4

Minima and quasi minima

Rationality of thought imposes
a limit on a person's concept
of his relation to the cosmos.
(J. F. Nash, Autobiography)

Convexity plays a key role in minimization. First of all, a local minimum is automatically a global one. Secondly, for convex functions, the classical Fermat necessary condition for a local extremum becomes sufficient to characterize a global minimum.

In this chapter we deal with the problem of existence of a minimum point, and thus we quite naturally begin with stating and commenting on the Weierstrass existence theorem. We also show that in reflexive (infinite dimensional) Banach spaces convexity is a very important property for establishing existence of a global minimum under reasonable assumptions. There are however several situations, for example outside reflexivity, where to have a general existence theorem for a wide class of functions is practically impossible. Thus it is important to know that at least for "many" functions in a prescribed class, an existence theorem can be provided. A fundamental tool for getting this type of result is the Ekeland variational principle, probably one of the most famous results in modern nonlinear analysis. So, in this chapter we spend some time in analyzing this variational principle, and deriving some of its interesting consequences, mainly in the convex setting.

The problem we were alluding to of identifying classes of functions for which "most" of the problems have solutions will be discussed in detail in Chapter 11. The chapter ends with the description of some properties of the level sets of a convex function, and with a taste of the algorithms that can be used in order to find the minima of a convex function, in a finite dimensional setting.

4.1 The Weierstrass theorem

The next result is the fundamental Weierstrass theorem.

Theorem 4.1.1 *Let* (X, τ) *be a topological space, and assume* $f\colon (X,\tau) \to (-\infty, \infty]$ *is* τ*-lower semicontinuous. Suppose moreover there is* $\bar{a} > \inf f$ *such that* $f^{\bar{a}}$ *is* τ*-compact. Then* f *has absolute minima:* $\operatorname{Min} f := \{\bar{x} : f(\bar{x}) \leq f(x), \forall x \in X\}$ *is a nonempty set.*

Proof.

$$\operatorname{Min} f = \bigcap_{\bar{a} > a > \inf f} f^a.$$

Each f^a is nonempty and τ-closed (due to τ-lower semicontinuity of f); hence

$$\{f^a : \bar{a} > a > \inf f\}$$

is a family of nonempty, nested, τ-compact sets, and this entails nonemptiness of their intersection. □

The previous theorem is surely a milestone in optimization. Thus, when we face an optimization problem, the challenge is to see if there is a topology τ on the set X in order to fulfill its assumptions. Observe that the two requested conditions, τ-lower semicontinuity of f, and having a τ-compact level set, go in opposite directions. Given a function f on X, in order to have f τ-lower semicontinuous we need many closed sets on X (i.e., the finer the topology τ with which we endow X, the better the situation), but to have a compact level set we need a topology rich in compact sets, which is the same as saying poor in open (and so, closed) sets. For instance, think of a continuous function (in the norm topology) defined on an infinite-dimensional Hilbert space. Clearly, each level set of f is a closed set. But also, no level set (at height greater than $\inf f$) is compact! To see this, observe that each f^a must contain a ball around a point x fulfilling $f(x) < a$. As is well known, compact sets in infinite-dimensional spaces do have empty interiors. Thus Weierstrass' theorem can *never* be applied in this setting, with the norm topology. Fortunately, we have other choices for the topology on the space. On the Banach space X, let us consider the weak topology. This is defined as the weakest topology making continuous all the elements of X^*, the continuous dual space of X. By the very definition, this topology is coarser than the norm topology, and strictly coarser in infinite dimensions, as it is not difficult to show. This implies that the weak topology will provide us more compact sets, but fewer closed sets. Thus, the following result is very useful.

Proposition 4.1.2 *Let* X *be a Banach space, and let* $F \subset X$ *be a norm closed and convex set. Then* F *is weakly closed.*

Proof. To prove the claim, we show that F^c, the complement of F, is weakly open. Remember that a subbasic family of open sets for the weak topology is given by

$$\{x \in X : \langle x^*, x\rangle < a,\ x^* \in X^*, a \in \mathbb{R}\}.$$

So, let $x \in F^c$. Being F closed and convex, we can strictly separate F from x (Theorem A.1.6): there are $x^* \in X^*$ and $a \in \mathbb{R}$ such that

$$F \subset \{x \in X : \langle x^*, x\rangle > a\} \text{ and } \langle x^*, x\rangle < a.$$

Thus the open set $\{x \in X : \langle x^*, x\rangle < a\}$ contains x and does not intersect F. □

As a consequence of the previous results we can prove, for instance, the following theorem (some simple variant of it can be formulated as well):

Theorem 4.1.3 *Let X be a reflexive Banach space, let $f \in \Gamma(X)$. Suppose $\lim_{\|x\|\to\infty} f(x) = \infty$. Then the problem of minimizing f over X has solutions.*

Proof. As a consequence of the Banach–Alaoglu theorem, reflexivity guarantees that a weakly closed and bounded set is weakly compact. □

Exercise 4.1.4 Let us take a nonempty closed convex set C in a Banach space X, and $x \in X$. The *projection* of x over C is the (possibly empty) set $p_C(x)$ of the points of C which are *nearest* to x:

$$p_C(x) = \{z \in C : \|z - x\| \le \|c - x\|, \forall c \in C\}.$$

Prove that $p_C(x) \neq \emptyset$, provided X is reflexive, and that it is a singleton if X is a Hilbert space. In this case, prove also that $y = P_C(x)$ if and only if $y \in C$ and

$$\langle x - y, c - y\rangle \le 0, \forall c \in C.$$

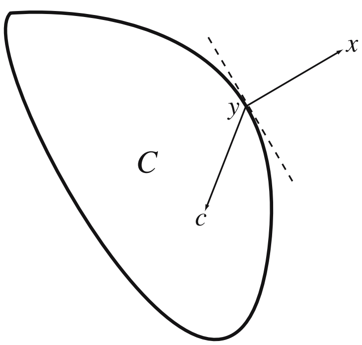


Figure 4.1. The projection y of x on the set C.

The concept of projection allows us to get a formula for the subdifferential of the distance function $d(\cdot\,, C)$, where C is a closed convex subset of a Hilbert space X.

Proposition 4.1.5 *Let X be a Hilbert space, C a nonempty closed convex subset of X, $x \in X$. Then*

$$\partial d(\cdot\,, C)(x) = \begin{cases} 0^* & \text{if } x \in \operatorname{int} C, \\ N_C(x) \cap B^* & \text{if } x \in \partial C, \\ \dfrac{x - P_C(x)}{\|x - P_C(x)\|} & \text{if } x \notin C, \end{cases}$$

where, as usual, $N_C(x)$ is the normal cone at x to C and $P_C(x)$ is the projection of x over C.

Proof. To prove the claim, we appeal to the fact that

$$d(x, C) = (\| \cdot \| \nabla I_C)(x),$$

that the inf-convolution is exact at any point, and to Proposition 3.2.11, which provides a formula for the subdifferential of the inf-convolution at a point where it is exact. Let $x \in \operatorname{int} C$. Setting $u = 0$, $v = x$, we have that $d(x, C) = \|u\| + I_C(v)$, $\partial\|u\| = B_{X^*}$, $\partial I_C(v) = \{0^*\}$, $\partial d(\cdot\,, C)(x) = \partial\|u\| \cap \partial I_C(v) = \{0^*\}$. Now, let us suppose x is in the boundary of C: $x \in \partial C$. Again take $u = 0$, $v = x$. This provides $\partial\|u\| = B_{X^*}$, $\partial I_C(v) = N_C(x)$, and thus $\partial d(\cdot\,, C)(x) = \partial\|u\| \cap \partial I_C(v) = B^* \cap N_C(x)$. Finally, let $x \notin C$. Then $d(x, C) = \|x - P_C(x)\| + I_C(p_C(x))$, $\partial\|x - P_C(x)\| = \frac{x - P_C(x)}{\|x - P_C(x)\|}$, $\partial I_C(P_C(x)) = N_C(P_C(x))$. But $\frac{x - P_C(x)}{\|x - P_C(x)\|} \in N_C(P_C(x))$, as it is seen in the Exercise 4.1.4, and this ends the proof. □

Exercise 4.1.6 Let X be a reflexive Banach space and let $f \colon X \to (-\infty, \infty]$ be a lower semicontinuous, lower bounded function. Let $\varepsilon > 0$, $r > 0$ and $\bar{x} \in X$ be such that $f(\bar{x}) \leq \inf_X f + r\varepsilon$. Then, there exists $\hat{x} \in X$ enjoying the following properties:

(i) $\|\hat{x} - \bar{x}\| \leq r$;
(ii) $f(\hat{x}) \leq f(\bar{x})$;
(iii) $f(\hat{x}) \leq f(x) + \varepsilon\|\hat{x} - x\| \ \ \forall x \in X$.

Hint. The function $g(x) = f(x) + \varepsilon\|\bar{x} - x\|$ has a minimum point $\hat{x}$. Check that $\hat{x}$ fulfills the required properties.

The following section is dedicated to extending the previous result to complete metric spaces.

4.2 The Ekeland variational principle

Due to the lack of a suitable topology to exploit the basic Weierstrass existence theorem, it is quite difficult, except for the reflexive case, to produce general existence results for minimum problems. So it is important to produce results guaranteeing existence at least in "many" cases. The word "many" of course can be given different meanings. The Ekeland variational principle, the fundamental result we describe in this section, allows us to produce a generic existence theorem. But its power goes far beyond this fact; its claim for the existence of a quasi minimum point with particular features has surprisingly many applications, not only in optimization, but also, for instance, in critical point and fixed point theory. Let us start by introducing a useful definition.

Definition 4.2.1 Let (X, d) be a metric space, let $f : X \to \mathbb{R}$ be lower semicontinuous. The *strong slope* of f at x, denoted by $|\nabla f|(x)$ is defined as

$$|\nabla f|(x) = \begin{cases} \limsup_{y \to x} \frac{f(x)-f(y)}{d(x,y)} & \text{if } x \text{ is not a local minimum,} \\ 0 & \text{if } x \text{ is a local minimum.} \end{cases}$$

The next is an estimation from above of the strong slope.

Proposition 4.2.2 *Let X be a metric space, let $f : X \to \mathbb{R}$ be locally Lipschitz at $x \in X$, with Lipschitz constant L. Then $|\nabla f|(x) \le L$.*

For a more regular function f we have:

Proposition 4.2.3 *Let X be a Banach space, let $f : X \to \mathbb{R}$ be Gâteaux differentiable at $x \in X$. Then $|\nabla f|(x) \ge \|\nabla f(x)\|_*$.*

Proof. Let $u \in X$ be such that $\|u\| = 1$ and $\langle \nabla f(x), -u \rangle \ge \|\nabla f(x)\|_* - \varepsilon$, for some small $\varepsilon > 0$. Then

$$\limsup_{y \to x} \frac{f(x) - f(y)}{d(x,y)} \ge \lim_{t \to 0} \frac{f(x) - f(x+tu)}{t} = \langle \nabla f(x), -u \rangle \ge \|\nabla f(x)\|_* - \varepsilon.$$

This allows us to complete the proof. □

Clearly, every function f which is discontinuous at a point x but Gâteaux differentiable at the same point, provides an example when the inequality in the above proposition is strict. But with a bit more regularity we get

Proposition 4.2.4 *Let X be a Banach space, let $f : X \to \mathbb{R}$ be Fréchet differentiable at $x \in X$. Then $|\nabla f|(x) = \|f'(x)\|_*$.*

Proof. Write, for $y \ne x$,

$$f(y) = f(x) + \langle f'(x), y - x \rangle + \varepsilon_y \|y - x\|,$$

where $\varepsilon_y \to 0$ if $y \to x$. Then we get

$$\frac{f(x)-f(y)}{\|y-x\|} = \langle -f'(x), \frac{y-x}{\|y-x\|}\rangle + \varepsilon_y \leq \|f'(x)\|_* + \varepsilon_y.$$

This shows that $|\nabla f|(x) \leq \|f'(x)\|$ and, by means of Proposition 4.2.3, we can conclude. □

Propositions 4.2.3 and 4.2.4 explain the importance of the notion of strong slope (and also the notation used). In particular, for a Fréchet differentiable function, it generalizes the notion of norm of the derivative, to a purely metric setting. Beyond this, it has also interesting connections with nonsmooth differentials of nonconvex functions.

We can now introduce the variational principle.

Theorem 4.2.5 *Let (X, d) be a complete metric space and let $f\colon X \to (-\infty, \infty]$ be a lower semicontinuous, lower bounded function. Let $\varepsilon > 0$, $r > 0$ and $\bar{x} \in X$ be such that $f(\bar{x}) \leq \inf_X f + r\varepsilon$. Then, there exists $\hat{x} \in X$ enjoying the following properties:*

(i) $d(\hat{x}, \bar{x}) \leq r$;
(ii) $f(\hat{x}) \leq f(\bar{x}) - \varepsilon d(\bar{x}, \hat{x})$;
(iii) $f(\hat{x}) < f(x) + \varepsilon d(\hat{x}, x) \;\; \forall x \neq \hat{x}$.

Proof. Let us define the following relation on $X \times X$:

$$x \preceq y \text{ if } f(x) \leq f(y) - \varepsilon d(x, y).$$

It is routine to verify that $\preceq$ is reflexive, antisymmetric and transitive. Moreover, lower semicontinuity of f guarantees that $\forall x_0 \in X$, the set $A := \{x \in X : x \preceq x_0\}$ is a closed set. Let us now define

$$x_1 = \bar{x}, \quad S_1 = \{x \in X : x \preceq x_1\},$$
$$x_2 \in S_1 \text{ such that } f(x_2) \leq \inf_{S_1} f + \frac{r\varepsilon}{4};$$

and recursively

$$S_n = \{x \in X : x \preceq x_n\},$$
$$x_{n+1} \in S_n \text{ such that } f(x_{n+1}) \leq \inf_{S_n} f + \frac{r\varepsilon}{2(n+1)}.$$

For all $n \geq 1$, S_n is a nonempty closed set, and $S_n \supset S_{n+1}$. Let us now evaluate the size of the sets S_n. Let $x \in S_n$, for $n > 1$. Then $x \preceq x_n$ and $x \in S_{n-1}$, hence

$$f(x) \leq f(x_n) - \varepsilon d(x, x_n),$$
$$f(x_n) \leq f(x) + \frac{r\varepsilon}{2n},$$

giving

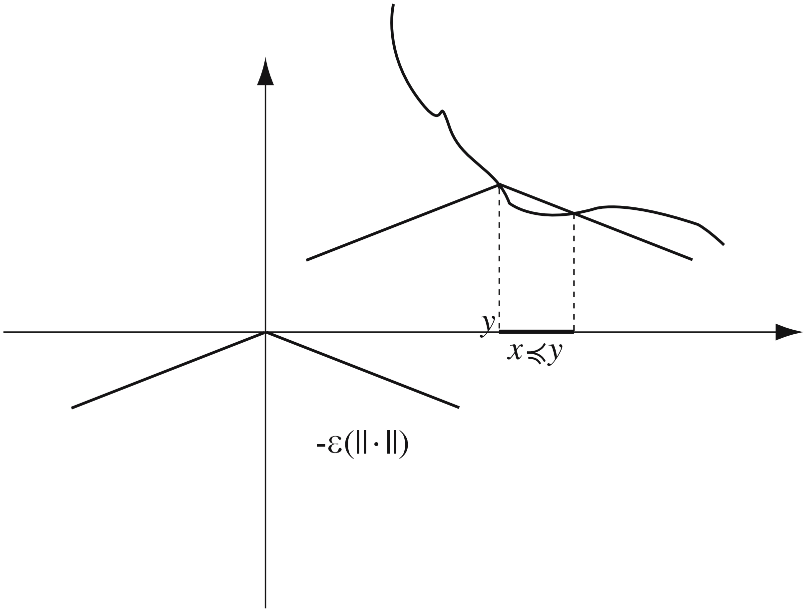


Figure 4.2. The $\preceq$ relation.

$$d(x, x_n) \leq \frac{r}{2n}.$$

In the same way it can be shown that if $x \in S_1$, then $d(x, x_1) = d(x, \bar{x}) \leq r$. Since X is a complete metric space and the sequence of the diameters of the sets S_n goes to zero, it follows that $\bigcap_{n\geq 1} S_n$ is a singleton (see Exercise 4.2.6). Let $\bigcap_{n\geq 1} S_n := \{\hat{x}\}$. Now, it is a pleasure to show that $\hat{x}$ has the required properties. The first and the second one immediately follow from the fact that $\hat{x} \in S_1$, while, to verify the third one, if we suppose the existence of $x \in X$ such that $f(\hat{x}) \geq f(x) + \varepsilon d(x, \hat{x})$, then $x \preceq \hat{x} \preceq x_n, \forall n$, implying $x \in \bigcap_{n\geq 1} S_n$ and so $x = \hat{x}$. □

Exercise 4.2.6 Let (X, d) be a complete metric space, let $\{S_n\}$ be a sequence of nested closed sets such that diam $S_n \to 0$. Prove that $\bigcap S_n$ is a singleton.

Hint. Take $x_n \in S_n$ for all n. Then $\{x_n\}$ is a Cauchy sequence. Thus $\bigcap S_n$ is nonempty. Moreover, it cannot contain more than one point, as diam $S_n \to 0$.

The third condition of the Ekeland principle has many interesting, and sometimes rather surprising, consequences. At first, it shows that the *approximate* solution $\hat{x}$ of the problem of minimizing f is, at the same time, also the *unique exact* solution of a minimum problem, close to the original one, in a sense we shall specify in Chapter 11. Moreover, this approximate solution enjoys an important property with respect to the strong slope, as we now see.

Corollary 4.2.7 *Let X be a complete metric space. Let $f\colon X \to (-\infty, \infty]$ be lower semicontinuous and lower bounded. Let $\varepsilon, r > 0$ and $\bar{x} \in X$ be such that $f(\bar{x}) < \inf_X f + \varepsilon r$. Then there exists $\hat{x} \in X$ with the following properties:*

(i) $d(\hat{x}, \bar{x}) < r$;
(ii) $f(\hat{x}) \leq f(\bar{x})$;
(iii) $|\nabla f|(\hat{x}) < \varepsilon$.

Proof. It is enough to apply the principle, with suitable $0 < \varepsilon_0 < \varepsilon$, $0 < r_0 < r$. The last condition implies $|\nabla f|(\hat{x}) \leq \varepsilon_0$, as is easy to see. □

From the previous results we deduce:

Corollary 4.2.8 *Let X be a Banach space, let $f\colon X \to \mathbb{R}$ be lower semicontinuous, lower bounded and Gâteaux differentiable. Given $\varepsilon, r > 0$ and $\bar{x} \in X$ such that $f(\bar{x}) < \inf_X f + \varepsilon r$, there exists $\hat{x} \in X$ with the following properties:*

(i) $d(\hat{x}, \bar{x}) < r$;
(ii) $f(\hat{x}) \leq f(\bar{x})$;
(iii) $\|\nabla f(\hat{x})\|_* < \varepsilon$.

Proof. From Proposition 4.2.3 and Corollary 4.2.7. □

Corollary 4.2.9 *Let X be a Banach space, let $f\colon X \to \mathbb{R}$ be lower semicontinuous, lower bounded and Gâteaux differentiable. Then there exists a sequence $\{x_n\} \subset X$ such that*

(i) $f(x_n) \to \inf f$;
(ii) $\nabla f(x_n) \to 0^*$.

Sequences $\{x_n\}$ such that $\nabla f(x_n) \to 0^*$ are known in the literature as *Palais–Smale sequences*, and at level a if it happens that $f(x_n) \to a$. A function f is said to satisfy the Palais–Smale condition (at level a) if every Palais–Smale sequence with bounded values (at level a) has a limit point. This is a compactness assumption crucial in every abstract existence theorem in critical point theory. And the notion of strong slope is the starting point for a purely metric critical point theory. The above corollary claims the existence of Palais–Smale sequences at level $\inf f$.

The Ekeland principle has interesting consequences for convex functions too.

Theorem 4.2.10 *Let X be a Banach space, let $f \in \Gamma(X)$. Let $x \in \operatorname{dom} f$, $\varepsilon, r, \sigma > 0$, $x^* \in \partial_{\varepsilon r} f(x)$. Then there are $\hat{x} \in \operatorname{dom} f$ and $\hat{x}^* \in X^*$, such that*

(i) $\hat{x}^* \in \partial f(\hat{x})$;
(ii) $\|x - \hat{x}\| \leq \frac{r}{\sigma}$;
(iii) $\|\hat{x}^* - x^*\|_* \leq \varepsilon\sigma$;
(iv) $|f(x) - f(\hat{x})| \leq r(\varepsilon + \frac{\|x^*\|_*}{\sigma})$.

Proof. As $x^* \in \partial_{\varepsilon r} f(x)$, it holds, $\forall y \in X$,

$$f(y) \geq f(x) + \langle x^*, y - x \rangle - \varepsilon r.$$

Setting $g(y) = f(y) - \langle x^*, y \rangle$, we get

$$g(x) \leq \inf_X g + (\varepsilon\sigma)\frac{r}{\sigma}.$$

Applying the principle to the function g (and replacing r by $\frac{r}{\sigma}$, ε by $\sigma\varepsilon$), we have then the existence of an element $\hat{x} \in \operatorname{dom} f$ satisfying condition (ii). Let us find the right element in its subdifferential. Condition (iii) of the principle says that $\hat{x}$ minimizes the function $g(\cdot) + \varepsilon\sigma\|\cdot - \hat{x}\|$, so that

$$0^* \in \partial(g(\cdot) + \varepsilon\sigma\|\cdot - \hat{x}\|)(\hat{x}).$$

We can use the sum Theorem 3.4.2. We then get

$$0^* \in \partial g(\hat{x}) + \varepsilon\sigma B_{X^*} = \partial f(\hat{x}) - x^* + \varepsilon\sigma B_{X^*}.$$

This is equivalent to saying that there exists an element $\hat{x}^* \in \partial f(\hat{x})$ such that $\|\hat{x}^* - x^*\|_* \leq \varepsilon\sigma$. Finally, condition (iv) routinely follows from (ii), (iii) and from $x^* \in \partial_{\varepsilon r} f(x)$, $\hat{x}^* \in \partial f(\hat{x})$. □

The introduction of a constant σ in the above result is not made with the intention of creating more entropy. For instance, the choice of $\sigma = \max\{\|x^*\|_*, 1\}$ allows controlling the variation of the function f, at the expense, of course, of controlling of the norm of $\hat{x}^*$. Thus the following useful result can be easily proved.

Corollary 4.2.11 *Let X be a Banach space, let $f \in \Gamma(X)$. Let $x \in \operatorname{dom} f$. Then there is a sequence $\{x_n\} \subset \operatorname{dom} \partial f$ such that*

$$x_n \to x \text{ and } f(x_n) \to f(x).$$

Proof. This follows from (ii) and (iv) of Theorem 4.2.10, with the above choice of σ, $\varepsilon = 1$, and $r = \frac{1}{n}$. □

Corollary 4.2.12 *Let X be a Banach space, let $f \in \Gamma(X)$ be lower bounded, let $\varepsilon, r > 0$ and $\bar{x} \in \operatorname{dom} f$ be such that $f(\bar{x}) < \inf f + \varepsilon r$. Then there exist $\hat{x} \in \operatorname{dom} f$ and $\hat{x}^* \in \partial f(\hat{x})$, such that*

(i) $\|\bar{x} - \hat{x}\| < r$;
(ii) $\|\hat{x}^*\|_* < \varepsilon$.

Proof. We apply Theorem 4.2.10 to the point $x = \bar{x}$, and with $\sigma = 1$. Observe that $0^* \in \partial_{\varepsilon_0 r_0} f(\bar{x})$, with suitable $\varepsilon_0 < \varepsilon$ and $r_0 < r$. □

Another very interesting consequence of the previous theorem is the following fact.

Corollary 4.2.13 *Let X be a Banach space and $f \in \Gamma(X)$. Then there exists a dense subset D of* $\operatorname{dom} f$ *such that $\partial f(x) \neq \emptyset$ for all $x \in D$.*

Proof. Fix any $r > 0$ and $x \in \operatorname{dom} f$. Find x^* in $\partial_{r/2} f(x)$. Apply Theorem 4.2.10 to x, x^*, with the choice of $\varepsilon = 1/2, \sigma = 1$. We get $\hat{x}$ such that $\partial f(\hat{x}) \neq \emptyset$ and such that $\|x - \hat{x}\| < r$, and this finishes the proof. □

The following proposition, beyond being interesting in itself, is useful in proving that the subdifferential of a function in $\Gamma(X)$ is a maximal monotone operator. Remember that in Theorem 3.5.14 we have already shown this result for a narrower class of functions. To prove it, we follow an idea of S. Simmons (see [Si]).

Proposition 4.2.14 *Let X be a Banach space, let $f \in \Gamma(X)$, and suppose $f(0) > \inf f$. Then there are $z \in \operatorname{dom} f$, $z^* \in \partial f(z)$ with the following properties:*

(i) $f(z) < f(0)$;
(ii) $\langle z^*, z \rangle < 0$.

Proof. Observe at first that (i) is an immediate consequence of (ii) and of the definition of subdifferential. So, let us establish the second property. Let $f(0) > a > \inf f$, and set

$$2k := \sup_{x \neq 0} \frac{a - f(x)}{\|x\|}.$$

It is obvious that $k > 0$. We shall prove later that $k < \infty$. By definition of k,

$$f(x) + 2k\|x\| \geq a, \ \forall x \in X.$$

Moreover, there exists $\bar{x}$ such that

$$k < \frac{a - f(\bar{x})}{\|\bar{x}\|},$$

providing

$$f(\bar{x}) + 2k\|\bar{x}\| < a + k\|\bar{x}\| \leq \inf\{f(x) + 2k\|x\| : x \in X\} + k\|\bar{x}\|.$$

We can then apply Corollary 4.2.12 with $\varepsilon = k$ ed $r = \|\bar{x}\|$. Hence there are $z \in \operatorname{dom} f$ and $w^* \in \partial(f(\cdot) + k\|\cdot\|)(z)$ such that

$$\|z - \bar{x}\| < \|\bar{x}\| \text{ and } \|w^*\| < k.$$

The first condition implies $z \neq 0$. By the sum Theorem 3.4.2 we also have

$$w^* = z^* + y^*,$$

with

$$z^* \in \partial f(z) \text{ and } y^* \in \partial(k\|\cdot\|)(z).$$

The last condition, by applying the definition of subdifferential, implies

$$0 \geq k\|z\| - \langle y^*, z\rangle,$$

whence

$$\langle y^*, z\rangle \geq k\|z\|.$$

We then get

$$\langle z^*, z\rangle = \langle w^*, z\rangle - \langle y^*, z\rangle < k\|z\| - k\|z\| \leq 0.$$

To conclude, we must verify that $k < \infty$. It is enough to consider the case when $f(x) < a$. Let $x^* \in X^*, \alpha \in \mathbb{R}$ be such that $f(y) \geq \langle x^*, y\rangle - \alpha$, $\forall y \in X$. The existence of such an affine function minorizing f relies on the fact that $f \in \Gamma(X)$ (Corollary 2.2.17). We then have

$$a - f(x) \leq |a| + |\alpha| + \|x^*\|_*\|x\|,$$

whence

$$\frac{a - f(x)}{\|x\|} \leq \frac{|a| + |\alpha|}{d(0, f^a)} + \|x^*\|_*,$$

and this ends the proof. □

Exercise 4.2.15 Prove the following generalization of Theorem 4.4.1. Let $f \in \Gamma(X)$. Then ∂f is a maximal monotone operator.

Hint. Use the proof of Theorem 4.1.1 and the previous proposition.

To conclude this section, we want to get a result on the characterization of the epigraph of $f \in \Gamma(X)$, which improves upon Theorem 2.2.21. There, it was proved that the epigraph can be characterized as the intersection of the epigraphs of all the affine functions minorizing f. Here we prove that we can just consider *very particular* affine functions minorizing f, in order to have the same characterization.

To prove our result, we first must show the following lemma.

Lemma 4.2.16 *Let C be a closed convex set, and $x \notin C$. Then, for every $k > 0$, there exist $c \in C$, $c^* \in \partial I_C(c)$ such that*

$$\langle c^*, x - c\rangle \geq k.$$

Proof. Let $d = d(x, C)$, let $\alpha > k + d + 2$ and let $\bar{x} \in C$ be such that $\|\bar{x} - x\| < d(1 + \frac{1}{\alpha})$. Let

$$S = \{(tx + (1-t)\bar{x}, t\alpha + (1-t)(-1)) : 0 \leq t \leq 1\}.$$

Then $S \cap \operatorname{epi} I_C = \emptyset$ and they can be strictly separated. Thus there exist $x^* \in X^*, r^* \in \mathbb{R}$ and $h \in \mathbb{R}$ such that $(x^*, r^*) \neq (0^*, 0)$ and

$$\langle (x^*, r^*), (c, r) \rangle \geq h > \langle (x^*, r^*), (u, \beta) \rangle,$$

for all $c \in C$, $r \geq 0$, $(u, \beta) \in S$. Taking any $c \in C$ and $r > 0$ big enough in the above inequalities shows that $r^* \geq 0$. And taking $c = \bar{x} = u$ shows that actually $r^* > 0$. Setting $y^* = -\frac{x^*}{r^*}$, and putting at first $(u, \beta) = (\bar{x}, -1)$ and then $(u, \beta) = (x, \alpha)$ in the above inequalities, we finally get

$$y^* \in \partial_1 I_C(\bar{x}) \text{ and } \langle y^*, x - c \rangle > \alpha, \quad \forall c \in C.$$

Thanks to Theorem 4.2.10 $(\varepsilon, r, \sigma = 1)$, we have the existence of $c \in C$, $c^* \in \partial I_C(c)$ such that

$$\|c - \bar{x}\| \leq 1 \text{ and } \|c^* - y^*\|_* \leq 1.$$

Thus

$$\begin{aligned} \langle c^*, x - c \rangle &= \langle c^* - y^*, x - c \rangle + \langle y^*, x - c \rangle > \alpha - (\|x - \bar{x}\| + \|\bar{x} - c\|) \\ &\geq \alpha - (d(1 + \frac{1}{\alpha}) + 1) > k. \end{aligned}$$

□

Theorem 4.2.17 *Let $f \in \Gamma(X)$. Then, for all $x \in X$,*

$$f(x) = \sup\{f(y) + \langle y^*, x - y \rangle : (y, y^*) \in \partial f\}.$$

Proof. Observe at first that from the previous lemma the conclusion easily follows for the indicator function of a given closed convex set. Next, let us divide the proof into two parts. At first we prove the claim for $\bar{x} \in \operatorname{dom} f$, then for $\bar{x}$ such that $f(\bar{x}) = \infty$, which looks a bit more complicated. Thus, given $\bar{x} \in \operatorname{dom} f$ and $\eta > 0$, we need to find $(y, y^*) \in \partial f$ such that

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

Fix ε such that $2\varepsilon^2 < \eta$ and separate $\operatorname{epi} f$ from $(\bar{x}, f(\bar{x}) - \varepsilon^2)$. We then find $x^* \in \partial_{\varepsilon^2} f(\bar{x})$ (using the standard separation argument seen for the first time in Lemma 2.2.16). From Theorem 4.2.10 we have the existence of y, $y^* \in \partial f(y)$, such that

$$\|x^* - y^*\| \leq \varepsilon \text{ and } \|\bar{x} - y\| \leq \varepsilon.$$

Thus

$$f(y) + \langle y^*, \bar{x} - y \rangle \geq f(\bar{x}) + \langle x^* - y^*, y - \bar{x} \rangle - \varepsilon^2 \geq f(\bar{x}) - \eta.$$

This shows the first part of the claim. Suppose now $f(\bar{x}) = \infty$, and fix $k > 0$. We need to find $(y, y^*) \in \partial f$ such that

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

We shall apply Lemma 4.2.16 to $C = \operatorname{epi} f$ and to $x = (\bar{x}, k)$. We then see that there exist $(x, r) \in \operatorname{epi} f$, $(x^*, r^*) \in \partial I_{\operatorname{epi} f}(x, r)$ such that

$$\langle (x^*, r^*), (\bar{x}, k) - (x, r) \rangle \geq 2. \tag{4.1}$$

Moreover, the condition $(x^*, r^*) \in \partial I_{\operatorname{epi} f}(x, r)$ amounts to saying that

$$\langle (x^*, r^*), (y, \beta) - (x, r) \rangle \leq 0, \tag{4.2}$$

for all $(y, \beta) \in \operatorname{epi} f$. From (4.2) it is easy to see that $r^* \leq 0$ and, with the choice of $(y, \beta) = (x, f(x))$, we see that $r = f(x)$. Suppose now $r^* < 0$. Then we can suppose, without loss of generality, that $r^* = -1$. Thus $(x^*, -1)$ supports $\operatorname{epi} f$ at $(x, f(x))$ and this means that $x^* \in \partial f(x)$. Moreover, from (4.1) we get

$$\langle x^*, \bar{x} - x \rangle + (-1)(k - f(x)) \geq 2,$$

i.e.,

$$f(x) + \langle x^*, \bar{x} - x \rangle \geq k + 2 > k,$$

so that we have shown the claim in the case $r^* < 0$. It remains to see the annoying case when $r^* = 0$. In such a case (4.1) and (4.2) become

$$\langle x^*, \bar{x} - x \rangle \geq 2, \langle x^*, y - x \rangle \leq 0, \forall y \in \operatorname{dom} f. \tag{4.3}$$

Set $d = \|x - \bar{x}\|$ and $a = \frac{1}{\|x^*\|_*}$. Let $y^* \in \partial_a f(x)$, and observe that from (4.3) we have that for all $t > 0$, $z_t^* := y^* + tx^* \in \partial_a f(x)$. From Theorem 4.2.10 there exist y_t, $y_t^* \in \partial f(y_t)$ such that

$$\|x - y_t\| \leq a, \text{ and } \|z_t^* - y_t^*\|_* \leq 1.$$

As $\{y_t : t > 0\}$ is a bounded set, there exists b such that $f(y_t) \geq b$ for all $t > 0$. We then get

$$\begin{aligned} f(y_t) + \langle y_t^*, \bar{x} - y_t \rangle &= f(y_t) + \langle y_t^* - z_t^*, \bar{x} - y_t \rangle + \langle z_t^*, \bar{x} - y_t \rangle \\ &\geq b - (d + a) - \|y^*\|(d + a) + t(\langle x^*, \bar{x} - x \rangle + \langle x^*, x - y_t \rangle) \\ &\geq b - (d + a) - \|y^*\|(d + a) + t. \end{aligned}$$

Then we can choose t big enough to make the following inequality be true:

$$b - (1 + \|y^*\|)(d + a) + t \geq k,$$

and this ends the proof. □

We conclude by improving the result of the Lemma 3.6.4, once again with a beautiful argument following from the Ekeland variational principle.

Lemma 4.2.18 *Let $f \colon X \to (-\infty, \infty]$ be convex. Let $\delta, a > 0$, $g \colon B(0, a) \to \mathbb{R}$ a Gâteaux function and suppose $|f(x) - g(x)| \leq \delta$ for $x \in B(0; a)$. Let*

$0 < r < R \leq a$, let x be such that $\|x\| \leq r$ and $x^* \in \partial f(x)$. Then both the following estimates hold:

$$d(x^*, \nabla g(B(x; R-r))) \leq \frac{2\delta}{R-r},$$
$$d(x^*, \nabla g(RB) \leq \frac{2\delta}{R-r}.$$

The same holds if g is convex and real valued, provided we replace ∇g with ∂g.

Proof. Without loss of generality we can suppose $f(x) = 0$ and $x^* = 0^*$. Then $g(x) < \delta$ and, if $\|u\| \leq R$, $g(u) > -\delta$ (since f is nonnegative). It follows that

$$g(x) < \inf g + \frac{2\delta}{R-r}(R-r),$$

on the ball rB. To conclude, it is enough to use Corollary 4.2.8 (or Corollary 4.2.12 for the convex case). □

4.3 Minimizing a convex function

In this section we want to analyze some properties of the level sets of a convex function, and to give a flavor of how one can proceed in looking for a minimum of a convex function defined on a Euclidean space. We do not go into the details of this topic; the interested reader is directed to excellent books treating this important problem in a systematic way, such as the one by Hiriart-Urruty–Lemaréchal [HUL]. We start by considering the level sets.

4.3.1 Level sets

We begin by establishing a result which actually could be derived by subsequent, more general statements, but which we prefer to present here, and to prove it with an elementary argument.

Proposition 4.3.1 *Let $f\colon \mathbb{R}^n \to (-\infty, \infty]$ be a convex, lower semicontinuous function. Suppose* $\operatorname{Min} f$ *is nonempty and compact. Then f^a is bounded for all $a > \inf f$ and $\forall \varepsilon > 0$ there exists $a > \inf f$ such that $f^a \subset B_\varepsilon[\operatorname{Min} f]$. Moreover, if $\{x_n\}$ is such that $f(x_n) \to \inf f$, then $\{x_n\}$ has a limit point which minimizes f. And if* $\operatorname{Min} f$ *is a singleton x, then $x_n \to x$.*

Proof. Let $r > 0$ be such that $\operatorname{Min} f \subset (r-1)B$ and, without loss of generality, suppose $0 \in \operatorname{Min} f$ and $f(0) = 0$. By contradiction, suppose there are $a > \inf f$ and $\{x_n\}$ such that $f(x_n) \leq a$ and $\|x_n\| \to \infty$. It is an easy matter to verify that the sequence $\{\frac{rx_n}{\|x_n\|}\}$ is such that $f(\frac{rx_n}{\|x_n\|}) \to 0$, as a consequence of convexity of f. Then $\{\frac{rx_n}{\|x_n\|}\}$ has a subsequence converging to a point $\bar{x}$ of norm

r, and $\bar{x}$ minimizes f, by lower semicontinuity of f. But this is impossible. Now suppose there is $\varepsilon > 0$ such that for all n there is x_n such that $f(x_n) \le \inf f + \frac{1}{n}$ and $d(x_n, \operatorname{Min} f) > \varepsilon$ for all n. Then $\{x_n\}$ is bounded, thus it has a cluster point which minimizes f, against the fact that $d(x_n, \operatorname{Min} f) > \varepsilon)$ for all n. To conclude, we must show that if $\operatorname{Min} f$ is a singleton, say x and $f(x_n) \to \inf f$, then $\{x_n\}$ converges to x. This is a purely topological argument. Suppose not; then there are $a > 0$ and a subsequence $\{y_n\}$ of $\{x_n\}$ such that $\|y_n - x\| \ge a$ for all n. As $\{y_n\}$ is bounded, it has a limit point which minimizes f, so that this limit point must be x, against the assumption $\|y_n - x\| \ge a$ for all n. □

The first result we present shows that the level sets of a convex lower semicontinuous function "cannot be too different". Next, we inquire about the connections between the local shape of the boundary of a level set, at a point x, the descent directions at the point x, and the subdifferential of f at x. For the first result, recall the definition of recession cone given in Definition 1.1.15.

Proposition 4.3.2 *Let $f \in \Gamma(X)$ and suppose $f^a, f^b \neq \emptyset$. Then $0^+(f^a) = 0^+(f^b)$.*

Proof. Let $z \in f^a$, $x \in 0^+(f^a)$ and fix $y \in f^b$. We must show that $f(x+y) \le b$. As $(1-\frac{1}{n})y + \frac{1}{n}(z+nx) \to y + x$, we have

$$f(y+x) \le \liminf f\Big(\Big(1-\frac{1}{n}\Big)y + \frac{1}{n}(z+nx)\Big) \le \liminf\Big(\Big(1-\frac{1}{n}\Big)b + \frac{1}{n}a\Big) = b,$$

and this ends the proof. □

Remark 4.3.3 Consider a separable Hilbert space with basis $\{e_n : n \in \mathbb{N}\}$, and the function

$$f(x) = \sum_{n=1}^{\infty} \frac{\langle x, e_n \rangle^2}{n^4}.$$

From the previous proposition (but it is easily seen directly, too), $0^+(f^a) = \{0\}\ \forall a > 0$, as $0^+(f^0) = \{0\}$. However f^a is unbounded for all $a > 0$, and this shows that Proposition 4.3.1 and Proposition 1.1.16 fail in infinite dimensions.

Proposition 4.3.4 *Let $f \colon X \to (-\infty, \infty]$ be convex and lower semicontinuous. Suppose there is $b > \inf f$ such that f^b is bounded. Then f^a is bounded for all $a > \inf f$.*

Proof. In the finite dimensional case the result is an immediate consequence of Proposition 4.3.2, since $0^+(f^a) = 0^+(f^b) = \{0\}$ and this is equivalent to saying that f^a is bounded (moreover, the condition $b > \inf f$ can be weakened to $f^b \neq \emptyset$). In the general case, let $a > b$, let r be such that $f^b \subset (r-1)B$ and take a point $\bar{x}$ such that $f(\bar{x}) < b$. With the usual translation of the axes we can suppose, without loss of generality, $\bar{x} = 0$, $f(0) = 0$ and consequently

$b > 0$. This clearly does not affect boundedness of the level sets. Let y be such that $\|y\| = \frac{r(a+1)}{b}$. Then $z = \frac{b}{a+1}y$ has norm r. It follows that

$$b < f(z) \leq \frac{b}{a+1} f(y),$$

whence $f(y) \geq a + 1$. This shows that $f^a \subset \frac{r(a+1)}{b}B$. □

The next proposition is quite simple.

Proposition 4.3.5 *Let $f\colon X \to (-\infty, \infty]$ be convex, lower semicontinuous. Let $b > \inf f$ be such that f^b is bounded. Then, for every $r > 0$ there exists $c > b$ such that $f^c \subset B_r[f^b]$.*

Proof. Without loss of generality, suppose $f(0) = 0$. Let $k > 0$ be such that $f^{b+1} \subset kB$, and $k > r(b+1)$. The choice of $c = b + \frac{rb}{k-r}$ works since, if $x \in f^c$, then $\frac{b}{c}x \in f^b$. Moreover,

$$\|x - \frac{b}{c}x\| \leq k\frac{c-b}{c} = r.$$

□

Exercise 4.3.6 Let $f\colon X \to \mathbb{R}$ be convex and continuous, where X is a Euclidean space. Let C be a closed convex subset of X. Let $a \in \mathbb{R}$ be such that $f^a \neq \emptyset$ and suppose $0^+(C) \cap 0^+(f^a) = \{0\}$. Then $f(C)$ is closed.

Hint. Suppose $\{y_n\} \subset f(C)$ and $y_n \to y$. Let $c_n \in C$ be such that $y_n = f(c_n)$. Show that $\{c_n\}$ must be bounded.

Exercise 4.3.7 Let $f \in \Gamma(X)$, X a Banach space. Suppose $a > \inf f$. Then $f^a = \text{cl}\{x : f(x) < a\}$.

Hint. Let x be such that $f(x) = a$ and let z be such that $f(z) < a$. Look at f on the segment $[x, z]$.

We now see that, given a point x, the directions y such that $f'(x; y) < 0$ are those for which the vector goes "into" the level set relative to x.

Proposition 4.3.8 *Let $f\colon X \to (-\infty, \infty]$ be convex and lower semicontinuous. Let x be a point where f is (finite and) continuous. Then*

$$\{y : f'(x; y) < 0\} = \{y : \exists\lambda > 0, z,\ f(z) < f(x) \text{ and } y = \lambda(z - x)\}.$$

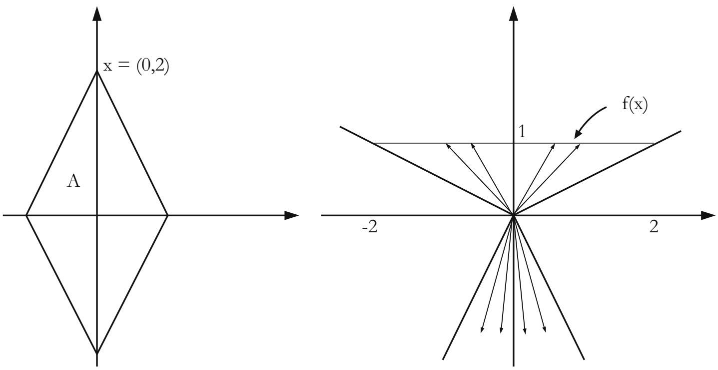
Proof. Let $A = \{y : f'(x; y) < 0\}$ and let $B = \{y : \exists\lambda > 0, z,\ f(z) < f(x) \text{ and } y = \lambda(z - x)\}$. Observe that both A and B are cones. Now, let $y \in B$. Then there are $\lambda > 0$ and z such that $y = \lambda(z - x)$ and $f(z) < f(x)$. Since A is a cone we can suppose, without loss of generality, $\lambda < 1$. We have that $f(\lambda z + (1-\lambda)x) < f(x)$ for all λ. Thus $f(x+y) - f(x) < 0$, which implies $f'(x; y) < 0$ so that $y \in A$. Now, let $y \in A$. Then $f(x + ty) - f(x) < 0$ for small $t > 0$. The conclusion follows. □

We now want to say something on the following topic. As is well known, if a function f is smooth, and one considers a point x where ∇f does not vanish, then $\nabla f(x)$ is perpendicular to the tangent plane to the level set at height $f(x)$. In the convex case, this means that the gradient is a vector in the normal cone at x to the level set at height $f(x)$. Moreover the direction of $\nabla f(x)$ is a descent direction. At least for small $t > 0$ we have $f(x - t\nabla f(x)) < f(x)$. But what happens in the nonsmooth case? The following example shows that things can be different.

Example 4.3.9 This is an example showing that in the nonsmooth case a direction opposite to one subgradient at a point of a given function is not necessarily a descent direction for the function itself, not even locally. Let

$$f(x, y) = 2|x| + |y|,$$

let $p = (0, 2)$, and let the direction v be $v = (1, 1)$. It is straightforward to see that $v \in \partial f(p)$ and that for no $t > 0$ does $p - tv$ belong to the level set relative to p.



$A = f^2$ Descent directions

Figure 4.3.

Also in the nonsmooth case, however, it is true that, if $x^* \in \partial f(x)$, then x^* is in the normal cone at x to the level set at height $f(x)$, as is easy to see. But actually it is possible to provide much more precise information, and this is what we are going to do now.

The result of the next exercise will be used in the proposition following it.

Exercise 4.3.10 Let X be a Banach space, $x \in X$, $0^* \neq x^*$. Set $H = \{z : \langle x^*, z\rangle \geq \langle x^*, x\rangle\}$. Prove that

$$N_H(x) = \mathbb{R}_-\{x^*\}.$$

Hint. Let $z^* \in N_H(x)$. Then $\langle z^*, u\rangle = \langle z^*, x+u-x\rangle \le 0$, for all u such that $\langle x^*, u\rangle = 0$. It follows that $\langle z^*, u\rangle = 0$, for all u such that $\langle x^*, u\rangle = 0$. Derive the conclusion.

Theorem 4.3.11 *Let X be a Banach space, let $f\colon X \to (-\infty, \infty]$ be convex and lower semicontinuous. Let x be a point where f is (finite and) continuous and suppose $f(x) = a > \inf f$. Then*

$$N_{f^a}(x) = \text{cone}\{\partial f(x)\}.$$

Proof. The fact that $N_{f^a}(x)$ contains the cone generated by the subdifferential of f on x is easy to see and is true also if $f(x) = \inf f$. To see the opposite inclusion, let $0^* \neq x^* \in N_{f^a}(x)$. Since $\langle x^*, z-x\rangle \le 0$ for all $z \in f^a$, it follows that $\langle x^*, z-x\rangle < 0$ for all $z \in \text{int}\, f^a$. Otherwise, for some $z \in \text{int}\, f^a$ we would have $\langle x^*, z-x\rangle = 0$. This would imply that x^* has a local maximum at z, but in this case it would be $x^* = 0^*$. From this we have that $f(z) < f(x)$ implies $\langle x^*, z\rangle < \langle x^*, x\rangle$ and this in turn implies that if $\langle x^*, z\rangle \ge \langle x^*, x\rangle$, then $f(z) \ge f(x)$. In other words, f has a minimum on x over the set $H = \{z : \langle x^*, z\rangle \ge \langle x^*, x\rangle\}$. It follows, by using the sum theorem (since f is continuous at x) that

$$0^* \in \partial(f + I_H)(x) = \partial f(x) + N_H(x).$$

Now, as suggested by Exercise 4.3.10, $N_H(x) = \mathbb{R}_-\{x^*\}$. Thus there are $t \ge 0$ and $z^* \in \partial f(x)$ such that $x^* = tz^*$, and this ends the proof. □

If X is finite dimensional, it is enough to assume that $\partial f(x) \neq \emptyset$, but in this case one must take the closure of the cone generated by $\partial f(x)$ (see [Ro, Theorem 23.7]).

4.3.2 Algorithms

Usually, even if we know that the set of the minima of a (convex) function is nonempty, it is not easy or even possible to directly find a minimum point (for instance by solving the problem $0^* \in \partial f(x)$.) For this reason, several algorithms were developed in order to build up sequences of points approximating a solution (in some sense). In this section we shall consider some of these procedures. We are then given a convex function $f\colon \mathbb{R}^n \to \mathbb{R}$ with a nonempty set of minimizers, and we try to construct sequences $\{x_k\}$ approximating $\text{Min}\, f$. The sequences $\{x_k\}$ will be built up in the following fashion:

$$x_0 \text{ arbitrary}, \quad x_{k+1} = x_k - \lambda_k d_k.$$

The vector d_k is assumed to be of norm one, so that λ_k is the *length* of the step at time k. Of course, both the choices of λ_k and d_k are crucial for good behavior of the algorithm. As far as λ_k is concerned, it is clear that it must not be too small, as in such a case the sequence $\{x_k\}$ could converge to something not minimizing the function. And if it converges to a solution, its convergence could be much too slow. On the other hand, it should not be too big, as in this case the algorithm need not converge. On the other side, $-d_k$ represents the *direction* along which we build up the element x_{k+1}, starting from x_k. Usually, it is a vector d_k such that $-d_k$ has the same direction as a vector $v_k \in \partial f(x_k)$. In the smooth case, this choice guarantees that the function decreases at each step, at least if λ is sufficiently small. In the nonsmooth case, we have seen in Example 4.3.9 that this does not always happen.

Theorem 4.3.12 *Let* $\{\lambda_k\}$ *be such that*

$$\lambda_k \mapsto 0, \tag{4.4}$$

$$\sum_{k=0}^{\infty} \lambda_k = \infty. \tag{4.5}$$

Let

$$v_k \in \partial f(x_k),$$

and let

$$d_k = \begin{cases} \frac{v_k}{\|v_k\|} & \text{if } v_k \neq 0, \\ 0 & \text{if } v_k = 0. \end{cases}$$

Moreover, suppose Min f *is a nonempty bounded set. Then*

$$\lim_{k \to +\infty} d(x_k, \operatorname{Min} f) = 0 \quad \text{and} \quad \lim_{k \to +\infty} f(x_k) = \inf f.$$

Proof. First, observe that if for some k it is $d_k = 0$, then we have reached a minimum point. In this case the sequence could possibly become constant, but it is not necessary to assume this. The result holds also in the case the algorithm does not stop. Simply observe that if $d_k = 0$, then $x_{k+1} = x_k$. Thus, we can assume, without loss of generality, that $d_k \neq 0$ for all k. Moreover, observe that the equality $\lim_{k \to +\infty} f(x_k) = \inf f$ is an easy consequence of the first part of the claim.

Now, suppose there are $a > 0$ and k such that

$$d(x_k, \operatorname{Min} f) \geq a > 0. \tag{4.6}$$

This implies, in view of Proposition 4.3.1, that there exists $c > 0$ such that $f(x_k) \geq \inf f + c$. Since, for all x,

$$f(x) \geq f(x_k) + \langle v_k, x - x_k \rangle,$$

we have that

$$\langle v_k, x - x_k \rangle \leq 0, \forall x \in f^{\inf f + c}.$$

Since f is continuous and Min f is compact, there exists $r > 0$ such that $B_r[\operatorname{Min} f] \subset f^{\inf f + c}$. Take $\bar{x} \in \operatorname{Min} f$ and consider the point $\bar{x} + r d_k \in B_r[\operatorname{Min} f]$. Then

$$\langle v_k, \bar{x} + rd_k - x_k \rangle \leq 0,$$

and also

$$\langle d_k, \bar{x} + rd_k - x_k \rangle \leq 0,$$

providing

$$\langle d_k, \bar{x} - x_k \rangle \leq -r.$$

Thus (4.6) implies

$$\begin{aligned} \|\bar{x} - x_{k+1}\|^2 &= \|\bar{x} - x_k\|^2 + 2\lambda_k \langle d_k, \bar{x} - x_k \rangle + \lambda_k^2 \\ &\leq \|\bar{x} - x_k\|^2 - 2r\lambda_k + \lambda_k^2 \\ &\leq \|\bar{x} - x_k\|^2 - r\lambda_k, \end{aligned} \tag{4.7}$$

eventually. From this we obtain in particular that, if (4.6) holds and k is large enough,

$$d(x_{k+1}, \operatorname{Min} f) \leq d(x_k, \operatorname{Min} f). \tag{4.8}$$

Now suppose, by contradiction, there is $a > 0$ such that, for all large k,

$$d(x_k, \operatorname{Min} f) \geq a > 0. \tag{4.9}$$

From (4.7) we then get

$$\|\bar{x} - x_{k+i}\|^2 \leq \|\bar{x} - x_k\|^2 - r \sum_{j=k}^{k+i-1} \lambda_j \to -\infty,$$

which is impossible. It follows that $\liminf d(x_k, \operatorname{Min} f) = 0$. Now, fix $a > 0$ and K such that $\lambda_k < a$ for $k \geq K$. There is $k > K$ such that $d(x_k, \operatorname{Min} f) < a$. This implies

$$d(x_{k+1}, \operatorname{Min} f) < 2a.$$

Now, two cases can occur:

(i) $d(x_{k+2}, \operatorname{Min} f) < a$;
(ii) $d(x_{k+2}, \operatorname{Min} f) \geq a$.

In the second case, from (4.8) we can conclude that

$$d(x_{k+2}, \operatorname{Min} f) \leq d(x_{k+1}, \operatorname{Min} f) < 2a.$$

Thus, in any case, we have that

$$d(x_{k+2}, \operatorname{Min} f) < 2a.$$

By induction, we conclude that $d(x_n, \operatorname{Min} f) \leq 2a$ for all large n, and this ends the proof. □

With some changes in the above proof, it can be seen that the same result holds if we take $v_k \in \partial_{\varepsilon_k} f(x_k)$, for any sequence $\{\varepsilon_k\}$ converging to zero.

The above result can be refined if $\operatorname{Min} f$ has interior points.

Corollary 4.3.13 *With the assumptions of Theorem* 4.3.12*, if moreover* $\operatorname{int}\operatorname{Min} f \neq \emptyset$*, then* $v_k = 0$ *for some* k*.*

Proof. Suppose, by way of contradiction, $v_k \neq 0$ for all k. Let $\bar{x} \in \operatorname{int}\operatorname{Min} f$. Then there is $r > 0$ such that $B[\bar{x}; r] \subset \operatorname{Min} f$. Let $\tilde{x}_k = \bar{x} + r d_k$. Then $\tilde{x}_k \in B[\bar{x}; r] \subset \operatorname{Min} f$, hence $f(\tilde{x}_k) = \inf f$. Moreover,

$$f(y) \geq f(x_k) + \langle v_k, y - x_k \rangle \quad \forall y \in \mathbb{R}^n,$$

providing

$$f(\tilde{x}_k) \geq f(x_k) + \langle v_k, \tilde{x}_k - x_k \rangle.$$

Moreover, $f(x_k) \geq \inf f = f(\tilde{x}_k)$, hence

$$\langle v_k, \tilde{x}_k - x_k \rangle \leq 0.$$

We repeat what we did in the first part of Theorem 4.3.12 to get that

$$\|x_{k+s} - \bar{x}\|^2 \leq \|x_k - \bar{x}\|^2 - r \sum_{i=k}^{k+s-1} \lambda_i \to -\infty,$$

which provides the desired contradiction. □

The results above concern the case when f has a nonempty and bounded set of minimizers. The next result instead takes into account the case when the set of the minimizers of f in unbounded. As we shall see, we must put an extra condition on the size of the length steps λ_k. Thus, we shall suppose as before

$$\lambda_k \to 0,$$
$$\sum_{k=0}^{+\infty} \lambda_k = \infty,$$
$$v_k \in \partial f(x_k),$$
$$d_k = \begin{cases} 0 & \text{if } v_k = 0, \\ \frac{v_k}{\|v_k\|} & \text{if } v_k \neq 0. \end{cases}$$

Moreover, suppose

$$\sum_{k=0}^{\infty} {\lambda_k}^2 < \infty. \tag{4.10}$$

Then, the following result holds:

Theorem 4.3.14 *If* Min f *is nonempty, then the sequence* $\{x_k\}$ *converges to an element belonging to the set* Min f.

Proof. As in Theorem 4.3.12, we consider the case when $d_k \neq 0$ for all k. Let $\bar{x} \in \operatorname{Min} f$. Then

$$\begin{aligned} \|x_{k+1} - \bar{x}\|^2 &= \|x_k - \bar{x} - \lambda_k d_k\|^2 \\ &= \|x_k - \bar{x}\|^2 + 2\langle x_k - \bar{x}, -\lambda_k d_k\rangle + {\lambda_k}^2 \\ &\leq \|x_k - \bar{x}\|^2 + 2\frac{\lambda_k}{\|v_k\|}\langle v_k, \bar{x} - x_k\rangle + {\lambda_k}^2. \end{aligned} \tag{4.11}$$

Moreover,

$$f(y) - f(x_k) \geq \langle v_k, y - x_k\rangle \quad \forall y \in \mathbb{R}^n \quad \forall k,$$

whence

$$0 \geq \inf f - f(x_k) \geq \langle v_k, \bar{x} - x_k\rangle \quad \forall k. \tag{4.12}$$

From (4.11) we get

$$\begin{aligned} \|x_{k+1} - \bar{x}\|^2 &\leq \|x_k - \bar{x}\|^2 + {\lambda_k}^2 \\ &\leq \|x_0 - \bar{x}\|^2 + \sum_{i=0}^{k} {\lambda_i}^2. \end{aligned} \tag{4.13}$$

From (4.13) and (4.10) we see that the sequence $\{x_k\}$ is bounded. This implies that the sequence $\{v_k\}$ is also bounded, as f is Lipschitz on a ball containing $\{x_k\}$. We see now that there is a subsequence $\{x_{k_j}\}$ such that

$$a_{k_j} := \langle v_{k_s}, \bar{x} - x_{k_s}\rangle \to 0. \tag{4.14}$$

Otherwise, from (4.12) there would be $b > 0$ and $K \in \mathbb{R}$ such that

$$a_k \leq -b \quad \forall k > K.$$

From (4.11) we get

$$\|x_{k+1} - \bar{x}\|^2 \leq \|x_0 - \bar{x}\|^2 + 2\sum_{i=0}^{k} \frac{\lambda_i}{\|v_i\|}\langle v_i, \bar{x} - x_i\rangle + \sum_{i=0}^{k} {\lambda_i}^2,$$

implying

$$\lim_{k\to\infty} \|x_{k+1} - \bar{x}\|^2 = -\infty,$$

which is impossible. Thus, from (4.14) and (4.12) we get that

$$f(x_{k_s j}) \to f(\bar{x}).$$

As $\{x_{k_j}\}$ is bounded, it has a subsequence (still labeled by k_j) converging to some element x^*. Hence

$$\lim_{j\to\infty} f(x_{k_j}) = f(x^*)$$

implying $x^* \in \operatorname{Min} f$. It remains to prove that the whole sequence $\{x_k\}$ converges to x^*. From the fact that $\bar{x}$ is arbitrary in (4.11) and (4.12), we can put x^* there instead of $\bar{x}$. Given $\varepsilon > 0$, there exists $K_1 \in \mathbb{R}$ such that, if $k_j > K_1$,

$$\left\|x_{k_j} - x^*\right\|^2 < \frac{\varepsilon}{2}, \quad \text{and} \quad \sum_{i=k_j}^{\infty} {\lambda_i}^2 < \frac{\varepsilon}{2}.$$

Then, from (4.11) and (4.12) we get that

$$\left\|x_{k_j+n} - x^*\right\|^2 \leq \left\|x_{k_j} - x^*\right\|^2 + \sum_{i=k_j}^{k_j+n-1} {\lambda_i}^2 < \varepsilon \quad \forall n \geq 1.$$

This implies $x_k \to x^*$. □