $\{u\}$). By assumption, f is continuous at the point u, and the qualification assumption (Q) of Theorem 9.4.1 is satisfied.

Hence, for any $v^* \in V^*$, the equality

$$(f+\delta_{\{u\}})^*(v^*)=(f^* \# \delta^*_{\{u\}})(v^*)$$

holds, and the infimum in the formulation of $(f^*\#\delta^*_{\{u\}})(v^*)$ is achieved. An elementary computation yields

$$(f+\delta_{\{u\}})^*(v^*)=\langle v^*,u\rangle - f(u),$$
$$\delta^*_{\{u\}}(w^*)=\langle w^*,u\rangle.$$

Hence, for any $v^* \in V^*$, there exists some $w^* \in V^*$ such that

$$\langle v^*,u\rangle - f(u) = f^*(v^*-w^*)+\langle w^*,u\rangle.$$

Equivalently,

$$f(u)+f^*(v^*-w^*)-\langle v^*-w^*,u\rangle = 0.$$

This is the Fenchel extremality relation. This is equivalent to

$$v^* - w^* \in \partial f(u),$$

which expresses that $\partial f(u) \neq \emptyset$.

Note that, as well, we may have applied Theorem 9.5.4 instead of Theorem 9.4.1 to obtain the above result.

As a general rule, the set $\partial f(u)$ is closed and convex. This is an immediate consequence of the definition of $\partial f(u)$. Let us now verify that under the continuity assumption of f at u, this set is bounded. Since f is continuous at u, it is bounded on a neighborhood of u. Let $r > 0$ and $M \geq 0$ be such that

$$f(u + rv) \leq M \quad \forall v \in \mathbf{B}(0,1).$$

Take $v^* \in \partial f(u)$. By definition of ∂f, we have for all $v \in \mathbf{B}(0,1)$

$$f(u + rv) \geq f(u) + r\langle v^*, v\rangle.$$

Hence

$$\langle v^*, v\rangle \leq \frac{1}{r}(M + |f(u)|).$$

This being true for any $v \in \mathbf{B}(0,1)$, we obtain

$$\|v^*\|_* \leq \frac{1}{r}(M + |f(u)|),$$

and, as a consequence, the set $\partial f(u)$ is bounded. $\square$

Let us now come to the central role played by the subdifferential calculus in convex optimization. The following result, despite its elementary proof (it is a straightforward consequence of the definition of ∂f), shows the role of the subdifferential optimality rule $\partial f(u) \ni 0$ as a substitute to the classical Fermat rule.

Proposition 9.5.3. *Let $(V, \|\cdot\|)$ be a normed space and let $f : V \to \mathbf{R} \cup \{+\infty\}$ be a closed convex and proper function. Then, for an element $u \in V$ the two following statements are equivalent:*

(i) $f(u) \leq f(v)$ *for all* $v \in V$;

(ii) $\partial f(u) \ni 0$.

Let us stress that the above proposition gives a necessary and sufficient condition for an element $u \in V$ to be a solution of the convex minimization problem

$$\min\left\{f(v) : v \in V\right\}.$$

This necessary and sufficient condition

$$\partial f(u) \ni 0$$

is an extension to nonsmooth convex functions of the classical first-order necessary and sufficient condition of optimality for convex $\mathbf{C}^1$ functions, namely,

$$\nabla f(u) = 0.$$

Thus, for a given convex optimization problem, the problem which consists in finding the optimal solutions can be attacked by using the subdifferential calculus and solving the generalized equation $\partial f(u) \ni 0$.

As we have stressed, Legendre–Fenchel calculus and subdifferential calculus are intimately connected; playing with both of them when passing from one formulation to the other gives a lot of flexibility and makes a rich calculus. This calculus is made even richer when exploiting some of its geometrical aspects (duality via polar cones, etc.).

Let us develop these ideas in the following general approach to optimization (both finite and infinite dimensional) problems. Let $f_0 : V \to \mathbf{R} \cup \{+\infty\}$ be a closed convex proper function (objective function) on a normed linear space V, and let $C \subset V$ be a closed convex nonempty subset of V (set of constraints). Consider the following optimization problem:

$$\inf\{f_0(v) : v \in C\}. \qquad (\mathscr{P})$$

It can be written in the equivalent form

$$\inf\{f(v) : v \in V\},$$

where $f := f_0 + \delta_C$.

An element $u \in V$ is an optimal solution of $(\mathscr{P})$ iff $\partial f(u) \ni 0$. To compute ∂f we assume that the qualification assumption (Q) is satisfied:

$$f_0 \text{ is continuous at a point of } C \text{ or } \operatorname{int} C \cap \operatorname{dom} f_0 \neq \emptyset. \qquad (Q)$$

Then, Theorem 9.5.4 tells us that it is equivalent to look for a solution of the equation

$$\partial f_0(u) + \partial \delta_C(u) \ni 0.$$

To describe the subdifferential of the indicator function of a closed convex set C, we need to introduce the notion of tangent and normal cone to C at a point $u \in C$.

Definition 9.5.2. *Let C be a closed convex nonempty subset of a normed space V and let $u \in C$.*

(a) *The tangent cone to C at u, denoted by $T_C(u)$, is defined by*

$$T_C(u) = \overline{\bigcup_{\lambda \geq 0} \lambda(C-u)}.$$

It is the closure of the cone spanned by $C-u$.

(b) *The normal cone (also called outward normal cone) $N_C(u)$ to C at $u \in C$ is the polar cone of the tangent cone:*

$$\begin{aligned} N_C(u) &= \{v^* \in V^* : \langle v^*, v\rangle \leq 0 \quad \forall\, v \in T_C(u)\} \\ &= \{v^* \in V^* : \langle v^*, v-u\rangle \leq 0 \quad \forall\, v \in C\}. \end{aligned}$$

Proposition 9.5.4. *Let C be a closed convex nonempty subset of a normed space V. Then, for every $u \in C$,*

$$\partial \delta_C(u) = N_C(u).$$

PROOF. By definition of the subdifferential

$$
\begin{aligned}
u^* \in \partial \delta_C(u) &\iff \delta_C(v) \geq \delta_C(u) + \langle u^*, v-u\rangle \quad \forall v \in C \\
&\iff \begin{cases} u \in C, \\ \langle u^*, v-u\rangle \leq 0 \quad \forall\, v \in C \end{cases} \\
&\iff \begin{cases} u \in C, \\ \langle u^*, v\rangle \leq 0 \quad \forall\, v \in T_C(u), \end{cases}
\end{aligned}
$$

that is, $u^* \in N_C(u)$. □

An equivalent and quite useful characterization of $N_C(u)$ is given by the Fenchel extremality relation:

$$
\begin{aligned}
u^* \in N_C(u) &\iff u^* \in \partial \delta_C(u) \\
&\iff \delta_C(u) + \delta_C^*(u^*) = \langle u^*, u\rangle \\
&\iff \sigma_C(u^*) = \langle u^*, u\rangle,
\end{aligned}
$$

where we have used that $\delta_C^* = \sigma_C$ (see Proposition 9.3.1). Let us formulate this result precisely.

Proposition 9.5.5. *Let C be a closed convex nonempty subset of a normed linear space V. For every $u \in C$ we have*

$$N_C(u) = \left\{ u^* \in V^* \,:\, \langle u^*, u\rangle = \max\{\langle u^*, v\rangle \,:\, v \in C\}\right\}.$$

Equivalently, an element u^* of $N_C(u)$ is characterized by the fact that the linear form $v \mapsto \langle u^*, v\rangle$ attains its maximum on C at the point u.

Let us come back to the convex constrained optimization problem $(\mathscr{P})$. We can summarize the previous results in the following statement.

Theorem 9.5.5. *Let $(V, \|\cdot\|)$ be a normed space, let $f_0 : V \to \mathbf{R} \cup \{+\infty\}$ be a closed convex and proper function, and let $C \subset V$ be a closed convex nonempty subset. We assume that one of the two following qualification assumptions (Q_1) or (Q_2) is satisfied:*

$$f_0 \text{ is continuous at some point of } C, \tag{Q_1}$$

$$\operatorname{dom} f_0 \cap \operatorname{int} C \neq \emptyset. \tag{Q_2}$$

Then the following statements are equivalent:

(i) *u is an optimal solution of the minimization problem $(\mathscr{P})$*

$$\min\left\{ f_0(v) \,:\, v \in C \right\}; \tag{$\mathscr{P}$}$$

(ii) *u is a solution of the equation*

$$\partial f_0(u) + N_C(u) \ni 0;$$

(iii) *there exists some $u^* \in V^*$ such that*

$$\begin{cases} u \in C, \\ u^* \in \partial f_0(u), \\ \langle u^*, v-u\rangle \geq 0 \quad \forall\, v \in C. \end{cases}$$

To go further we need to enrich the model and give more information on the structure of the set of constraints C. Because of its practical importance, in the next subsection we are going to pay particular attention to the mathematical convex programming theory (and in particular to linear programming) and the theory of multipliers. We will see how the notion of the dual problem naturally occurs.

When f_0 is a smooth convex function, say, $f_0 \in \mathbf{C}^1(V, \mathbf{R})$, Theorem 9.5.5 takes the following simpler equivalent form: u is an optimal solution of the above minimization problem $(\mathscr{P})$ iff

$$\text{(iii)} \qquad u \in C \text{ and } \langle \nabla f_0(u), v-u\rangle \geq 0 \text{ for every } v \in C.$$

Problem (iii) is a particular case of the following general *variational inequality problem*: given an operator $A : V \to V^*$ and $z \in V^*$

$$\begin{cases} \text{find } u \in C \text{ such that} \\ \langle Au, v-u\rangle \geq \langle z, v-u\rangle \quad \forall\, v \in C. \end{cases}$$

Note that when $C = V$ (i.e., there are no constraints), the above problem reduces to the standard equation $Au = z$.

As an example, let us examine the important case where C is a closed convex cone such that $C \cap (-C) = \{0\}$. Then, C is equal to the positive cone for the partial ordering $v \geq u \iff v - u \in C$. Then problem (iii) takes the following equivalent form:

$$\begin{cases} \nabla f_0(u) \geq 0, \\ u \geq 0, \\ \langle \nabla f_0(u), u\rangle = 0. \end{cases}$$

(The last equality is obtained by taking successively $v = 0$ and $v = 2u$ in (iii).) This type of problem is called a *complementarity problem*.

Take now a closely related problem where $C = \{v \in V \ : \ v \geq g\}$, where $g \in V$ is given. One can easily obtain that (iii) becomes

$$\begin{cases} \nabla f_0(u) \geq 0, \\ u \geq g, \\ \langle \nabla f_0(u), u-g\rangle = 0. \end{cases}$$

When $V = H_0^1(\Omega)$ and $f_0(v) = \frac{1}{2}\int_\Omega |\nabla v(x)|^2\, dx$ is the Dirichlet integral, we obtain

$$\begin{cases} -\Delta u \geq 0, \\ u \geq g, \\ \langle \Delta u, u-g\rangle = 0. \end{cases}$$

The first condition expresses that $-\Delta u = \mu \geq 0$ is a nonnegative Radon measure. The last condition (complementary condition) can be recognized as

$$\int_\Omega (\tilde{u} - \tilde{g})\, d\mu = 0,$$

where $\tilde{u}$ and $\tilde{g}$ are the quasi-continuous representatives of u and g. It expresses that $\mu = -\Delta u$ does not charge the set where $\tilde{u} > \tilde{g}$.

In other words, μ is concentrated on the *contact set* $\omega = \{\tilde{u} = \tilde{g}\}$ and we have to solve the free boundary value problem:

$$\begin{cases} -\Delta u = 0 & \text{on } \Omega \setminus \omega, \\ u = g & \text{on } \omega, \\ u = 0 & \text{on } \partial\Omega. \end{cases}$$

9.6 ▪ Mathematical programming: Multipliers and duality

In this section, $(V, ||.||)$ is a normed space. Mathematical programming is concerned with optimization problems of the form

$$\min\left\{f_0(v) \,:\, f_1(v) \le 0, \dots, f_n(v) \le 0\right\}, \tag{$\mathscr{P}$}$$

where f_i $(i = 1, \dots, n)$ are given functions from V into $\mathbf{R}$.

Thus, a mathematical programming problem is an optimization problem where the constraint C has the following specific form:

$$C = \left\{v \in V \,:\, f_i(v) \le 0, \quad i = 1, \dots, n\right\}.$$

This problem is of fundamental importance; a large number of problems in decision sciences, engineering, and so forth can be written as mathematical programming problems.

The mathematical analysis of this kind of problem depends heavily on the geometrical properties of the functions f_i $(i = 0, \dots, n)$. When the functions f_i are affine, $(\mathscr{P})$ is called a linear programming problem. When f_i $(i = 1, \dots, n)$ are affine and f_0 is quadratic, $(\mathscr{P})$ is called a quadratic programming problem.

In this section, we study the situation where $f_0, f_1, \dots, f_n$ are supposed to be convex functions. Thus $(\mathscr{P})$ is a convex minimization problem (f_0 and C are convex); it is called a convex mathematical programming problem.

9.6.1 ▪ Karush–Kuhn–Tucker optimality conditions

The following theorem, which is the central result of this section, will be obtained by applying Theorem 9.5.5 to our situation.

Because of the specific form of the constraint C, the constraint qualification assumption (Q) takes a quite simple form (this is the Slater qualification assumption). The computation of the normal cone $N_C(u)$ provides, as fundamental mathematical objects, the Karush–Kuhn–Tucker optimality conditions and the corresponding Lagrange multipliers.

Theorem 9.6.1. *Suppose that V is a normed space, $f_0 : V \to \mathbf{R} \cup \{+\infty\}$ is closed convex proper, and $f_1, \dots, f_n : V \to \mathbf{R}$ are convex and continuous. Suppose moreover that the following Slater qualification assumption is satisfied:*

There exists some $v_0 \in V$ such that $f_0(v_0) < +\infty$ and such that $f_i(v_0) < 0$ $\forall i = 1, \dots, n$.

Then the following statements are equivalent:

(i) *u is a solution of problem $(\mathscr{P})$ above;*

(ii) *there exist $\lambda_1, \lambda_2, \dots, \lambda_n$ in $\mathbf{R}^+$ such that*

$$\begin{cases} \partial f_0(u) + \lambda_1 \partial f_1(u) + \cdots + \lambda_n \partial f_n(u) \ni 0, \\ \lambda_i \ge 0 \quad \forall i = 1, \dots, n, \\ f_i(u) \le 0 \quad \forall i = 1, \dots, n, \\ \lambda_i f_i(u) = 0 \quad \forall i = 1, \dots, n. \end{cases}$$

The central point of the proof of Theorem 9.6.1 is the computation of the normal cone $N_C(u)$. We are going to do it first when C is a closed half-space (that is, when $C=\{v\in V : f(v)\le 0\}$ with f affine continuous) and then in the general case.

Lemma 9.6.1. *Let $(V,||.||)$ be a normed space and $u^*\in V^*$ with $u^*\neq 0$. Let us consider the closed half-space*

$$\mathscr{H}=\{v\in V : \langle u^*,v-u\rangle\le 0\}.$$

Then, $N_{\mathscr{H}}(u)=\mathbf{R}_+u^*$.

In other words, $v^*\in V^*$ belongs to the normal cone to $\mathscr{H}$ at u iff there exists some $\lambda\ge 0$ such that $v^*=\lambda u^*$.

PROOF. The inclusion $\mathbf{R}_+u^*\subset N_{\mathscr{H}}(u)$ is immediate: by the definition of $\mathscr{H}$, we have $\langle u^*,v-u\rangle\le 0$ for all $v\in\mathscr{H}$. Hence $u^*\in N_{\mathscr{H}}(u)$ and $\mathbf{R}_+u^*\subset N_{\mathscr{H}}(u)$.

Conversely, take $v^*\in N_{\mathscr{H}}(u)$, $v^*\neq 0$ (the case $v^*=0$ is trivial). By definition of $N_{\mathscr{H}}(u)$, we have

$$\langle v^*,v-u\rangle\le 0\quad\forall\, v\in\mathscr{H}. \tag{9.9}$$

As a particular subset of $\mathscr{H}$, let us consider the affine subspace

$$W=\{v\in V : \langle u^*,v-u\rangle=0\}.$$

We have $W=u+M$, where $M=\ker u^*$ is the hyperspace

$$M=\{v\in V : \langle u^*,v\rangle=0\}.$$

By taking in (9.9) elements v belonging to $W=u+M$, we obtain

$$\langle v^*,v\rangle\le 0\quad\forall\, v\in M.$$

Then, replace v by $-v$ (M is a subspace) to obtain

$$\langle v^*,v\rangle=0\quad\forall\, v\in M.$$

We now follow a standard device in linear algebra. Take an arbitrary element $w\in V, w\notin M$; noticing that

$$\left\langle u^*, v-\frac{\langle u^*,v\rangle}{\langle u^*,w\rangle}w\right\rangle=0,$$

we deduce that for every $v\in V$,

$$v-\frac{\langle u^*,v\rangle}{\langle u^*,w\rangle}w\in M=\ker u^*.$$

Since $v^*=0$ on M, we have

$$\langle v^*,v\rangle=\left\langle v^*, \frac{\langle u^*,v\rangle}{\langle u^*,w\rangle}w\right\rangle,$$

that is,

$$\langle v^*,v\rangle=\left\langle \frac{\langle v^*,w\rangle}{\langle u^*,w\rangle}u^*, v\right\rangle.$$

This being true for all $v \in V$, we finally obtain

$$v^* = \frac{\langle v^*, w\rangle}{\langle u^*, w\rangle} u^*,$$

i.e., $v^* = t u^*$ for some $t \in \mathbf{R}$.

Until now, we have exploited only a part of the information given by (9.9). Returning to (9.9), t must satisfy

$$t\langle u^*, v-u\rangle \leq 0 \quad \forall\, v \in \mathcal{H}.$$

Since for all $v \in \mathcal{H}$ $\langle u^*, v-u\rangle \leq 0$, we necessarily have $t \geq 0$. $\square$

Let us now examine the situation where $C = \{v \in V \ : \ f(v) \leq 0\}$ and compute the normal cone $N_C(u)$ at an arbitrary point u of C.

Proposition 9.6.1. *Suppose that $f : V \to \mathbf{R}$ is a convex continuous function on a normed linear space V. Set*

$$C = \{v \in V \ : \ f(v) \leq 0\}$$

and assume that C satisfies the following Slater property:

there exists some $v_0 \in C$ such that $f(v_0) < 0$.

Then, for every $u \in C$

$$N_C(u) = \begin{cases} \{0\} & \text{if } f(u) < 0, \\ \mathbf{R}^+ \partial f(u) & \text{if } f(u) = 0. \end{cases}$$

As a consequence,

$$u^* \in N_C(u) \iff \exists \lambda \geq 0 \text{ such that } u^* \in \lambda \partial f(u) \text{ and } \lambda f(u) = 0.$$

PROOF. Take $u \in C$. If $f(u) < 0$, because of the continuity of f, we have $u \in \operatorname{int} C$, which yields $T_C(u) = V$ and hence $N_C(u) = \{0\}$.

If on the contrary $f(u) = 0$, let us prove that $N_C(u) = \mathbf{R}_+ \partial f(u)$. The inclusion $\mathbf{R}_+ \partial f(u) \subset N_C(u)$ is quite easy to verify: take $u^* \in \partial f(u)$; by definition of the subdifferential $\partial f(u)$ of f at u

$$\forall v \in V \quad f(v) \geq f(u) + \langle u^*, v-u\rangle.$$

Noticing that $f(u) = 0$ and $f(v) \leq 0$ for all $v \in C$, we obtain

$$\langle u^*, v-u\rangle \leq 0 \quad \forall\, v \in C,$$

i.e., $u^* \in N_C(u)$. Since $N_C(u)$ is a cone, we obtain $\mathbf{R}_+ \partial f(u) \subset N_C(u)$.

Let us now prove the opposite inclusion, which is the delicate part of the proof: $N_C(u) \subset \mathbf{R}_+ \partial f(u)$. Equivalently, we have to prove that if $f(u) = 0$ and $u^* \in N_C(u)$, then there exists some $\lambda \geq 0$ such that $u^* \in \lambda \partial f(u)$. The case $u^* = 0$ is trivial, so we assume in the following that $u^* \neq 0$. We are going to prove the existence of such λ by using a variational argument. As a direct consequence of the definition of the normal cone, we have (see Proposition 9.5.5) the equivalence

$$u^* \in N_C(u)$$
$$\Updownarrow$$
$$\text{the linear form } v \mapsto \langle u^*, v\rangle \text{ attains its maximal value on } C \text{ at } u \in C.$$

As a general property of a linear form, the maximum of the linear form $v \mapsto \langle u^*, v\rangle$ on C is attained on its boundary and

$$v \in \operatorname{int} C \Longrightarrow \langle u^*, v\rangle < \langle u^*, u\rangle.$$

Hence

$$f(v) < 0 \Longrightarrow \langle u^*, v\rangle < \langle u^*, u\rangle.$$

Therefore

$$\langle u^*, v-u\rangle \geq 0 \Longrightarrow f(v) \geq 0,$$

that is, on the closed half-space $\mathscr{H} = \left\{v \in V \ : \ \langle u^*, v-u\rangle \geq 0\right\}$ we have $f(v) \geq 0$. Noticing that $u \in \mathscr{H}$ and $f(u) = 0$, we have the following variational property:

"f achieves its minimal value on the half-space $\mathscr{H}$ at the point u."

Hence, $\partial(f + \delta_{\mathscr{H}})(u) \ni 0$. Since f is continuous, we can apply Theorem 9.5.5 to obtain

$$\partial f(u) + N_{\mathscr{H}}(u) \ni 0.$$

We are in the situation described in Lemma 9.6.1. Noticing that

$$\mathscr{H} = \left\{v \in V \ : \ \langle -u^*, v-u\rangle \leq 0\right\},$$

we thus have $N_{\mathscr{H}}(u) = \mathbf{R}_+(-u^*) = \mathbf{R}_-(u^*)$.

As a consequence, there exists some $t \leq 0$ such that

$$\partial f(u) + t u^* \ni 0.$$

Let us finally prove that $t < 0$. Otherwise, $t = 0$ and $\partial f(u) \ni 0$, which expresses that f attains its minimal value at u. This is impossible because $f(u_0) < 0$ (Slater condition) and $f(u) = 0$. Thus $t < 0$, and, dividing by t the above relation, we obtain $u^* \in -\frac{1}{t}\partial f(u)$, i.e., $u^* \in \mathbf{R}_+ \partial f(u)$. □

We have now all the elements to prove Theorem 9.6.1.

PROOF OF THEOREM 9.6.1. Let us first verify that all the assumptions of Theorem 9.5.5 are satisfied. Since the functions f_i are continuous, the Slater condition implies that $v_0 \in \operatorname{int} C$. Since $f_0(v_0) < +\infty$, we have $\operatorname{dom} f \cap \operatorname{int} C \neq \emptyset$ and the qualification assumption (Q_2) is satisfied.

Thus u is a solution of the convex programming problem $(\mathscr{P})$ iff

$$\partial f_0(u) + N_C(u) \ni 0.$$

Then notice that $C = \bigcap_{i=1}^n C_i$, where $C_i = \left\{v \in V \ : \ f_i(v) \leq 0\right\}$, which is equivalent to saying that $\delta_C = \delta_{C_1} + \cdots + \delta_{C_n}$. The Slater condition implies that each of the closed convex functions $f_i = \delta_{C_i}$ is continuous at the point v_0. Thus, the subdifferential rule for the sum of convex functions (see Theorem 9.5.4) gives

$$\partial \delta_C = \partial \delta_{C_1} + \cdots + \partial \delta_{C_n},$$

that is, for any $u \in C$,

$$N_C(u) = N_{C_1}(u) + \cdots + N_{C_n}(u).$$

We now combine these results with Proposition 9.6.1 to obtain the existence of real numbers $\lambda_1 \geq 0, \dots, \lambda_n \geq 0$ such that

$$\begin{aligned}&\partial f_0(u) + \lambda_1 \partial f_1(u) + \cdots + \lambda_n \partial f_n(u) \ni 0,\\ &\lambda_i = 0 \quad \text{if } f_i(u) < 0.\end{aligned}$$

Thus, in all cases $\lambda_i f_i(u) = 0$. □

9.6.2 ▪ The marginal approach to multipliers

Let us first restate Theorem 9.6.1 in a variational way.

Proposition 9.6.2. *Assume that the hypotheses of Theorem* 9.6.1 *are satisfied. Let u be an optimal solution of the minimization problem*

$$\min\left\{f_0(v) \,:\, f_i(v) \leq 0, \qquad i = 1, \dots, n\right\}. \tag{$\mathscr{P}$}$$

(a) *Then, there exists some vector $\lambda \in \mathbf{R}^n_+$ such that u is a solution of the unconstrained minimization problem:*

$$\min\left\{f_0(v) + \sum_{i=1}^{n} \lambda_i f_i(v) \,:\, v \in V\right\}. \tag{$\mathscr{P}_\lambda$}$$

Moreover, the complementarity slackness condition holds:

$$\lambda_i f_i(u) = 0, \qquad i = 1, \dots, n.$$

(b) *Conversely, if for some $\lambda \in \mathbf{R}^n_+$, u is a solution of the unconstrained minimization problem $(\mathscr{P}_\lambda)$ and*

$$\begin{cases} f_i(u) \leq 0, & i = 1, \dots, n,\\ \lambda_i f_i(u) = 0, & i = 1, \dots, n,\end{cases}$$

then u is an optimal solution of the minimization problem $(\mathscr{P})$.

PROOF. Just notice that, since for all $i = 1, \dots, n$ the functions f_i are supposed to be continuous, the additivity rule for subdifferentials holds,

$$\partial f_0 + \lambda_1 \partial f_1 + \cdots + \lambda_n \partial f_n = \partial (f_0 + \lambda_1 f_1 + \cdots + \lambda_n f_n),$$

and the Karush–Kuhn–Tucker condition can be written in the form

$$\partial \left(f_0 + \sum_{i=1}^{n} \lambda_i f_i \right)(u) \ni 0.$$

This expresses that u is a solution of the convex unconstrained minimization problem $(\mathscr{P}_\lambda)$. □

Definition 9.6.1. *Let u be an optimal solution of the minimization problem $(\mathscr{P})$ above. We call a vector $\lambda \in \mathbf{R}^n_+$ a Lagrange multiplier vector for u if*

$$\partial f_0(u) + \sum_{i=1}^{n} \lambda_i \partial f_i(u) \ni 0$$

and

$$\lambda_i f_i(u) = 0, \qquad i = 1, \ldots, n.$$

The determination of Lagrange multipliers is a central question since, if we are able to compute a Lagrange multiplier $\lambda(u)$ of an optimal solution u, then u can be obtained as a solution of the unconstrained minimization problem

$$\min\left\{ f_0(v) + \sum_{i=1}^{n} \lambda_i f_i(v) \ : \ v \in V \right\}. \qquad (\mathscr{P}_\lambda)$$

Let us first notice that the set of Lagrange multipliers does not depend on the solution u, i.e., if u_1 and u_2 are two solutions of the minimization problem $(\mathscr{P})$, then $M(u_1) = M(u_2)$, where $M(u_i)$ is the set of Lagrange multipliers of the solution u_i. Indeed, this is a consequence of the following characterization of Lagrange multipliers.

Proposition 9.6.3. *Let u be an optimal solution of the minimization problem*

$$\min\left\{ f_0(v) \ : \ f_i(v) \leq 0, \qquad i = 1, \ldots, n \right\}. \qquad (\mathscr{P})$$

Then, the set of Lagrange multipliers for u is equal to

$$M = \left\{ \lambda \in \mathbf{R}^n_+ \ : \ \inf_C f_0 = \inf_V \left(f_0 + \sum_{i=1}^{n} \lambda_i f_i \right) \right\},$$

where $C = \left\{ v \in V \ : \ f_i(v) \leq 0 \text{ for all } i = 1, \ldots, n \right\}$ is the set of constraints.

PROOF. Take a Lagrange multiplier λ for u. Then u is a solution of the unconstrained minimization problem

$$f_0(u) + \sum_{i=1}^{n} \lambda_i f_i(u) = \inf_V \left\{ f_0(v) + \sum_{i=1}^{n} \lambda_i f_i(v) \ : \ v \in V \right\}.$$

Because of the complementarity slackness property we deduce

$$f_0(u) = \inf_V \left\{ f_0 + \sum_{i=1}^{n} \lambda_i f_i \right\}.$$

On the other hand, since u is an optimal solution of $(\mathscr{P})$, we have

$$f_0(u) = \inf\left\{ f_0 + \delta_C \right\},$$

which proves that $\lambda \in M$.

Conversely, let us suppose that $\lambda \in M$. Then,

$$f_0(u) \leq f_0(u) + \sum_{i=1}^{n} \lambda_i f_i(u)$$

and $\sum_{i=1}^n \lambda_i f_i(u) \geq 0$. Since $\lambda_i \geq 0$ and $f_i(u) \leq 0$, this implies $\lambda_i f_i(u) = 0$ for all $i = 1, \dots, n$. Hence

$$f_0(u) + \sum_{i=1}^n \lambda_i f_i(u) = \inf_V \left\{ f_0(v) + \sum_{i=1}^n \lambda_i f_i(v) \; : \; v \in V \right\},$$

which expresses that u is a solution of the unconstrained minimization problem

$$\min \left\{ f_0(v) + \sum_{i=1}^n \lambda_i f_i(v) : v \in V \right\}.$$

As a consequence,

$$\partial f_0(u) + \sum_{i=1}^n \lambda_i f_i(u) \ni 0,$$

which, together with $\lambda_i \geq 0$, $f_i(u) \leq 0$, and $\lambda_i f_i(u) = 0$, tells us that λ is a Lagrange multiplier vector for u. $\square$

Clearly, the set M is independent of u solution of $(\mathcal{P})$. Thus we can speak of the set of Lagrange multipliers of a convex program. Indeed, the definition of the set M makes sense, and the set M may be nonempty, even when there is no solution of the convex program $(\mathcal{P})$. This leads us to give the following definition.

Definition 9.6.2. *For a given convex program*

$$\inf \left\{ f_0(v) \; : \; f_i(v) \leq 0, \qquad i = 1, \dots, n \right\}, \tag{$\mathcal{P}$}$$

the set M of generalized Lagrange multiplier vectors is defined by

$$M = \left\{ \lambda \in \mathbf{R}^n_+ \; : \; \inf_V (f_0 + \delta_C) = \inf_V \left(f_0 + \sum_{i=1}^n \lambda_i f_i \right) \right\},$$

where $C = \left\{ v \in V \; : \; f_i(v) \leq 0, \; i = 1, \dots, n \right\}$. When the problem $(\mathcal{P})$ has a solution, then M is the set of Lagrange multiplier vectors for $(\mathcal{P})$.

Without ambiguity, in what follows we will omit the word "generalized." We are going to characterize the set M by using marginal analysis.

Definition 9.6.3. *The value function attached to a convex program*

$$\inf \left\{ f_0(v) \; : \; f_i(v) \leq 0 \; \forall \; i = 1, \dots, n \right\} \tag{$\mathcal{P}$}$$

is the function $p : \mathbf{R}^n \to \overline{\mathbf{R}}$ which is defined, for every $y = (y_1, y_2 \dots, y_n) \in \mathbf{R}^n$, by

$$p(y) := \inf \left\{ f_0(v) \; : \; f_i(v) \leq y_i \; \forall \; i = 1, \dots, n \right\}.$$

The function p is also called the marginal function.

Let us observe that the value function is the optimal value of the perturbed convex program $(\mathcal{P}_y)$

$$\inf \left\{ f_0(v) \; : \; f_i(v) \leq y_i \; \forall \; i = 1, \dots, n \right\}. \tag{$\mathcal{P}_y$}$$

The initial problem, or unperturbed problem, corresponds to the case $y = 0$, i.e., $(\mathscr{P}) = (\mathscr{P}_0)$. We also notice that the value function may take the value $-\infty$, which may be a source of difficulties.

We are going to show that Lagrange multiplier vectors for problem $(\mathscr{P})$ correspond to subgradients of the value function p.

Theorem 9.6.2. *Consider the convex minimization problem*

$$\inf\{f_0(v) \,:\, f_i(v) \le 0 \ \forall\, i = 1,\dots,n\} \tag{$\mathscr{P}$}$$

and its value function $p : \mathbf{R}^n \to \overline{\mathbf{R}}$

$$p(y) = \inf\{f_0(v) \,:\, f_i(v) \le y_i \ \forall\, i = 1,\dots,n\}.$$

Then the following properties hold:

(a) *the value function p is convex;*

(b) *if $p(0) \in \mathbf{R}$, then $M = -\partial p(0)$, i.e., the set of generalized Lagrange multiplier vectors for $(\mathscr{P})$ is equal to the opposite of the subdifferential of p at the origin;*

(c) *if $p(0) \in \mathbf{R}$ and the Slater qualification assumption is satisfied, then p is continuous at the origin and, as a consequence, M is a nonempty, closed, convex, bounded set in $\mathbf{R}^n_+$.*

PROOF. (a) We notice that

$$p(y) = \inf_{v\in V} f(v,y),$$

where $f(v,y) = f_0(v) + \delta_{C(y)}(v)$ and $C(y) = \{v \in V \,:\, f_i(v) \le y_i \text{ for all } i = 1,\dots,n\}$. Let us verify that the mapping $(v,y) \mapsto \delta_{C(y)}(v)$ is convex.

We just need to verify that for every $(u,z) \in V \times \mathbf{R}^n$ and $(v,y) \in V \times \mathbf{R}^n$ such that $u \in C(z)$ and $v \in C(y)$, we still have $\lambda u + (1-\lambda)v \in C(\lambda z + (1-\lambda)y)$ for all $\lambda \in [0,1]$.

Indeed, this is an immediate consequence of the convexity of functions f_i: we have

$$\begin{aligned} f_i(\lambda u + (1-\lambda)v) &\le \lambda f_i(u) + (1-\lambda) f_i(v) \\ &\le \lambda z_i + (1-\lambda) y_i = (\lambda z + (1-\lambda)y)_i. \end{aligned}$$

Since f_0 is convex, we obtain that f is convex with respect to the pair (v,y). The convexity of the value function p is then a consequence of Proposition 9.2.3.

(b) We first prove that every generalized Lagrange multiplier vector $\lambda \in \mathbf{R}^n_+$ satisfies $-\lambda \in \partial p(0)$. Equivalently, we need to prove that

$$\forall y \in \mathbf{R}^n \quad p(y) \ge p(0) - \sum_{i=1}^n \lambda_i y_i.$$

By definition of p and by Definition 9.6.2 of generalized Lagrange multiplier vectors, we have

$$\begin{aligned} p(0) &= \inf\{f_0 + \delta_C\} \\ &= \inf\left\{ f_0 + \sum_{i=1}^n \lambda_i f_i \right\}. \end{aligned}$$

Take an arbitrary $y \in \mathbf{R}^n$ and denote by $C(y)$ the set

$$C(y) = \{v \in V \,:\, f_i(v) \le y_i, \qquad i = 1,\dots,n\}.$$

For every $v \in C(y)$ we have $\sum_{i=1}^n \lambda_i f_i(v) \le \sum_{i=1}^n \lambda_i y_i$ (recall that $\lambda_i \ge 0$ for all $i = 1,\dots,n$). Hence, for all $v \in C(y)$,

$$p(0) \le f_0(v) + \sum_{i=1}^n \lambda_i y_i.$$

As a consequence, by taking the infimum with respect to $v \in C(y)$, we obtain

$$p(0) \le p(y) + \sum_{i=1}^n \lambda_i y_i.$$

Let us now prove that, conversely, if $-\lambda \in \partial p(0)$, then λ is a generalized Lagrange multiplier vector for $(\mathscr{P})$.

We first prove that $\lambda \in \mathbf{R}^n_+$. Indeed, for every $y \in \mathbf{R}^n_+$ we have $C \subset C(y)$, and as a consequence

$$p(y) \le p(0).$$

Combining this inequality and the subdifferential inequality

$$p(y) \ge p(0) - \sum_{i=1}^n \lambda_i y_i,$$

we obtain

$$\sum_{i=1}^n \lambda_i y_i \ge 0.$$

This being true for all $y \in \mathbf{R}^n_+$, we obtain that $\lambda \in \mathbf{R}^n_+$.

Let us now prove that

$$\inf_V (f_0 + \delta_C) = \inf_V \left(f_0 + \sum_{i=1}^n \lambda_i f_i \right).$$

Equivalently, we need to prove that

$$p(0) = \inf_V \left(f_0 + \sum_{i=1}^n \lambda_i f_i \right).$$

The inequality $p(0) \ge \inf_V \left(f_0 + \sum_{i=1}^n \lambda_i f_i\right)$ is always true for arbitrary $\lambda \in \mathbf{R}^n_+$: indeed, for every $v \in C$, we have $f_i(v) \le 0$ and hence $\lambda_i f_i(v) \le 0$. This immediately yields

$$\begin{aligned} \inf_V \left(f_0 + \sum_{i=1}^n \lambda_i f_i \right) &\le \inf_C \left(f_0 + \sum_{i=1}^n \lambda_i f_i \right) \\ &\le \inf_C f_0 = p(0). \end{aligned}$$

The opposite inequality $p(0) \le \inf\left(f_0 + \sum_{i=1}^n \lambda_i f_i\right)$ relies on the fact that $-\lambda \in \partial p(0)$. We thus have for each $y \in \mathbf{R}^n$

$$p(y) + \sum_{i=1}^n \lambda_i y_i \ge p(0).$$

Take an arbitrary $v \in V$ and choose correspondingly $y_i = f_i(v)$ for all $i = 1, \dots, n$. Thus, we have $v \in C(y)$ and $p(y) \le f_0(v)$. As a consequence,

$$f_0(v) + \sum_{i=1}^{n} \lambda_i f_i(v) \ge p(0).$$

This being true for all $v \in V$, by taking the infimum with respect to v, we obtain

$$\inf_V \left(f_0 + \sum_{i=1}^{n} \lambda_i f_i \right) \ge p(0).$$

Finally, we have proved that $\lambda \in \mathbf{R}^n_+$ and

$$\inf_V (f_0 + \delta_C) = \inf_V \left(f_0 + \sum_{i=1}^{n} \lambda_i f_i \right).$$

By Definition 9.6.2, λ is a generalized Lagrange multiplier vector.

(c) By the Slater qualification assumption, there exists some $v_0 \in \operatorname{dom} f_0$ such that $f_i(v_0) < 0$ for all $i = 1, \dots, n$. Thus, we can find a neighborhood of the origin in $\mathbf{R}^n$, say, $B(0, r)$ with $r > 0$, such that

$$\forall y \in B(0, r),\ \forall\ i = 1, \dots, n \quad f_i(v_0) < y_i.$$

(It is enough to take, for example, $r = \frac{1}{2} \inf \left\{ |f_i(v_0)| \ :\ i = 1, \dots, n \right\}$.)

By definition of the value function p, we have

$$\forall y \in B(0, r) \quad p(y) \le f_0(v_0).$$

Since $f_0(v_0) < +\infty$, p is bounded from above on the ball $B(0, r)$. Let us prove that this property, together with $p(0) \in \mathbf{R}$, implies

$$\forall y \in \mathbf{R}^n \quad p(y) > -\infty.$$

We first formulate the properties above in terms of epigraphs. We have

$$B(0, r) \times [f_0(v_0), +\infty[\subset \operatorname{epi} p.$$

If $p(y) = -\infty$ for some $y \in \mathbf{R}^n$, we would have

$$\{y\} \times \mathbf{R} \subset \operatorname{epi} p.$$

Take $\xi = -\alpha y$ with $\alpha > 0$ to have $|\xi| < r$, for example, $\alpha = r/(2|y|)$. We can write $\alpha = (1-\lambda)/\lambda$ for some $0 < \lambda < 1$, which gives $\lambda \xi + (1-\lambda) y = 0$. Then we observe that

$$\begin{cases} (\xi, f_0(v_0)) \in B(0, r) \times [f_0(v_0), +\infty[, \\ (y, t) \in \{y\} \times \mathbf{R} \quad \text{for every } t \in \mathbf{R}. \end{cases}$$

By the convexity of epi p we obtain

$$\big(\lambda \xi + (1-\lambda) y, \lambda f_0(v_0) + (1-\lambda) t \big) \in \operatorname{epi} p,$$

i.e.,

$$p(0) \le \lambda f_0(v_0) + (1-\lambda) t \quad \text{for every } t \in \mathbf{R}.$$

Since $0 < \lambda < 1$, this implies $p(0) = -\infty$, a contradiction.

We can now apply Theorem 9.2.2: the function $p : \mathbf{R}^n \to \mathbf{R} \cup \{+\infty\}$ is convex and majorized on a neighborhood of the origin. Hence, p is continuous at the origin. By Proposition 9.5.2, $\partial p(0) \neq \emptyset$, and the set $M = -\partial p(0)$ is a nonempty closed convex bounded set in $\mathbf{R}^n_+$. □

Let us now introduce a dual minimization problem $(\mathscr{P}^*)$ to the convex program $(\mathscr{P})$ and show that the generalized Lagrange multiplier vectors are the solutions of this dual problem $(\mathscr{P}^*)$.

Theorem 9.6.3 (dual convex program). *Let us consider a convex program*

$$\inf\{f_0(v) \,:\, f_i(v) \leq 0 \;\forall\, i = 1, \ldots, n\} \tag{$\mathscr{P}$}$$

and let $p : \mathbf{R}^n \to \overline{\mathbf{R}}$ be its value function. We assume that $p(0) \in \mathbf{R}$ and that the Slater qualification assumption holds. Then the following hold:

(a) *The generalized Lagrange multiplier vectors of $(\mathscr{P})$ are the solutions of the maximization problem*

$$\sup\{-p^*(-\lambda) \,:\, \lambda \in \mathbf{R}^n_+\}, \tag{$\mathscr{P}^*$}$$

which is called the dual problem of $(\mathscr{P})$. The set of solutions of $(\mathscr{P}^*)$ is a nonempty closed convex bounded subset of $\mathbf{R}^n_+$.

(b) *For every $\lambda \in \mathbf{R}^n_+$, the equality*

$$-p^*(-\lambda) = \inf_{v \in V} \left\{ f_0(v) + \sum_{i=1}^{n} \lambda_i f_i(v) \right\}$$

holds and, as a consequence, the dual problem $(\mathscr{P}^*)$ can be written in the form

$$\sup_{\lambda \in \mathbf{R}^n_+} \inf_{v \in V} \left\{ f_0(v) + \sum_{i=1}^{n} \lambda_i f_i(v) \right\}. \tag{$\mathscr{P}^*$}$$

PROOF. By Theorem 9.6.2, we have the following equivalence:

$$\lambda \in M \iff -\lambda \in \partial p(0).$$

We know that p is a convex function. Thus, by using the Fenchel extremality relation (Proposition 9.5.1), we obtain

$$\lambda \in M \iff p(0) + p^*(-\lambda) = 0.$$

Theorem 9.6.2 also tells us that p is continuous at the origin. By Proposition 9.3.2, we thus have $p(0) = p^{**}(0)$. Noticing that $p^{**}(0) = \sup\{-p^*(\mu) \,:\, \mu \in \mathbf{R}^n\}$, we obtain

$$\begin{aligned} \lambda \in M \iff -p^*(-\lambda) &= \sup_{\mu \in \mathbf{R}^n} -p^*(\mu) \\ &= \sup_{\mu \in \mathbf{R}^n} -p^*(-\mu). \end{aligned}$$

Thus, $\lambda \in M$ iff λ is a solution of the maximization problem

$$\sup_{\mu \in \mathbf{R}^n} -p^*(-\mu).$$

Let us now compute p^*:

$$\begin{aligned}p^*(\mu) &= \sup_y \left\{ \langle \mu, y\rangle - p(y) \right\} \\ &= \sup_y \left\{ \langle \mu, y\rangle - \inf\left\{ f_0(v) \,:\, f_i(v) \le y_i \ \forall\, i=1,\dots,n \right\} \right\} \\ &= \sup_y \left\{ \langle \mu, y\rangle - f_0(v) \,:\, y\in \mathbf{R}^n,\ f_i(v) \le y_i \ \forall\, i=1,\dots,n \right\}.\end{aligned}$$

If for some $i \in \{1,\dots,n\}$ we have $\mu_i > 0$, then $p^*(\mu) = +\infty$. Otherwise, when $-\mu \in \mathbf{R}^n_+$ we have

$$\begin{aligned}p^*(\mu) &= \sup_{v\in V} \left\{ -f_0(v) + \sup_{y_i \ge f_i(v)} \sum_{i=1}^n \mu_i y_i \right\} \\ &= \sup_{v\in V} \left\{ \sum_{i=1}^n \mu_i f_i(v) - f_0(v) \right\}.\end{aligned}$$

Hence

$$-p^*(-\mu) = \begin{cases} \inf_{v\in V} \left\{ f_0(v) + \sum_{i=1}^n \mu_i f_i(v) \right\} & \text{if } \mu \in \mathbf{R}^n_+, \\ -\infty & \text{otherwise,} \end{cases}$$

and the dual problem can be written

$$\sup_{\mu\in\mathbf{R}^n_+} \inf_{v\in V} \left\{ f_0(v) + \sum_{i=1}^n \mu_i f_i(v) \right\}. \qquad \square$$

9.6.3 ▪ The Lagrangian approach to duality

In the framework of convex problems, and thanks to the Legendre–Fenchel transform, we have seen that a large number of mathematical objects can be paired with a dual one. Indeed we are going to go further and see how to realize the duality of optimization problems themselves. In the previous section, we introduced a variational problem $(\mathscr{P}^*)$, called the dual problem of $(\mathscr{P})$. We are going to justify this terminology and explore how the primal problem $(\mathscr{P})$ and its dual $(\mathscr{P}^*)$ are related to each other in remarkable ways.

Let us introduce some basic notations and concepts. The convex program

$$\inf\left\{ f_0(v) \,:\, f_i(v) \le 0 \ \forall\, i=1,\dots,n \right\} \qquad (\mathscr{P})$$

is called the *primal problem*.

The key notion in the duality theory for optimization problems is the *Lagrangian*.

Definition 9.6.4. *The Lagrangian function attached to the convex program* $(\mathscr{P})$ *is the function* $\mathbf{L} : V \times \mathbf{R}^n_+ \to \mathbf{R} \cup \{+\infty\}$ *defined by*

$$\mathbf{L}(v,\lambda) = f_0(v) + \sum_{i=1}^n \lambda_i f_i(v).$$

We already noticed that this expression plays a central role in the theory of Lagrange multipliers. The new aspect in the definition above is to consider this expression as a bivariate function, i.e., $\mathbf{L}$ is a function of the two variables v and λ. This is a big step, since we are no longer concerned only with the primal problem, its solutions, and the characterization of its solutions: we choose to give from the very beginning an equivalent status to the variables v and λ.

Let us first observe that the Lagrangian function $\mathbf{L}$ encapsulates all the information of the primal problem $(\mathscr{P})$. Clearly

$$\sup_{\lambda \in \mathbf{R}^n_+} \mathbf{L}(v,\lambda) = f_0(v) + \delta_C(v),$$

where $C = \left\{ v \in V \ : \ f_i(v) \le 0 \text{ for all } i = 1,\dots,n \right\}$ is the set of constraints. Thus, the primal problem can be equivalently written as an inf-sup problem, namely,

$$\inf_{v \in V} \sup_{\lambda \in \mathbf{R}^n_+} \mathbf{L}(v,\lambda). \tag{$\mathscr{P}$}$$

Let us denote by $\alpha \in \overline{\mathbf{R}}$ the optimal value of $(\mathscr{P})$

$$\alpha := \inf_{v \in V} \sup_{\lambda \in \mathbf{R}^n_+} \mathbf{L}(v,\lambda);$$

α is called the *primal value*.

This formulation of $(\mathscr{P})$ makes it rather natural to consider the associated variational problem

$$\sup_{\lambda \in \mathbf{R}^n_+} \inf_{v \in V} \mathbf{L}(v,\lambda), \tag{$\mathscr{P}^*$}$$

called the *dual problem*, which is obtained by interchanging the order of the sup and inf operators.

Indeed, this formulation fits perfectly with the conclusion of Theorem 9.6.3, where it is shown that, under some assumptions, Lagrange multipliers are solutions of the dual problem $(\mathscr{P}^*)$. Let us denote by $\beta \in \overline{\mathbf{R}}$ the optimal value of $(\mathscr{P}^*)$

$$\beta := \sup_{\lambda \in \mathbf{R}^n_+} \inf_{v \in V} \mathbf{L}(v,\lambda);$$

β is called the *dual value*.

The dual problem therefore consists in maximizing over vectors $\lambda \in \mathbf{R}^n_+$ the *dual function*

$$d(\lambda) := \inf_{v \in V} \mathbf{L}(v,\lambda).$$

Note that the dual problem $(\mathscr{P}^*)$ is well defined without any assumptions on the functions f_i, $i = 1,\dots,n$.

One has always $\beta \le \alpha$, because $\sup_X \inf_Y \le \inf_Y \sup_X$ is always true. It can happen that the primal value α is strictly larger than the dual value β. In this case, we say that there is a *duality gap*.

A basic question is to find conditions ensuring that there is no duality gap. When this is the case, the primal and the dual problem are connected through a rich calculus involving value functions, Legendre–Fenchel transform and subdifferentials, and minimax and saddle value problems.

The notion of the saddle value and the saddle point of the Lagrangian function is also fundamental: it permits us to treat in a unifying way the primal and the dual aspects of optimization problems and will allow us to develop all these ideas in a far more general setting in the next section.

Definition 9.6.5. *Let $L : X \times Y \to \overline{\mathbf{R}}$ be a bivariate function where X and Y are arbitrary spaces. A point $(\overline{x},\overline{y}) \in X \times Y$ is called a saddle point of L if*

$$\max_{y\in Y} L(\overline{x},y) = L(\overline{x},\overline{y}) = \min_{x\in X} L(x,\overline{y}).$$

Equivalently, $(\overline{x},\overline{y})$ is a saddle point of L if

$$L(\overline{x},y) \le L(\overline{x},\overline{y}) \le L(x,\overline{y}) \qquad \forall\, x \in X,\ y \in Y.$$

Another way to say this is

$$\begin{cases} \text{(a) } \overline{x} \text{ is a solution of the minimization problem } \inf_{x\in X} L(x,\overline{y}), \\ \text{(b) } \overline{y} \text{ is a solution of the maximization problem } \max_{y\in Y} L(\overline{x},y). \end{cases}$$

Note that the existence of a saddle point $(\overline{x},\overline{y})$ implies that there is no duality gap. This follows from the equalities

$$\sup_{y\in Y} L(\overline{x},y) = L(\overline{x},\overline{y}) = \inf_{x\in X} L(x,\overline{y}),$$

which imply

$$\alpha = \inf_x \sup_y L(x,y) \le L(\overline{x},\overline{y}) \le \sup_y \inf_x L(x,y) = \beta.$$

Since $\alpha \ge \beta$ is always true we obtain

$$L(\overline{x},\overline{y}) = \inf_x \sup_y L(x,y) = \sup_y \inf_x L(x,y).$$

The converse is not true in general: it is possible to have no duality gap without the existence of saddle points.

We can now reformulate the conclusions of Theorem 9.6.3 in the following form.

Theorem 9.6.4. *Consider a convex program $(\mathscr{P})$ and assume that the Slater condition holds. Then the following facts hold true:*

(a) *There is no duality gap, i.e., the primal and the dual values are equal; let us call it the optimal value.*

(b) *(dual attainment) Assuming moreover that the optimal value is finite, then the set of solutions of the dual problem $(\mathscr{P}^*)$ is nonempty: it is the set of generalized Lagrange multipliers of problem $(\mathscr{P})$, and it is convex and bounded.*

(c) *(saddle point formulation of primal solutions) The following assertions are equivalent:*

(i) *u is a solution of the primal problem $(\mathscr{P})$;*

(ii) *there exists a vector $\lambda \in \mathbf{R}^n_+$ such that (u,λ) is a saddle point of the Lagrangian function* $\mathbf{L}$.

If (ii) *is satisfied, then λ is a Lagrange multiplier of the optimal solution u, and it is a solution of the dual problem $(\mathscr{P}^*)$.*

PROOF. (a) If $\alpha=-\infty$, there is nothing to prove, since we know that $\alpha\geq\beta$. Otherwise, if α is finite, we are in the situation which was studied in Theorem 9.6.2: the Slater condition implies $\partial p(0)\neq\emptyset$ and the set M of generalized Lagrange multiplier vectors is nonempty. For every $\overline{\lambda}\in M$ we have

$$\begin{aligned}\alpha&=\inf_v\sup_\lambda \mathbf{L}(v,\lambda)\\&=\inf_v \mathbf{L}(v,\overline{\lambda})\\&\leq\sup_\lambda\inf_v \mathbf{L}(v,\lambda)=\beta.\end{aligned}$$

Since $\alpha\geq\beta$ is always true, we obtain $\alpha=\beta$.

(b) It is just a reformulation of Theorem 9.6.3: the set of solutions of the dual problem $(\mathscr{P}^*)$ has been characterized with the help of the value function:

$$\begin{aligned}\lambda \text{ solution of } (\mathscr{P}^*) &\iff -\lambda\in\partial p(0)\\&\iff \lambda \text{ generalized Lagrange multiplier of } (\mathscr{P}).\end{aligned}$$

(c) Let us first prove the implication (i) $\Longrightarrow$ (ii).

Let u be a solution of problem $(\mathscr{P})$. Then, the Slater condition implies the existence of a Lagrange multiplier $\overline{\lambda}$ associated to u (Theorem 9.6.1 and Proposition 9.6.3). Therefore,

$$\mathbf{L}(u,\overline{\lambda})=\min_{v\in V}\mathbf{L}(v,\overline{\lambda}).$$

On the other hand, because of the complementary slackness property ($\overline{\lambda}_i f_i(u)=0$ for all $i=1,\dots,n$), we have for any $\lambda\in\mathbf{R}^n_+$

$$\begin{aligned}\mathbf{L}(u,\overline{\lambda})&=f_0(u)+\sum_{i=1}^n\overline{\lambda}_i f_i(u)\\&=f_0(u)\\&\geq f_0(u)+\sum_{i=1}^n\lambda_i f_i(u)\end{aligned}$$

(use that $\lambda_i\geq 0$ and $f_i(u)\leq 0$). Hence

$$\mathbf{L}(u,\overline{\lambda})\geq\sup_{\lambda\in\mathbf{R}^n_+}\mathbf{L}(u,\lambda).$$

Finally

$$\inf_{v\in V}\mathbf{L}(v,\overline{\lambda})\geq\mathbf{L}(u,\overline{\lambda})\geq\sup_{\lambda\in\mathbf{R}^n_+}\mathbf{L}(u,\lambda),$$

which expresses that $(u,\overline{\lambda})$ is a saddle point of $\mathbf{L}$.

Let us now prove the implication (ii) $\Longrightarrow$ (i). If $(u,\overline{\lambda})$ is a saddle point of $\mathbf{L}$ on $V\times\mathbf{R}^n_+$, we have

$$\begin{aligned}f_0(u)+\delta_C(u)=\sup_{\lambda\in\mathbf{R}^n_+}\mathbf{L}(u,\lambda)&\leq\mathbf{L}(u,\overline{\lambda})\\&\leq\inf_v\mathbf{L}(v,\overline{\lambda})\leq\sup_{\lambda\in\mathbf{R}^n_+}\inf_v\mathbf{L}(v,\lambda)\\&\leq\inf_v\sup_{\lambda\in\mathbf{R}^n_+}\mathbf{L}(v,\lambda)=\inf_v\big(f_0(v)+\delta_C(v)\big).\end{aligned}$$

Hence, u is a solution of the primal problem $(\mathcal{P})$ and

$$\inf_v \big(f_0+\delta_C\big)=\inf_v\left(f_0+\sum_{i=1}^n \overline{\lambda}_i f_i\right),$$

which expresses (see Proposition 9.6.3) that $\overline{\lambda}$ is a Lagrange multiplier and, hence, $\overline{\lambda}$ is a solution of the dual problem. □

9.6.4 ▪ Duality for linear programming

Take $V=\mathbf{R}^n$. Given vectors $a^1, a^2, \dots, a^m$, c in $\mathbf{R}^n$, and a vector b in $\mathbf{R}^m$ consider the primal linear program

$$\inf\big\{\langle c,x\rangle \ : \ \langle a^i,x\rangle - b_i \le 0,\ i=1,\dots,m\big\}, \tag{$\mathcal{P}$}$$

where $\langle\cdot,\cdot\rangle$ is the usual Euclidean scalar product in $\mathbf{R}^n$. This is clearly a problem of linear programming, with

$$\begin{cases} f_0(x)=\langle c,x\rangle, \\ f_i(x)=\langle a^i,x\rangle-b_i, \qquad i=1,\dots,m. \end{cases}$$

Indeed, f_0 is linear and $C=\big\{x\in\mathbf{R}^n \ : \ \langle a^i,x\rangle-b_i\le 0\big\}$ is a polyhedral set (finite intersection of closed half-spaces). The Lagrangian function $\mathbf{L}:\mathbf{R}^n\times\mathbf{R}^m_+\to\mathbf{R}$ is given by

$$\mathbf{L}(x,\lambda)=\langle c,x\rangle+\sum_{i=1}^m \lambda_i\big(\langle a^i,x\rangle-b_i\big).$$

Therefore, the primal problem can be rewritten as

$$\inf_x \sup_{\lambda\in\mathbf{R}^m_+} \mathbf{L}(x,\lambda) \tag{$\mathcal{P}$}$$

and the dual problem $(\mathcal{P}^*)$ is given by

$$\sup_{\lambda\in\mathbf{R}^m_+} \inf_x \mathbf{L}(x,\lambda). \tag{$\mathcal{P}^*$}$$

Let us compute the dual function

$$\begin{aligned} d(\lambda)&=\inf_x \mathbf{L}(x,\lambda)\\ &=\inf_{x\in\mathbf{R}^n}\left(\left\langle x, c+\sum_{i=1}^m \lambda_i a^i\right\rangle-\sum_{i=1}^m \lambda_i b_i\right). \end{aligned}$$

We find

$$d(\lambda)=\begin{cases} -\sum_{i=1}^m \lambda_i b_i & \text{if } c+\sum_{i=1}^m \lambda_i a^i=0,\\ -\infty & \text{otherwise.}\end{cases}$$

The dual problem $(\mathcal{P}^*)$ is then given by

$$\begin{cases} \sup -\langle b,\lambda\rangle \\ \text{subject to} \quad \begin{cases} \sum_{i=1}^m \lambda_i a^i=-c, \\ \lambda\in\mathbf{R}^m_+, \end{cases} \end{cases} \tag{$\mathcal{P}^*$}$$

and the Kuhn–Tucker optimality conditions are the following:

$$\begin{cases} \sum_{i=1}^m \lambda_i a^i = -c, \\ \lambda_i \geq 0, \ x \in \mathbf{R}^n, \\ \langle a^i, x\rangle - b_i \leq 0, \\ \lambda_i\big(\langle a^i, x\rangle - b_i\big) = 0. \end{cases}$$

9.7 ▪ A general approach to duality in convex optimization

In Section 9.6, we developed a duality theory for convex programs. This is an important class of convex optimization problems, but it is far from covering the whole field of convex optimization. Thus a number of natural questions arise: Is it possible to develop a duality theory for general convex optimization, and if yes, is there a unique dual minimization problem? What are the relations between primal and dual problems, and what is the interpretation of the solutions of the dual problem?

At the center of all these questions is the notion of the Lagrangian function which we now introduce. Let us consider the primal problem

$$\inf\big\{f(v) \,:\, v \in V\big\}, \tag{$\mathscr{P}$}$$

where $f : V \to \mathbf{R}\cup\{+\infty\}$ is a general convex, lower semicontinuous, and proper function whose definition usually includes the constraints.

The basic idea is to introduce a bivariate function $\mathbf{L} : V \times W \to \overline{\mathbf{R}}$ which satisfies

$$\forall v \in V \quad f(v) = \sup_{w\in W} \mathbf{L}(v, w).$$

We say that $\mathbf{L}$ is a Lagrangian function associated to problem $(\mathscr{P})$. In this way, the primal problem can be written as an inf-sup problem:

$$\inf_{v\in V} \sup_{w\in W} \mathbf{L}(v, w). \tag{$\mathscr{P}$}$$

As we did in Section 9.6, one can associate to problem $(\mathscr{P})$ a dual problem $(\mathscr{P}^*)$ which is obtained by interchanging the order of inf and sup:

$$\sup_{w\in W} \inf_{v\in V} \mathbf{L}(v, w). \tag{$\mathscr{P}^*$}$$

Equivalently, $(\mathscr{P}^*)$ can be written as a maximization problem, in the form

$$\sup_{w\in W} d(w) \tag{$\mathscr{P}^*$}$$

with $d : W \to \overline{\mathbf{R}}$ being defined by

$$d(w) := \inf_{v\in V} \mathbf{L}(v, w).$$

We call $d(\cdot)$ the dual function (attached to the Lagrangian function $\mathbf{L}$). The interesting situation occurs when there is no duality gap, i.e.,

$$\inf(\mathscr{P}) = \sup(\mathscr{P}^*),$$

which is equivalent to saying

$$\inf_{v\in V}\sup_{w\in W}\mathbf{L}(v,w)=\sup_{w\in W}\inf_{v\in V}\mathbf{L}(v,w).$$

In this general abstract setting, we have the following result.

Proposition 9.7.1. *Let* $\mathbf{L}: V\times W\to\overline{\mathbf{R}}$ *be a general bivariate function. Then, the following facts are equivalent:*

(i) $(\overline{v},\overline{w})$ *is a saddle point of* $\mathbf{L}$*;*

(ii) $\overline{v}$ *is a solution of the primal problem* $(\mathscr{P})$*,* $\overline{w}$ *is a solution of the dual problem* $(\mathscr{P}^*)$*, and there is no duality gap:* $\inf(\mathscr{P})=\sup(\mathscr{P}^*)$*.*

PROOF. (i) $\Longrightarrow$ (ii) By definition of saddle point

$$\mathbf{L}(\overline{v},w)\le\mathbf{L}(\overline{v},\overline{w})\le\mathbf{L}(v,\overline{w})\qquad\forall v\in V,\ \forall w\in W.$$

Hence, for every $v\in V$

$$\sup_{w\in W}\mathbf{L}(\overline{v},w)\le\mathbf{L}(\overline{v},\overline{w})\le\mathbf{L}(v,\overline{w})\le\sup_{w\in W}\mathbf{L}(v,w).$$

This being true for all $v\in V$, we obtain that $\overline{v}$ is a solution of the minimization problem

$$\inf_{v\in V}\Big(\sup_{w\in W}\mathbf{L}(v,w)\Big),$$

which is precisely the primal problem $(\mathscr{P})$.

In a similar way, for every $w\in W$

$$\inf_{v\in V}\mathbf{L}(v,w)\le\mathbf{L}(\overline{v},w)\le\mathbf{L}(\overline{v},\overline{w})\le\inf_{v\in V}\mathbf{L}(v,\overline{w}).$$

Hence, $\overline{w}$ is a solution of the maximization problem

$$\sup_{w\in W}\Big(\inf_{v\in V}\mathbf{L}(v,w)\Big),$$

which is the dual problem $(\mathscr{P}^*)$.

(ii) $\Longrightarrow$ (i) Since $\overline{v}$ is a solution of the primal problem $(\mathscr{P})$, we have

$$\sup_{w\in W}\mathbf{L}(\overline{v},w)=\inf_{v\in V}\sup_{w\in W}\mathbf{L}(v,w).$$

Similarly, since $\overline{w}$ is a solution of the dual problem $(\mathscr{P}^*)$, we have

$$\inf_{v\in V}\mathbf{L}(v,\overline{w})=\sup_{w\in W}\inf_{v\in V}\mathbf{L}(v,w).$$

Since there is no duality gap, $\inf\sup=\sup\inf$ and we obtain

$$\sup_{w\in W}\mathbf{L}(\overline{v},w)=\inf_{v\in V}\mathbf{L}(v,\overline{w}),$$

that is,

$$\forall v\in V,\ \forall w\in W\quad\mathbf{L}(\overline{v},w)\le\mathbf{L}(v,\overline{w}).$$

This clearly implies $\mathbf{L}(\overline{v},\overline{w}) \le \mathbf{L}(v,\overline{w})$ for all $v \in V$, and $\mathbf{L}(\overline{v},w) \le \mathbf{L}(\overline{v},\overline{w})$ for all $w \in W$, i.e.,

$$\mathbf{L}(\overline{v},\overline{w}) = \min_{v} \mathbf{L}(v,\overline{w}) = \max_{w} \mathbf{L}(\overline{v},w),$$

which expresses that $(\overline{v},\overline{w})$ is a saddle point of $\mathbf{L}$. □

As we described above, duality theory for minimization problems follows very naturally from the Lagrangian formulation: it just consists in the permutation of the inf and the sup.

We stress the fact that the primal and the dual problems are intimately paired as soon as there is no duality gap and there exist saddle points of $\mathbf{L}$.

Thus the question is, for which class of bivariate functions $\mathbf{L}$ can one expect to have such properties? This is a central question in game theory, fixed point theory, economics, and so forth. Let us quote the celebrated Von Neumann's minimax theorem. Indeed, we give a slightly more general formulation to recover, as a particular case, the existence theorem for convex minimization problems (see Aubin [64], for example).

Theorem 9.7.1 (von Neumann's minimax theorem). *Let V and W be two reflexive Banach spaces and let $M \subset V$ and $N \subset W$ be two closed convex nonempty sets. Let $\mathbf{L} : M \times N \to \mathbf{R}$ be a bivariate function which satisfies the following properties:*

$$\begin{cases} (\mathrm{i_a})\ \forall w \in N,\ v \mapsto \mathbf{L}(v,w) \text{ is convex and lower semicontinuous,} \\ (\mathrm{i_b})\ \forall v \in M\ w \mapsto \mathbf{L}(v,w) \text{ is concave and upper semicontinuous.} \end{cases}$$

$$\begin{cases} (\mathrm{ii_a})\ M \text{ is bounded or there exists some } w_0 \in N \text{ such that } v \mapsto \mathbf{L}(v,w_0) \text{ is coercive,} \\ (\mathrm{ii_b})\ N \text{ is bounded or there exists some } v_0 \in M \text{ such that } w \mapsto -\mathbf{L}(v_0,w) \text{ is coercive.} \end{cases}$$

Then $\mathbf{L}$ possesses a saddle point $(\overline{v},\overline{w}) \in M \times N$, i.e.,

$$\min_{v\in M} \mathbf{L}(v,\overline{w}) = \mathbf{L}(\overline{v},\overline{w}) = \max_{w\in N} \mathbf{L}(\overline{v},w).$$

In particular $\inf_v \sup_w \mathbf{L}(v,w) = \sup_w \inf_v \mathbf{L}(v,w)$, i.e., there is no duality gap.

It follows from the previous results that the key property to developing a duality theory for optimization problems is the possibility to write the function f in the following form:

$$f(v) = \sup_{w\in W} \mathbf{L}(v,w)$$

with $\mathbf{L} : V \times W \to \overline{\mathbf{R}}$ a convex-concave bivariate function.

We are going to see how the Legendre–Fenchel transform permits us to produce such convex-concave Lagrangian functions in a systematic and elegant way.

The idea is first to introduce a perturbation function $F : V \times Y \to \mathbf{R} \cup \{+\infty\}$ such that $F(v,0) = f(v)$ for all $v \in V$. The primal problem $(\mathscr{P})$ can now be written as

$$\inf\big\{F(v,0) \,:\, v \in V\big\}. \qquad (\mathscr{P})$$

The key property which allows us to produce a convex-concave Lagrangian function from F is that F is convex with respect to $(v,w) \in V \times W$. For example, in convex programming, the duality scheme that was studied in Section 9.6 is associated with the perturbation function:

$$F(v,y) = \begin{cases} f_0(v) & \text{if } f_i(v) \le y_i, \quad i = 1,\dots,n, \\ +\infty & \text{otherwise.} \end{cases}$$

One can easily verify that when f_0 and f_i $(i = 1,\dots,n)$ are convex functions, so is F.

Let us now describe how one can associate a Lagrangian function to F.

Proposition 9.7.2 (definition of Lagrangian). *Let $F : V \times Y \to \mathbf{R} \cup \{+\infty\}$ be a convex function. We associate to F a Lagrangian function $\mathbf{L} : V \times Y^* \to \overline{\mathbf{R}}$ by the following formula:*

$$\forall v \in V,\ \forall y^* \in Y^* \quad -\mathbf{L}(v,y^*) = \sup_{y \in Y} \big\{ \langle y^*, y\rangle - F(v,y) \big\},$$

i.e., $-\mathbf{L}(v,\cdot)$ is the Legendre–Fenchel conjugate of $F(v,\cdot)$. We have that $\mathbf{L}$ is a convex-concave function. More precisely,

$$\begin{cases} (1)\ \textit{for all } v \in V,\ y^* \mapsto \mathbf{L}(v,y^*) \textit{ is concave, upper semicontinuous on } Y^*; \\ (2)\ \textit{for all } y^* \in Y^*,\ v \mapsto \mathbf{L}(v,y^*) \textit{ is convex.} \end{cases}$$

PROOF. The proof is immediate; just note that (2) follows from Proposition 9.2.3 and from the fact that the function $(v,y) \mapsto -\langle y^*, y\rangle + F(v,y)$ is convex. □

As expected, we can reformulate problem $(\mathscr{P})$ by using the Lagrangian function $\mathbf{L}$ attached to the (convex) perturbation function F.

Proposition 9.7.3. *Let $F : V \times Y \to \mathbf{R} \cup \{+\infty\}$ be a convex, lower semicontinuous, proper function and let $\mathbf{L} : V \times Y^* \to \overline{\mathbf{R}}$ be the corresponding Lagrangian function, given by*

$$-\mathbf{L}(v,y^*) = \sup_{y \in Y} \big\{ \langle y^*, y\rangle - F(v,y) \big\}.$$

(a) Primal problem: *we have*

$$\forall v \in V \quad F(v,0) = \sup_{y^* \in Y^*} \mathbf{L}(v,y^*).$$

As a consequence, with the notation $f(v) = F(v,0)$, the primal problem

$$\inf \big\{ f(v) \ :\ v \in V \big\} \tag{$\mathscr{P}$}$$

can be written as

$$\inf_{v \in V} \sup_{y^* \in Y^*} \mathbf{L}(v,y^*). \tag{$\mathscr{P}$}$$

(b) Dual problem: *the dual problem, which by definition is*

$$\sup_{y^* \in Y^*} \inf_{v \in V} \mathbf{L}(v,y^*), \tag{$\mathscr{P}^*$}$$

can be written as

$$\sup_{y^* \in Y^*} d(y^*), \tag{$\mathscr{P}^*$}$$

where the dual function $d : Y^* \to \mathbf{R} \cup \{+\infty\}$ is given by

$$\begin{aligned} d(y^*) &= \inf_{v \in V} \mathbf{L}(v,y^*) \\ &= -F^*(0,y^*), \end{aligned}$$

and F^* is the Legendre–Fenchel conjugate of F with respect to (v,y).

PROOF. (a) Since F is closed convex and proper on $V \times Y$, for every $v \in V$ the function $\varphi_v : y \mapsto F(v, y)$ is closed convex on Y. Hence, for all $v \in V$

$$\begin{aligned}\varphi_v(y) &= \varphi_v^{**}(y)\\ &= \sup_{y^* \in Y^*} \big\{\langle y^*, y\rangle - \varphi_v^*(y^*)\big\}.\end{aligned}$$

By definition of $\mathbf{L}$ we have

$$\begin{aligned}\varphi_v^*(y^*) &= \sup_{y \in Y} \big\{\langle y^*, y\rangle - \varphi_v(y)\big\}\\ &= \sup_{y \in Y} \big\{\langle y^*, y\rangle - F(v, y)\big\}\\ &= -\mathbf{L}(v, y^*).\end{aligned}$$

Hence,

$$\varphi_v(y) = \sup_{y^* \in Y^*} \big\{\langle y^*, y\rangle + \mathbf{L}(v, y^*)\big\}.$$

Take now $y = 0$ to obtain

$$f(v) = F(v, 0) = \varphi_v(0) = \sup_{y^* \in Y^*} \mathbf{L}(v, y^*).$$

(b) By definition, the dual function $d : Y^* \to \mathbf{R} \cup \{+\infty\}$ is equal to

$$d(y^*) = \inf_{v \in V} \mathbf{L}(v, y^*).$$

By definition of $\mathbf{L}$,

$$\begin{aligned}\mathbf{L}(v, y^*) &= -\sup_{y} \big\{\langle y^*, y\rangle - F(v, y)\big\}\\ &= \inf_{y} \big\{-\langle y^*, y\rangle + F(v, y)\big\}.\end{aligned}$$

Hence,

$$d(y^*) = \inf_{v \in V, y \in Y} \big\{-\langle y^*, y\rangle + F(v, y)\big\}.$$

Thus,

$$\begin{aligned}d(y^*) &= -\sup_{v \in V, y \in Y} \big\{\langle 0, v\rangle + \langle y^*, y\rangle - F(v, y)\big\}\\ &= -F^*(0, y^*),\end{aligned}$$

where F^* is the conjugate of F with respect to (v, y). □

The other fundamental mathematical object which is attached to the perturbation function F is the value function.

Proposition 9.7.4 (definition of the value function). *Let $F : V \times Y \to \mathbf{R} \cup \{+\infty\}$ be a convex function. The value function (also called marginal function) attached to F is the function $p : Y \to \mathbf{R} \cup \{+\infty\}$, which is defined by*

$$\forall y \in Y \quad p(y) := \inf_{v \in V} F(v, y).$$

It is a convex function. Moreover, for every $y^* \in Y^*$

$$p^*(y^*) = F^*(0, y^*).$$

Thus, the dual problem $(\mathscr{P}^*)$ can be formulated in terms of p as follows:

$$\sup_{y^* \in Y^*} (-p^*(y^*)) = -\inf_{y^* \in Y^*} p^*(y^*). \qquad (\mathscr{P}^*)$$

PROOF. The convexity of p follows from the convexity of F and by applying Proposition 9.2.3. For $y^* \in Y^*$ we have

$$\begin{aligned}
p^*(y^*) &= \sup_{y \in Y} \left\{ \langle y^*, y \rangle - p(y) \right\} \\
&= \sup_{y \in Y} \left\{ \langle y^*, y \rangle - \inf_{v \in V} F(v, y) \right\} \\
&= \sup_{(v,y) \in V \times Y} \left\{ \langle 0, v \rangle + \langle y^*, y \rangle - F(v, y) \right\} \\
&= F^*(0, y^*) \\
&= -d(y^*),
\end{aligned}$$

where d is the dual function introduced in Proposition 9.7.3(b). □

We have now all the ingredients for developing a general convex duality theory. The following theorem may be proved in the same way as Theorems 9.6.2 and 9.6.3. Let us notice that the qualification assumption

"there exists some $v_0 \in V$ such that $F(v_0, .)$ is finite and continuous at the origin"

plays the role of the Slater qualification assumption in convex programming. For this reason, we call it generalized Slater.

Theorem 9.7.2. *Let $f : V \to \mathbf{R} \cup \{+\infty\}$ be a closed convex and proper function such that* $\inf f > -\infty$. *Let $F : V \times Y \to \mathbf{R} \cup \{+\infty\}$ be a perturbation function which satisfies the following conditions:*

(i) $F(v, 0) = f(v) \quad \forall v \in V$.

(ii) *F is a closed convex proper function.*

(iii) *(Generalized Slater) there exists some $v_0 \in V$ such that $y \mapsto F(v_0, y)$ is finite and continuous at the origin.*

Then the following properties hold:

(a) *The value function p is continuous at the origin. As a consequence, the set of solutions of the dual problem $(\mathscr{P}^*)$, which is equal to $\partial p(0)$, is nonempty. Indeed, it is a nonempty closed convex and bounded subset of Y^*.*

(b) *There is no duality gap, i.e.,* $\inf(\mathscr{P}) = \max(\mathscr{P}^*)$.

(c) *Let $\overline{u}$ be a solution of the primal problem $(\mathscr{P})$. Then, for every element $\overline{y}^*$ which is a solution of the dual problem (by property* (a) *there exist such elements), $(\overline{u}, \overline{y}^*)$ is a saddle point of the Lagrangian function $\mathbf{L}$ associated to F. Conversely, if $(\overline{u}, \overline{y}^*)$ is a saddle point of $\mathbf{L}$, then $\overline{u}$ is a solution of $(\mathscr{P})$ and $\overline{y}$ is a solution of $(\mathscr{P}^*)$.*

(d) *$(\overline{u}, \overline{y}^*)$ is a saddle point of $\mathbf{L}$ iff it satisfies the extremality relation:*

$$F(\overline{u}, 0) + F^*(0, \overline{y}^*) = 0.$$

PROOF. (a) The generalized Slater condition implies the existence of some neighborhood of 0 in Y in which the function $y \mapsto F(v_0, y)$ is bounded from above: let $r > 0$ and $M \in \mathbf{R}$ be such that

$$F(v_0, y) \leq M \quad \forall\, y \in B_Y(0, r).$$

As a consequence, the value function p satisfies

$$p(y) = \inf_{v \in V} F(v, y) \leq F(v_0, y) \leq M \quad \forall\, y \in B_Y(0, r).$$

The function p also satisfies $p(0) = \inf(\mathscr{P})$ and p is bounded from above on a neighborhood of the origin. By Theorem 9.3.2 and Proposition 9.5.2 we obtain that p is subdifferentiable at 0, i.e., $\partial p(0) \neq \emptyset$. We know by Proposition 9.7.4 that the dual problem $(\mathscr{P}^*)$ can be expressed in terms of the value function p. Indeed, the fact that $\overline{y}^*$ is a solution of $(\mathscr{P}^*)$ is equivalent to saying that $\overline{y}^*$ is a solution of the minimization problem

$$\inf_{y^* \in Y^*} p^*(y^*).$$

Thus,

$$\begin{aligned} \overline{y}^* \text{ is a solution of } (\mathscr{P}^*) &\iff \partial p^*(y^*) \ni 0 \\ &\iff y^* \in \partial p(0) \quad \text{(Theorem 9.5.1)}. \end{aligned}$$

By the argument above, $\partial p(0)$ is a closed convex bounded nonempty subset of Y^*. Thus, there exist solutions of the dual problem $(\mathscr{P}^*)$ and the set of these solutions is a closed convex bounded subset of Y^*.

(b) Let $\overline{y}^*$ be any solution of the dual problem $(\mathscr{P}^*)$. We know by (a) that there exist such elements and that they are characterized by the relation $\overline{y}^* \in \partial p(0)$ or, equivalently, by the Fenchel extremality relation

$$p(0) + p^*(\overline{y}^*) = \langle \overline{y}^*, 0 \rangle = 0.$$

Hence

$$\begin{aligned} \inf(\mathscr{P}) = p(0) &= -p^*(\overline{y}^*) \\ &= \sup_{y^* \in Y^*} -p^*(y^*) = \sup(\mathscr{P}^*). \end{aligned}$$

(c) We use the Lagrangian formulation of problem $(\mathscr{P})$ and $(\mathscr{P}^*)$ given by Proposition 9.7.3:

$$\begin{cases} (\mathscr{P}) & \inf_{v \in V} \sup_{y^* \in Y^*} \mathbf{L}(v, y^*), \\ (\mathscr{P}^*) & \sup_{y^* \in Y^*} \inf_{v \in V} \mathbf{L}(v, y^*). \end{cases}$$

The characterization of pairs of optimal solutions $(\overline{u}, \overline{y}^*)$ of problems $(\mathscr{P})$ and $(\mathscr{P}^*)$, respectively, as saddle points of the Lagrangian $\mathbf{L}$ is a direct consequence of Proposition 9.7.1.

(d) Let $(\overline{u}, \overline{y}^*)$ be a saddle point of $\mathbf{L}$. Let us reformulate the extremality relation

$$p(0) + p^*(\overline{y}^*) = 0 \quad \text{(see (b) above)}$$

in terms of the function F: we have

$$\begin{cases} p(0) = \inf(\mathscr{P}) = F(\overline{u}, 0), \\ p^*(\overline{y}^*) = \sup(\mathscr{P}^*) = F^*(0, \overline{y}^*) \quad \text{(Proposition 9.7.4)}. \end{cases}$$

Hence
$$F(\overline{u},0)+F^*(0,\overline{y}^*)=0,$$
that is, $(0,\overline{y}^*)\in\partial F(\overline{u},0)$. $\square$

9.8 ▪ Duality in the calculus of variations: First examples

As a model example, let us consider the Dirichlet problem: given $h\in L^2(\Omega)$, find $u\in H_0^1(\Omega)$ such that
$$\begin{cases} -\Delta u = h & \text{in } \Omega,\\ u=0 & \text{on } \partial\Omega.\end{cases}$$
The variational formulation of this problem has been extensively studied in Chapter 5: the solution u of the Dirichlet problem is the minimizer, on the Sobolev space $H_0^1(\Omega)$, of the functional
$$f(v)=\frac{1}{2}\int_\Omega |\nabla v(x)|^2\,dx-\int_\Omega h(x)v(x)\,dx.$$
The primal problem $(\mathscr{P})$ can be expressed as
$$\min_{v\in H_0^1(\Omega)}\left\{\frac{1}{2}\int_\Omega |\nabla v(x)|^2\,dx-\int_\Omega h(x)v(x)\,dx\right\}. \qquad (\mathscr{P})$$
We now introduce the perturbation function
$$F(v,y)=\frac{1}{2}\int_\Omega |\nabla v(x)+y(x)|^2\,dx-\int_\Omega h(x)v(x)\,dx,$$
which we consider as a function $F: H_0^1(\Omega)\times L^2(\Omega)^N\to\mathbf{R}$.

To compute the Lagrangian function $\mathbf{L}$ associated to F and describe the corresponding dual problem, we start to analyze the structure of this problem.

The primal function f can be written as
$$f(v)=\Phi(Av)+\Psi(v),$$
where $\Phi: L^2(\Omega)^N\to\mathbf{R}$ is the convex integral functional
$$\Phi(w)=\frac{1}{2}\int_\Omega |w(x)|^2\,dx$$
and A is the gradient operator, which can be viewed as a linear continuous operator from $H_0^1(\Omega)$ into $L^2(\Omega)^N$. The functional $\Psi: H_0^1(\Omega)\to\mathbf{R}$ is the linear and continuous mapping
$$\Psi(v)=-\int_\Omega h(x)v(x)\,dx.$$
The perturbation function $F: H_0^1(\Omega)\times L^2(\Omega)^N\to\mathbf{R}$ can then be written as
$$F(v,y)=\Phi(Av+y)+\Psi(v).$$

Theorem 9.8.1. *Let V and Y be two Banach spaces and suppose $\Phi\in\Gamma_0(Y)$, $\Psi\in\Gamma_0(V)$, and $A\in L(V,Y)$ (A is a linear continuous operator). Consider the primal problem*
$$\inf_{v\in V}\left\{\Phi(Av)+\Psi(v)\right\} \qquad (\mathscr{P})$$

and the perturbation function $F : V \times Y \to \mathbf{R} \cup \{+\infty\}$ *defined by*

$$F(v,y) = \Phi(Av + y) + \Psi(v).$$

Then the following facts hold:

(a) *The Lagrangian function* $L : V \times Y^* \to \overline{\mathbf{R}}$ *associated to* F *is given by*

$$\mathbf{L}(v,y^*) = \Psi(v) + \langle y^*, Av\rangle - \Phi^*(y^*)$$

and the dual problem $(\mathscr{P}^*)$ *is equal to*

$$\sup_{y^* \in Y^*} \left\{ -\Psi^*(-A^* y^*) - \Phi^*(y^*) \right\}, \qquad (\mathscr{P}^*)$$

where A^* *is the adjoint operator of* A.

(b) *Let us assume that there exists some* $v_0 \in V$ *such that* $\Psi(v_0) < +\infty$, $\Phi(A(v_0)) < +\infty$ *and* Φ *is continuous at* Av_0. *Then* $(\mathscr{P}^*)$ *has at least one solution* $\overline{y}^*$. *If* $\overline{u}$ *is a solution of* $(\mathscr{P})$, *one has the following extremality relations:*

$$\begin{cases} -A^*\overline{y}^* \in \partial\Psi(\overline{u}), \\ \quad \overline{y}^* \in \partial\Phi(A\overline{u}). \end{cases}$$

PROOF. (a) By the definition of Lagrangian (Proposition 9.7.2),

$$\begin{aligned}
\mathbf{L}(v,y^*) &= \inf_{y\in Y} \left\{ -\langle y^*, y\rangle + F(v,y) \right\} \\
&= \inf_{y\in Y} \left\{ -\langle y^*, y\rangle + \Phi(Av+y) + \Psi(v) \right\} \\
&= \Psi(v) - \sup_{y\in Y} \left\{ \langle y^*, y\rangle - \Phi(Av+y) \right\} \\
&= \Psi(v) - \sup_{y\in Y} \left\{ \langle y^*, Av+y\rangle - \Phi(Av+y) - \langle y^*, Av\rangle \right\} \\
&= \Psi(v) + \langle y^*, Av\rangle - \Phi^*(y^*).
\end{aligned}$$

The perturbation function F is convex and lower semicontinuous: this is an immediate consequence of the facts that $\Phi \in \Gamma_0(Y)$, $\Psi \in \Gamma_0(V)$, and $A \in L(V,Y)$. Therefore, the primal and the dual problems can be expressed in terms of the Lagrangian function L and we have (Proposition 9.7.3)

$$\begin{cases} (\mathscr{P}) \quad \inf_{v\in V} \sup_{y^*\in Y^*} \mathbf{L}(v,y^*), \\ (\mathscr{P}^*) \quad \sup_{y^*\in Y^*} \inf_{v\in V} \mathbf{L}(v,y^*). \end{cases}$$

Let us compute the dual function d:

$$\begin{aligned}
d(y^*) &= \inf_{v\in V} \mathbf{L}(v,y^*) \\
&= \inf_{v\in V} \left\{ \Psi(v) + \langle y^*, Av\rangle - \Phi^*(y^*) \right\} \\
&= -\Phi^*(y^*) - \sup_{v\in V} \left\{ \langle -y^*, Av\rangle - \Psi(v) \right\} \\
&= -\Phi^*(y^*) - \sup_{v\in V} \left\{ \langle -A^*y^*, v\rangle - \Psi(v) \right\} \\
&= -\Phi^*(y^*) - \Psi^*(-A^*y^*).
\end{aligned}$$

Therefore, from the analysis made in Section 9.7, part (a) is proved.

(b) The existence of $v_0 \in V$ such that $\Psi(v_0) < +\infty$ with Φ continuous at Av_0 clearly implies that the generalized Slater condition is satisfied. Hence $(\mathscr{P}^*)$ admits at least a solution (Theorem 9.7.2). Let $\overline{y}^*$ be such a solution. Let $\overline{u}$ be a solution of the primal problem $(\mathscr{P})$. We know that there is no duality gap. Hence $\inf(\mathscr{P}) = \sup(\mathscr{P}^*)$, i.e.,

$$\Phi(A\overline{u}) + \Psi(\overline{u}) = -\Phi(\overline{y}^*) - \Psi^*(-A^*\overline{y}^*).$$

Thus

$$\Phi(A\overline{u}) + \Phi^*(\overline{y}^*) + \Psi(\overline{u}) + \Psi^*(-A^*\overline{y}^*) = 0.$$

Equivalently,

$$\big(\Phi(A\overline{u}) + \Phi^*(\overline{y}^*) - \langle \overline{y}^*, A\overline{u}\rangle\big) + \big(\Psi(\overline{u}) + \Psi^*(-A^*\overline{y}^*) + \langle A^*\overline{y}^*, \overline{u}\rangle\big) = 0.$$

Since by the Fenchel inequality the quantities $\Phi(A\overline{u}) + \Phi^*(\overline{y}^*) - \langle \overline{y}^*, A\overline{u}\rangle$ and $\Psi(\overline{u}) + \Psi^*(-A^*\overline{y}^*) + \langle A^*\overline{y}^*, \overline{u}\rangle$ are nonnegative, we obtain

$$\begin{cases} \Phi(A\overline{u}) + \Phi^*(\overline{y}^*) - \langle \overline{y}^*, A\overline{u}\rangle = 0, \\ \Psi(\overline{u}) + \Psi^*(-A^*\overline{y}^*) + \langle A^*\overline{y}^*, \overline{u}\rangle = 0. \end{cases}$$

These are the Fenchel extremality relations, which are equivalent to

$$\begin{cases} \overline{y}^* \in \partial\Phi(A\overline{u}), \\ -A^*\overline{y}^* \in \partial\Psi(\overline{u}). \end{cases} \qquad \square$$

Remark 9.8.1. It is often more convenient to write the dual problem as a minimization problem

$$\inf_{y^* \in Y^*} \big\{\Phi^*(y^*) + \Psi^*(-A^*y^*)\big\}.$$

Let us come back to the Dirichlet problem and apply the above results: we recall that $V = H_0^1(\Omega)$ and $Y = L^2(\Omega)^N$.

(a) $A : H_0^1(\Omega) \to L^2(\Omega)^N$ is the gradient operator. The adjoint operator

$$A^* : L^2(\Omega)^N \to H^{-1}(\Omega) = H_0^1(\Omega)^*$$

is defined by

$$\begin{aligned} \langle A^*y, v\rangle_{(H^{-1}, H_0^1)} &= \langle y, Av\rangle_{L^2(\Omega)^N} \\ &= \sum_{i=1}^N \int_\Omega y_i \frac{\partial v}{\partial x_i}\, dx \\ &= \left\langle -\sum_{i=1}^N \frac{\partial y_i}{\partial x_i}, v \right\rangle_{(\mathscr{D}'(\Omega), \mathscr{D}(\Omega))}, \end{aligned}$$

where in the last equality v varies in $\mathscr{D}(\Omega)^N$, that is, $A^*y = -\sum_{i=1}^N D_i y_i = -\operatorname{div} y$ in the distribution sense, i.e., A^* is the opposite of the divergence operator.

(b) We know by Theorem 9.3.3 that the conjugate of the function

$$\Phi(y) = \frac{1}{2}\int_\Omega |y(x)|^2\, dx$$

is equal to

$$\Phi^*(y^*) = \frac{1}{2}\int_\Omega |y^*(x)|^2\,dx.$$

(c) Let $\Psi(v) = -\int_\Omega h(x)v(x)\,dx$. An easy computation gives

$$\begin{aligned}\Psi^*(v^*) &= \sup_{v\in H_0^1(\Omega)} \left\{\langle v^*, v\rangle_{(H^{-1},H_0^1)} - \langle -h, v\rangle_{L^2(\Omega)}\right\}\\ &= \sup_{v\in H_0^1(\Omega)} \langle v^* + h, v\rangle_{(H^{-1},H_0^1)}\\ &= \begin{cases} 0 & \text{if } v^* + h = 0,\\ +\infty & \text{otherwise.}\end{cases}\end{aligned}$$

We now collect all these results to obtain, thanks to Theorem 9.8.1, the description of the dual problem $(\mathscr{P}^*)$ of the Dirichlet problem

$$\sup\left\{-\frac{1}{2}\int_\Omega |y^*(x)|^2\,dx \;:\; y^* \in L^2(\Omega)^N,\ \operatorname{div} y^* = h\right\}. \tag{$\mathscr{P}^*$}$$

Clearly, the generalized Slater condition is satisfied (F is everywhere continuous!). Thus there exists a solution $\overline{y}^*$ of $(\mathscr{P}^*)$ and this solution is unique because of the strict convexity of the mapping $y^* \mapsto \int_\Omega |y^*(x)|^2\,dx$. On the other hand, we know that $(\mathscr{P})$ admits a unique solution u. The extremality relations yield

$$\begin{cases} \overline{y}^* = Au = \operatorname{grad} u,\\ \operatorname{div}\overline{y}^* = -h,\end{cases}$$

i.e., we have $-\operatorname{div}(\operatorname{grad} u) = h$ which is in accordance with the definition of u.