3

The derivatives and the subdifferential

Something must still happen, but my strength is over,
my fingers empty gloves,
nothing extraordinary in my eyes,
nothing driving me.
(M. Atwood, *Surfacing*)

In the previous chapter we have seen that convex functions enjoy nice properties from the point of view of continuity. Here we see that the same happens with directional derivatives. The limit involved in the definition of directional derivative always exists, and thus in order to claim the existence of the directional derivative at a given point and along a fixed direction, it is enough to check that such a limit is a real number. Moreover, the directional derivative at a given point is a sublinear function, i.e., a very particular convex function, with respect to the direction.

We then introduce and study the very important concept of gradient. Remember that we are considering extended real valued functions. Thus it can happen that the interior of the effective domain of a function is empty. This would mean that a concept of derivative would be useless in this case. However, we know that a convex function which is differentiable at a given point enjoys the property that its graph lies above that tangent line at that point, a remarkable *global* property. This simple remark led to the very useful idea of subgradient for a convex function at a given point. The definition does not require that the function be real valued at a neighborhood of the point, keeps most of the important properties of the derivative (in particular, if zero belongs to the subdifferential of f at a given point x, then x is a global minimizer for f), and if f is smooth, then it reduces to the classical derivative of f. The subdifferential of f at a given point, i.e., the set of its subgradients at that point, is also related to its directional derivatives.

Clearly, an object such as the subdifferential is more complicated to handle than a derivative. For instance, the simple formula that the derivative of the sum of two functions f and g is the sum of the derivatives of f and g must be rewritten here, and its proof is not obvious at all. Moreover, studying continuity of the derivative here requires concepts of continuity for multivalued functions, which we briefly introduce. We also briefly analyze concepts of twice differentiability for convex functions, to see that the theory can be extended beyond the smooth case. Thus, the subdifferential calculus introduced and analyzed in this chapter is of the utmost importance in the study of convex functions.

3.1 Properties of the directional derivatives

We shall now see that the same happens with directional derivatives. In particular, the limit in the definition of the directional derivative at a given point and for a fixed direction always exists. Thus, to claim existence of a directional derivative it is enough to check that such a limit is a real number.

Definition 3.1.1 Let $f \in \Gamma(X), x, d \in X$. The *directional derivative* of f at x along the vector d, denoted by $f'(x;d)$, is the following limit:

$$f'(x;d) = \lim_{t \to 0^+} \frac{f(x+td) - f(x)}{t},$$

whenever it is finite.

Proposition 3.1.2 *Let $f \in \Gamma(X), x, d \in X$. The directional derivative of f at x along the vector d exists if and only if the quotient*

$$\frac{f(x+td) - f(x)}{t}$$

is finite for some $\bar{t} > 0$ and is lower bounded in $(0, \infty)$.

Proof. Let $x, d \in X$. We know from Proposition 1.2.11 that the function

$$0 < t \mapsto g(t;d) := \frac{f(x+td) - f(x)}{t},$$

is increasing. This implies that $\lim_{t \to 0^+} g(t;d)$ always exists and

$$\lim_{t \to 0^+} g(t;d) = \inf_{t > 0} g(t).$$

If there is $\bar{t} > 0$ such that $g(\bar{t}) \in \mathbb{R}$ and if g is lower bounded, then the limit must be finite. □

Of course, $\lim_{t \to 0^+} \frac{f(x+td)-f(x)}{t} = \infty$ if and only if $f(x+td) = \infty$ for all $t > 0$. Note that we shall use the word directional derivative, even if d is not a unit vector.

The next estimate for the directional derivative is immediate.

Proposition 3.1.3 *Let $f \in \Gamma(X)$ be Lipschitz with constant k in a neighborhood V of x. Then*

$$|f'(x;d)| \le k, \ \forall d \in X : \|d\| = 1.$$

Proposition 3.1.4 *Let $f \in \Gamma(X)$, and let $x \in \operatorname{dom} f$. Then $X \ni d \mapsto \lim_{t \to 0^+} \frac{f(x+td)-f(x)}{t}$ is a sublinear function.*

Proof. We shall prove that $X \ni d \mapsto g(t;d)$ is convex and positively homogeneous.

$$\begin{aligned} f(x + t(\lambda d_1 + (1-\lambda)d_2)) &= f(\lambda(x+td_1) + (1-\lambda)(x+td_2)) \\ &\le \lambda f(x+td_1) + (1-\lambda)f(x+td_2), \end{aligned}$$

providing convexity of $d \mapsto \lim_{t\to 0^+} \frac{f(x+td)-f(x)}{t}$. It is immediate to verify that it is positively homogeneous. □

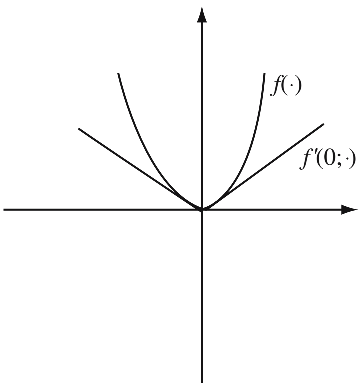


Figure 3.1.

The following example shows that the limit in the definition of the directional derivative can assume value $-\infty$.

$$f(x) = \begin{cases} -\sqrt{x} & \text{if } x \ge 0, \\ \infty & \text{elsewhere.} \end{cases}$$

If there exists d such that the limit in the definition is $-\infty$, as $f'(x;0) = 0$, then $d \mapsto \lim_{t\to 0^+} \frac{f(x+td)-f(x)}{t}$ is never lower semicontinuous, because a convex lower semicontinuous function assuming value $-\infty$ never assumes a real value (prove it, remembering Remark 1.2.6).

The next theorem provides a condition under which $d \mapsto f'(x;d) \in \Gamma(X)$.

Theorem 3.1.5 *Let $f \in \Gamma(X)$. Let $x_0 \in \operatorname{dom} f$. Suppose moreover,*

$$F := \mathbb{R}_+(\operatorname{dom} f - x_0)$$

is a closed vector space of X. Then $d \mapsto f'(x_0; d) \in \Gamma(X)$.

Proof. By translation, we can suppose that $x_0 = 0$. It is easy to show that

$$F = \bigcup_{n=1}^{\infty} nf^n.$$

As nf^n is a closed set for each $n \in \mathbb{R}$, and since F is a complete metric space, it follows from Baire's theorem that there exists $\bar{n}$ such that $\operatorname{int}_{|F} \bar{n} f^{\bar{n}}$ (hence $\operatorname{int}_{|F} f^{\bar{n}}) \neq \emptyset$. Thus f, restricted to F, is upper bounded on a neighborhood of a point $\bar{x}$. As $-t\bar{x} \in \operatorname{dom} f$ for some $t > 0$, it follows that $f_{|F}$ is upper bounded on a neighborhood of 0 (see the proof of Theorem 2.1.2), whence continuous and locally Lipschitz (Corollary 2.2.19) on a neighborhood of 0. It follows that $F \ni d \mapsto f'(0; d)$ is upper bounded on a neighborhood of zero and, by Proposition 2.1.5, is everywhere continuous. As $f'(0; d) = \infty$ if $d \notin F$ and F is a closed set, we conclude that $d \mapsto f'(x_0; d) \in \Gamma(X)$. □

Corollary 3.1.6 *Let $f \in \Gamma(X)$. Let $x_0 \in \operatorname{int} \operatorname{dom} f$. Then $d \mapsto f'(x_0; d)$ is a convex, positively homogeneous and everywhere continuous function.*

3.2 The subgradient

We now introduce the notion of subgradient of a function at a given point. It is a generalization of the idea of derivative, and it has several nice properties. It is a useful notion, both from a theoretical and a computational point of view.

Definition 3.2.1 Let $f \colon X \to (-\infty, \infty]$. $x^* \in X^*$ is said to be a *subgradient* of f at the point x_0 if $x_0 \in \operatorname{dom} f$ and $\forall x \in X$,

$$f(x) \geq f(x_0) + \langle x^*, x - x_0 \rangle.$$

The *subdifferential* of f at the point x_0, denoted by $\partial f(x_0)$, is the possibly empty set of all subgradients of f at the point x_0.

The above definition makes sense for any function f. However, a definition of derivative, as above, requiring a *global* property, is useful mainly in the convex case.

Definition 3.2.2 Let $A \subset X$ and $x \in A$. We say that $0^* \neq x^* \in X^*$ *supports A at x* if

$$\langle x^*, x \rangle \geq \langle x^*, a \rangle, \ \forall a \in A.$$

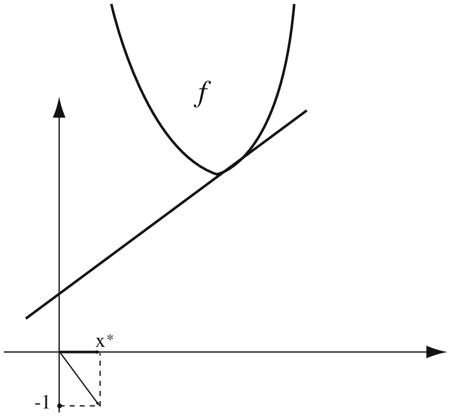


Figure 3.2. x^* is a subgradient of f at the point x_0.

Remark 3.2.3 $x^* \in \partial f(x_0)$ if and only if the pair $(x^*, -1)$ supports epi f at the point $(x_0, f(x_0))$. For, $\forall x \in X$

$$\langle x^*, x_0\rangle - f(x_0) \geq \langle x^*, x\rangle - r, \forall r \geq f(x) \Longleftrightarrow f(x) \geq f(x_0) + \langle x^*, x - x_0\rangle.$$

Example 3.2.4 Here are some examples of subgradients:

- $f(x) = |x|$. Then $\partial f(x) = \{\frac{x}{|x|}\}$ if $x \neq 0$, $\partial f(0) = [-1, 1]$ (try to extend this result to the function $f(x) = \|x\|$ defined on a Hilbert space X);
- $f\colon \mathbb{R} \to [0, \infty]$, $f(x) = I_{\{0\}}(x)$. Then $\partial f(0) = (-\infty, \infty)$;
- Let C be a closed convex set. $x^* \in \partial I_C(x) \Longleftrightarrow x \in C$ and $\langle x^*, c\rangle \leq \langle x^*, x\rangle, \forall c \in C$. That is, if $x^* \neq 0^*$, then $x^* \in \partial I_C(x)$ if and only if x^* supports C at x; $\partial I_C(x)$ is said to be the *normal cone* of C at x and it is sometimes indicated also by $N_C(x)$.

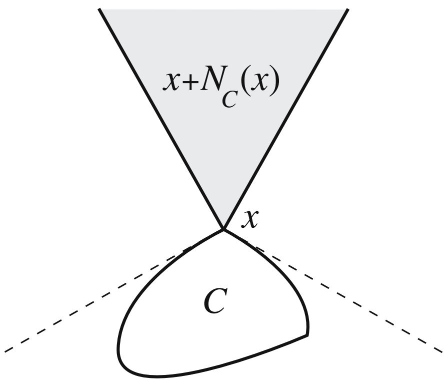


Figure 3.3. The normal cone to C at x.

- Let
$$f(x)=\begin{cases}-\sqrt{x} & \text{if } x\geq 0,\\ \infty & \text{otherwise .}\end{cases}$$
Then $\partial f(0)=\emptyset$, $\partial f(x)=\{-\frac{1}{2\sqrt{x}}\}$ if $x>0$.

Exercise 3.2.5 Let $f\colon \mathbb{R}^2 \to \mathbb{R}$ be the following function: $f(x,y) = \max\{|x|,|y|\}$. Find the subdifferential of f at the points $(0,0)$, $(0,1)$, $(1,1)$.

Hint. $\{(x^*,y^*):|x^*|+|y^*|\leq 1\}$,$\{(0,1)\}$, $\{(x^*,y^*):x^*\geq 0, y^*\geq 0, x^*+y^*=1\}$ respectively.

Definition 3.2.6 Given a Banach space X, the *duality mapping* $\delta\colon X\to X^*$ is defined as

$$\delta(x) := \{x^*\in X^* : \|x^*\|_* = 1 \text{ and } \langle x^*,x\rangle = \|x\|\}.$$

It is well known that $\delta(x)\neq\emptyset$ for all $x\in X$. The proof of this relies on the fact that the function $x^*\mapsto\langle x^*,x\rangle$ is weak* continuous.

Example 3.2.7 Let X be a Banach space, let $f(x)=\|x\|$. Then, for all $x\neq 0$,
$$(\partial\|\cdot\|)(x)=\delta(x).$$
We leave as an exercise the proof that $\delta(x)\subset(\partial\|\cdot\|)(x)$. To show the opposite inclusion, let $x^*\in\partial(\|x\|)$. Then, for all y,

$$\|y\|\geq\|x\|+\langle x^*,y-x\rangle. \tag{3.1}$$

The choices of $y=0$ and $y=2x$ show that

$$\langle x^*,x\rangle=\|x\|. \tag{3.2}$$

From (3.1) and (3.2) we get that

$$\|y\|\geq\langle x^*,y\rangle,\ \forall y\in X.$$

Combining this with (3.2), we conclude that $\|x^*\|_*=1$ and so $x^*\in\delta(x)$.

Exercise 3.2.5 shows that δ can be multivalued at some point. Those Banach spaces having a norm which is smooth outside the origin (in this case δ must be single valued) are important. We shall discuss this later.

Example 3.2.8 Let $X=l^2$, with $\{e_n\}_{n\in\mathbb{N}}$ the canonical basis, and $C=\{x\in l^2, x=(x_1,x_2,\dots,x_n,\dots):|x_n|\leq 2^{-n}\}$. Let

$$f(x)=\begin{cases}-\sum_{n=1}^{\infty}\sqrt{2^{-n}+x_n} & \text{if } x\in C,\\ \infty & \text{elsewhere.}\end{cases}$$

Then f is convex and its restriction to the set C is a continuous function. An easy calculation shows that $f'(0; e_n) = -2^{\frac{n-2}{2}}$. Now suppose $x^* \in \partial f(0)$. Then

$$f(2^{-n}e_n) \geq f(0) + \langle x^*, 2^{-n}e_n \rangle, \forall n \in \mathbb{N},$$

whence

$$(1-\sqrt{2})2^{\frac{n}{2}} \geq \langle x^*, e_n \rangle, \forall n \in \mathbb{N}.$$

Thus f has all directional derivatives at 0, but $\partial f(0) = \emptyset$. Observe that this cannot happen in finite dimensions, as Exercise 3.2.13 below shows.

Remark 3.2.9 Let $x \in \operatorname{dom} f$, $x^* \in \partial f(x)$, u^* in the normal cone to $\operatorname{dom} f$ at x ($\langle u^*, x - u \rangle \leq 0, \forall u \in \operatorname{dom} f$). Then $x^* + u^* \in \partial f(x)$. This does not provide any information if $x \in \operatorname{int} \operatorname{dom} f$, for instance if f is continuous at x, as the normal cone to $\operatorname{dom} f$ at x reduces to 0^*. However this information is interesting if $x \notin \operatorname{int} \operatorname{dom} f$. In many situations, for instance if X is finite-dimensional or if $\operatorname{dom} f$ has interior points, there exists at least a $0^* \neq u^*$ belonging to the normal cone at x, which thus is an unbounded set (the existence of such a $0^* \neq u^*$ in the normal cone follows from the fact that there is a hyperplane supporting $\operatorname{dom} f$ at x. The complete argument is suggested in Exercise 3.2.10). Hence, in the boundary points of $\operatorname{dom} f$ it can happen that the subdifferential of f is either empty or an unbounded set.

Exercise 3.2.10 Let X be a Banach space and let $\operatorname{int} \operatorname{dom} f \neq \emptyset$. Let $x \in \operatorname{dom} f \setminus \operatorname{int} \operatorname{dom} f$. Prove that the normal cone to $\operatorname{dom} f$ at the point x is unbounded.

Hint. Use Theorem A.1.5 by separating x from $\operatorname{int} \operatorname{dom} f$.

We now see how to evaluate the subdifferential of the inf convolution, at least in a particular case.

Proposition 3.2.11 *Let X be a Banach space, let $f, g \in \Gamma(X)$, let $x \in X$ and let u, v be such that*

$$u + v = x \text{ and } (f \nabla g)(x) = f(u) + g(v).$$

Then

$$\partial (f \nabla g)(x) = \partial f(u) \cap \partial g(v).$$

Proof. Let $x^* \in \partial f(u) \cap \partial g(v)$. Thus, for all $y \in X$ and $z \in X$

$$f(y) \geq f(u) + \langle x^*, y - u \rangle, \tag{3.3}$$

$$g(z) \geq g(v) + \langle x^*, z - v \rangle. \tag{3.4}$$

Let $w \in X$ and let $y, z \in X$ be such that $y + z = w$. Summing up (3.3) and (3.4) we get

$$f(y) + g(z) \geq (f \nabla g)(x) + \langle x^*, w - x \rangle. \tag{3.5}$$

By taking, in the left side of (3.5), the infimum over all y, z such that $y+z = w$, we can conclude that $x^* \in \partial(f \nabla g)(x)$. Conversely, suppose for all $y \in X$,

$$(f \nabla g)(y) \geq f(u) + g(v) + \langle x^*, y - (u+v) \rangle. \tag{3.6}$$

Then, given any $z \in X$, put $y = z + v$ in (3.6). We get

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

showing that $x^* \in \partial f(u)$. The same argument applied to $y = z + u$ shows that $x^* \in \partial g(v)$ and this ends the proof. □

The above formula applies to points where the inf-convolution is exact. A much more involved formula, involving approximate subdifferentials, can be shown to hold at any point. We shall use the above formula to calculate, in a Euclidean space, the subdifferential of the function $d(\cdot\,, C)$, where C is a closed convex set.

In the next few results we investigate the connections between the subdifferential of a function at a given point and its directional derivatives at that point.

Proposition 3.2.12 *Let $f \in \Gamma(X)$ and $x \in \operatorname{dom} f$. Then*

$$\partial f(x) = \{x^* \in X^* : \langle x^*, d \rangle \leq f'(x; d), \forall d \in X\}.$$

Proof. $x^* \in \partial f(x)$ if and only if

$$\frac{f(x+td) - f(x)}{t} \geq \langle x^*, d \rangle, \; \forall d \in X, \forall t > 0,$$

if and only if, taking the inf for $t > 0$ in the left side of the above inequality,

$$f'(x; d) \geq \langle x^*, d \rangle, \; \forall d \in X.$$

□

Exercise 3.2.13 If $f \in \Gamma(\mathbb{R}^n)$, if $f'(x; d)$ exists and is finite for all d, then $\partial f(x) \neq \emptyset$.

Hint. $f'(x; d)$ is sublinear and continuous. Now apply a corollary to the Hahn–Banach theorem (Corollary A.1.2) and Proposition 3.2.12.

Theorem 3.2.14 *Let $f \in \Gamma(X)$ and $x \in \operatorname{dom} f$. If*

$$F := \mathbb{R}_+(\operatorname{dom} f - x)$$

is a closed vector space, then

$$d \mapsto f'(x; d) = \sup\{\langle x^*, d \rangle : x^* \in \partial f(x)\}.$$

Proof. The function $d \mapsto f'(x;d)$ is sublinear (Proposition 3.1.4). From Theorem 3.1.5 $d \mapsto f'(x;d) \in \Gamma(X)$. Hence $d \mapsto f'(x;d)$ is the pointwise supremum of all linear functionals minorizing it (Corollary 2.2.22):

$$d \mapsto f'(x;d) = \sup\{\langle x^*, d\rangle : \langle x^*, d\rangle \le f'(x;d), \forall d \in X\}.$$

We conclude by Proposition 3.2.12, since $\langle x^*, d\rangle \le f'(x;d), \forall d \in X$ if and only if $x^* \in \partial f(x)$. □

The next theorem shows that the subdifferential is nonempty at "many" points.

Theorem 3.2.15 *Let $f \in \Gamma(X)$. Then $\partial f(x) \ne \emptyset, \forall x \in \operatorname{int}\operatorname{dom} f$.*

Proof. If $x \in \operatorname{int}\operatorname{dom} f$, then $\mathbb{R}_+(\operatorname{dom} f - x) = X$. Now apply Theorem 3.2.14. □

If X is finite dimensional, the previous result can be refined (same proof) since $\partial f(x) \ne \emptyset\ \forall x \in \operatorname{ri}\operatorname{dom} f$. In infinite dimensions it can be useless, since $\operatorname{dom} f$ could possibly have no interior points. But we shall show later that every function $f \in \Gamma(X)$ has a nonempty subdifferential on a dense subset of $\operatorname{dom} f$ (see Corollary 4.2.13).

From Propositions 3.1.3 and 3.2.12 we immediately get the following result providing an estimate from above of the norm of the elements in ∂f.

Proposition 3.2.16 *Let $f \in \Gamma(X)$ be Lipschitz with constant k in an open set $V \ni x$. Then*

$$\|x^*\| \le k, \forall x^* \in \partial f(x).$$

As a last remark we observe that the subdifferential keeps a fundamental property of the derivative of a convex function.

Proposition 3.2.17 *Let $f \in \Gamma(X)$. Then $0^* \in \partial f(\bar{x})$ if and only if $\bar{x}$ minimizes f on X.*

Proof. Obvious from the definition of subdifferential. □

3.3 Gâteaux and Fréchet derivatives and the subdifferential

Definition 3.3.1 Let $f\colon X \to (-\infty, \infty]$ and $x \in \operatorname{dom} f$. Then f is said to be *Gâteaux differentiable* at x if there exists $x^* \in X^*$ such that

$$f'(x;d) = \langle x^*, d\rangle,\ \forall d \in X.$$

And f is said to be *Fréchet differentiable* at x if there exists $x^* \in X^*$ such that

$$\lim_{d\to 0} \frac{f(x+d) - f(x) - \langle x^*, d\rangle}{\|d\|} = 0.$$

Gâteaux differentiability of f at x implies in particular that all the tangent lines to the graph of f at the point $(x, f(x))$, along all directions, lie in the same plane; Fréchet differentiability means that this plane is "tangent" to the graph at the point $(x, f(x))$.

Exercise 3.3.2 Show that if f is Gâteaux differentiable at x, the functional $x^* \in X^*$ given by the definition is unique. Show that Fréchet differentiability of f at x implies Gâteaux differentiability of f at x and that f is continuous at x. The opposite does not hold in general, as the example below shows.

Example 3.3.3 Let

$$f(x,y) = \begin{cases} 1 & \text{if } y \geq x^2 \text{or } y = 0, \\ 0 & \text{otherwise }. \end{cases}$$

Then all directional derivatives of f vanish at the origin, but f is not continuous at $(0,0)$, so that it is not Fréchet differentiable at the origin.

However, for convex functions in finite dimensions, the notions of Fréchet and Gâteaux differentiability agree, as we shall see.

We shall usually denote by $\nabla f(x)$ the unique $x^* \in X^*$ in the definition of Gâteaux differentiability. If f is Fréchet differentiable at x, we shall preferably use the symbol $f'(x)$ to indicate its Fréchet derivative at x.

Now a first result about Gâteaux differentiability in the convex case. Remember that the limit defining the directional derivative exists for every direction d; thus, in order to have Gâteaux differentiability, we only need to show that the limit is finite in any direction, and that there are no "angles".

Proposition 3.3.4 *Let $f \in \Gamma(X)$. Then f is Gâteaux differentiable at $x \in X$ if and only if $d \mapsto f'(x; d)$ upper bounded in a neighborhood of the origin and*

$$\lim_{t \to 0} \frac{f(x+td) - f(x)}{t}, \ \forall d \in X,$$

exists and is finite (as a two-sided limit).

Proof. The "only if" part is obvious. As far as the other one is concerned, observe that the equality between the right and left limits above means that $f'(x;-d) = -f'(x,d)$. Thus the function $d \mapsto f'(x;d)$, which is always sublinear, is in this case linear too. Upper boundedness next guarantees that $d \mapsto f'(x;d)$ is also continuous, and we conclude. □

The next exercise shows that Fréchet and Gâteaux differentiability do not agree in general for convex functions.

Exercise 3.3.5 Let $X = l^1$ with the canonical norm and let $f(x) = \|x\|$. Then f is Gâteaux differentiable at a point $x = (x_1, x_2, \dots,)$ if and only if $x_i \neq 0 \ \forall i$, and it is never Fréchet differentiable. $\nabla f(x) = x^* = (x_1^*, x_2^*, \dots)$, where $x_n^* = \frac{x_n}{|x_n|} := \operatorname{sgn} x_n$.

Hint. If, for some i, $x_i = 0$, then the limit

$$\lim_{t\to 0} \frac{f(x+te_i) - f(x)}{t},$$

does not exist, since the right limit is different from the left one. If $x_i \neq 0\ \forall i$, then for $\varepsilon > 0$, let N be such that $\sum_{i>N} |d_i| < \varepsilon$. For every small t,

$$\operatorname{sgn}(x_i + td_i) = \operatorname{sgn}(x_i), \quad \forall i \leq N.$$

Then

$$\Big|\frac{\|x+td\| - \|x\|}{t} - \sum_{i\in\mathbb{N}} d_i \operatorname{sgn} x_i\Big| < 2\varepsilon.$$

On the other hand, let x be such that $x_i \neq 0$ for all i and consider $d^n = (0, \dots, -2x_n, \dots)$. Then $d_n \to 0$, while

$$\Big|\|x+d^n\| - \|x\| - \sum_{i\in\mathbb{N}} d_i^n \operatorname{sgn} x_i\Big| = \|d^n\|,$$

showing that f is not Fréchet differentiable in x.

The concept of subdifferential extends the idea of derivative, in the sense explained in the following results.

Proposition 3.3.6 *Let $f \in \Gamma(X)$. If f is Gâteaux differentiable at x, then $\partial f(x) = \{\nabla f(x)\}$.*

Proof. By definition, $\forall d \in X$,

$$\lim_{t\to 0} \frac{f(x+td) - f(x)}{t} = \langle \nabla f(x), d\rangle.$$

As the function $0 < t \mapsto \frac{f(x+td)-f(x)}{t}$ is increasing,

$$\frac{f(x+td) - f(x)}{t} \geq \langle \nabla f(x), d\rangle,$$

whence

$$f(x+td) \geq f(x) + \langle \nabla f(x), td\rangle,\ \forall td \in X,$$

showing that $\partial f(x) \ni \nabla f(x)$. Now, let $x^* \in \partial f(x)$. Then

$$f(x+td) \geq f(x) + \langle x^*, td\rangle,\ \forall d \in X, \forall t > 0,$$

hence

$$\lim_{t\to 0^+} \frac{f(x+td) - f(x)}{t} := \langle \nabla f(x), d\rangle \geq \langle x^*, d\rangle,\ \forall d \in X,$$

whence $x^* = \nabla f(x)$. □

Proposition 3.3.7 *Let $f \in \Gamma(X)$. If f is continuous at x and if $\partial f(x) = \{x^*\}$, then f is Gâteaux differentiable at x and $\nabla f(x) = x^*$.*

Proof. First, observe that $d \mapsto f'(x;d)$ is everywhere continuous as $x \in \operatorname{int}\operatorname{dom} f$. Next, let $X \ni d$ be a (norm one) fixed direction. Let us consider the linear functional, defined on span $\{d\}$,

$$l_d(h) = af'(x;d) \quad \text{if } h = ad.$$

Then $l_d(h) \le f'(x;h)$ for all h in span $\{d\}$. The equality holds for $h = ad$ and $a > 0$, while $l_d(-d) = -f'(x;d) \le f'(x;-d)$. By the Hahn–Banach theorem (see Theorem A.1.1), there is a linear functional $x_d^* \in X^*$ agreeing with l_d on span $\{d\}$, and such that $\langle x_d^*, h\rangle \le f'(x;h)$ $\forall h \in X$. Then $x_d^* \in \partial f(x)$, so that $x_d^* = x^*$. As by construction $\langle x^*, d\rangle = f'(x;d)$ $\forall d \in X$, it follows that f is Gâteaux differentiable at x and $x^* = \nabla f(x)$. □

It may be worth noticing that in the previous result the assumption that f is continuous at x cannot be dropped. A set A (with empty interior) can have at a point x the normal cone reduced to the unique element zero (see Exercise A.1.8). Thus the indicator function of A is not Gâteaux differentiable at x, but $\partial I_A(x) = \{0\}$. Observe also that if $\operatorname{dom} f$ does have interior points, it is not possible that at a point x where f is not continuous, $\partial f(x)$ is a singleton (see Remark 3.2.9).

Recall that, denoting by $\{e_1, \dots, e_n\}$ the canonical basis in $\mathbb{R}^n$, the *partial derivatives* of f at x are defined as follows:

$$\frac{\partial f}{\partial x_i}(x) = \lim_{t \to 0} \frac{f(x + te_i) - f(x)}{t},$$

whenever the limit exists and is finite. Then we have the following proposition.

Proposition 3.3.8 *Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be convex. Then f is (Gâteaux) differentiable at $x \in \mathbb{R}^n$ if and only if the partial derivatives $\frac{\partial f}{\partial x_i}(x)$, $i = 1, \dots, n$ exist.*

Proof. Suppose there exist the partial derivatives of f at x. As f is continuous, $\partial f(x) \ne \emptyset$. Let $x^* \in \partial f(x)$, and write $x_i^* = \langle x^*, e_i\rangle$. Then $\forall t \ne 0$, $f(x + te_i) - f(x) \ge tx_i^*$, hence

$$\frac{\partial f}{\partial x_i}(x) = \lim_{t \to 0^+} \frac{f(x + te_i) - f(x)}{t} \ge x_i^*,$$

$$\frac{\partial f}{\partial x_i}(x) = \lim_{t \to 0^-} \frac{f(x + te_i) - f(x)}{t} \le x_i^*,$$

providing $x_i^* = \frac{\partial f}{\partial x_i}(x)$. Thus $\partial f(x)$ is a singleton, and we conclude with the help of Proposition 3.3.7. The opposite implication is an immediate consequence of Proposition 3.3.4. □

We shall see in Corollary 3.5.7 that Fréchet and Gâteaux differentiability actually agree for a convex function defined in a Euclidean space. The above proposition in turn shows that differentiability at a point is equivalent to the existence of the partial derivatives of f at the point.

3.4 The subdifferential of the sum

Let us consider the problem of minimizing a convex function f on a convex set C. This can be seen as the unconstrained problem of minimizing the function $f+I_C$. And $\bar{x} \in C$ is a solution of this problem if and only if $0 \in \partial(f+I_C)(\bar{x})$. Knowing this is not very useful unless $\partial(f+I_C) \subset \partial f + \partial I_C$. In such a case, we could claim the existence of a vector $x^* \in \partial f(\bar{x})$ such that $-x^*$ belongs to the normal cone of C at the point $\bar{x}$, a property that, at least when f is differentiable at $\bar{x}$, has a clear geometrical meaning. Unfortunately in general only the opposite relation holds true:

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

In the next exercise it can be seen that the desired relation need not be true.

Exercise 3.4.1 In $\mathbb{R}^2$ consider

$$A := \{(x,y) : y \geq x^2\},$$

$$B := \{(x,y) : y \leq 0\},$$

and their indicator functions I_A, I_B. Evaluate the subdifferential of I_A, I_B and of $I_A + I_B$ at the origin.

However, in some cases we can claim the desired result. Here is a first example:

Theorem 3.4.2 *Let $f, g \in \Gamma(X)$ and let $\bar{x} \in \operatorname{int}\operatorname{dom} f \cap \operatorname{dom} g$. Then, for all $x \in X$*

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

Proof. If $\partial(f+g)(x) = \emptyset$, there is nothing to prove. Otherwise, let $x^* \in \partial(f+g)(x)$. Then

$$f(y) + g(y) \geq f(x) + g(x) + \langle x^*, y-x \rangle, \ \forall y \in X. \tag{3.7}$$

Writing (3.7) in the form

$$f(y) - \langle x^*, y-x \rangle - f(x) \geq g(x) - g(y),$$

we see that

$$A := \{(y,a) : f(y) - \langle x^*, y-x \rangle - f(x) \leq a\},$$

$$B := \{(y, a) : g(x) - g(y) \geq a\}$$

are closed convex sets such that $\operatorname{int} A \neq \emptyset$ and $\operatorname{int} A \cap B = \emptyset$. From the Hahn–Banach theorem A.1.5, $\operatorname{int} A$ and B can be separated by a hyperplane that is not vertical, as is easy to see. Thus, there is an affine function $l(y) = \langle y^*, y\rangle + k$ such that

$$g(x) - g(y) \leq \langle y^*, y\rangle + k \leq f(y) - \langle x^*, y - x\rangle - f(x), \forall y \in X.$$

Setting $y = x$ we see that $k = \langle -y^*, x\rangle$, whence $\forall y \in X$,

$$g(y) \geq g(x) + \langle -y^*, y - x\rangle,$$

which gives $-y^* \in \partial g(x)$. Moreover, $\forall y \in X$,

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

so that $x^* + y^* \in \partial f(x)$. We thus have $x^* = -y^* + (x^* + y^*)$, with $-y^* \in \partial g(x)$ and $x^* + y^* \in \partial f(x)$. □

Exercise 3.4.3 Let $f : X \to \mathbb{R}$ be convex and lower semicontinuous and let C be a closed convex set. Then $\bar{x} \in C$ is a solution of the problem of minimizing f over C if and only if there is $x^* \in \partial f(\bar{x})$ such that $-x^*$ is in the normal cone to C at $\bar{x}$.

In the chapter dedicated to duality, the previous result will be specified when the set C is characterized by means of inequality constraints; see Theorem 5.4.2.

3.5 The subdifferential multifunction

In this section we shall investigate some properties of the subdifferential of f, considered as a multivalued function (multifunction) from X to X^*.

Proposition 3.5.1 *Let $f \in \Gamma(X)$ and $x \in X$. Then $\partial f(x)$ is a (possibly empty) convex and weakly* closed subset of X^*. Moreover, if f is continuous at x, then ∂f is bounded on a neighborhood of x.*

Proof. Convexity follows directly from the definition. Now, let $x^* \notin \partial f(x)$. This means that there is $y \in X$ such that

$$f(y) - f(x) < \langle x^*, y - x\rangle.$$

By the definition of weak* topology, it follows that for each z^* in a suitable (weak*) neighborhood of x^*, the same inequality holds. This shows that $\partial f(x)$ is weakly* closed. Finally, if f is continuous at x, it is upper and lower bounded around x, and thus it is Lipschitz in a neighborhood of x (Corollary 2.2.19). From Proposition 3.2.16 we get local boundedness of ∂f. □

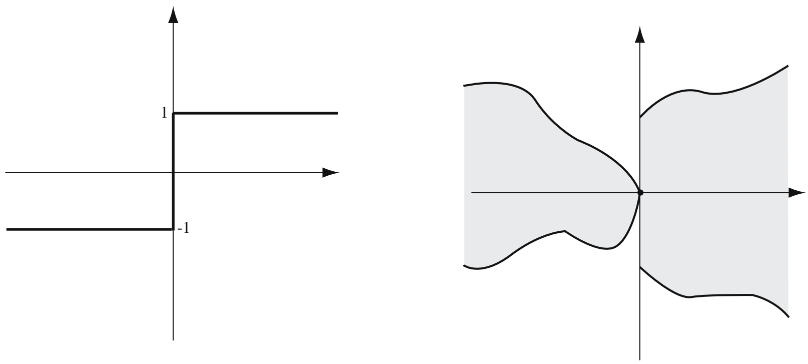
As a consequence of this, the *multifunction* $x \mapsto \partial f(x)$ is convex, weakly* closed valued, possibly empty valued at some x and locally bounded around x if x is a continuity point of f. We investigate now some of its continuity properties, starting with a definition.

Definition 3.5.2 Let (X, τ), (Y, σ) be two topological spaces and let $F\colon X \to Y$ be a given multifunction. Then F is said to be $\tau - \sigma$ *upper semicontinuous* at $\bar{x} \in X$ if for each open set V in Y such that $V \supset F(\bar{x})$, there is an open set $I \subset X$ containing $\bar{x}$ such that, $\forall x \in I$,

$$F(x) \subset V.$$

F is said to be $\tau - \sigma$ *lower semicontinuous* at $\bar{x} \in X$ if for each open set V in Y such that $V \cap F(\bar{x}) \neq \emptyset$, there is an open set $I \subset X$ containing $\bar{x}$ such that, $\forall x \in I$,

$$F(x) \cap V \neq \emptyset.$$



An upper semicontinuous multifunction not lower semicontinuous at 0.

A lower semicontinuous multifunction not upper semicontinuous at 0.

Figure 3.4.

Remark 3.5.3 The following facts are elementary to prove:

- If F is upper semicontinuous and if $F(x)$ is a singleton, then each *selection* of F (namely each function f such that $f(x) \in F(x), \forall x$) is continuous at x.
- Suppose $F(x)$ is a singleton for all x. Then if F is either upper semicontinuous or lower semicontinuous at a point, then it is continuous at that point, if it is considered as a *function*.

Exercise 3.5.4 Let X be a topological space and $f\colon X \to \mathbb{R}$ be a given function. Define the multifunction F on X as

$$F(x) = \{r \in \mathbb{R} : r \geq f(x)\},$$

i.e., the graph of F is the epigraph of f. Then F is upper semicontinuous at x if and only if f is lower semicontinuous at x.

The easy example of $f(x) = |x|$ shows that we cannot expect, in general, that ∂f be a lower semicontinuous multifunction. Instead, it enjoys upper semicontinuity properties, as we shall see in a moment.

Proposition 3.5.5 *Let $f \in \Gamma(X)$ be continuous and Gâteaux differentiable at x. Then the multifunction ∂f is norm-weak* upper semicontinuous at x.*

Proof. Let V be a weak* open set such that $V \supset \nabla f(x)$ and suppose there are a sequence $\{x_n\}$ converging to x and $x_n^* \in \partial f(x_n)$ such that $x_n^* \notin V$. As $\{x_n^*\}$ is bounded (see Proposition 3.5.1), it has a weak* limit x^* (it should be noticed that x^* is not necessarily limit of a subsequence). Now it is easy to show that $x^* \in \partial f(x) \subset V$, which is impossible. □

Proposition 3.5.6 *Let $f \in \Gamma(X)$ be Fréchet differentiable at x. Then the multifunction ∂f is norm-norm upper semicontinuous at x.*

Proof. Setting

$$g(\cdot) = f(\cdot + x) - f(x) - \langle f'(x), \cdot - x \rangle,$$

we have that $\partial g(\cdot) = \partial f(\cdot + x) - f'(x)$. Clearly, ∂g enjoys the same continuity properties at zero as ∂f at x. Thus we can suppose, without loss of generality, that $x = 0$, $f(x) = 0$, $f'(x) = 0^*$. By way of contradiction, suppose there are $\varepsilon > 0$, $\{x_n\}$ converging to 0, $x_n^* \in \partial f(x_n)$ for all n, such that $\{x_n^*\}$ is bounded and $\|x_n^*\| > 3\varepsilon$. Then there are $d_n \in X$ such that $\|d_n\| = 1$ and

$$\langle x_n^*, d_n \rangle > 3\varepsilon.$$

By definition of Fréchet differentiability, there is $\delta > 0$ such that

$$|f(x)| \leq \varepsilon \|x\|,$$

for all x such that $\|x\| \leq \delta$. As $x_n^* \in \partial f(x_n)$, then

$$\langle x_n^*, x \rangle \leq f(x) - f(x_n) + \langle x_n^*, x_n \rangle, \quad \forall x \in X.$$

Set $y_n = \delta d_n$, with n so large that $|f(y_n)| < \varepsilon\delta$, $|\langle x_n^*, x_n \rangle| < \varepsilon\delta$. Then

$$3\varepsilon\delta < \langle x_n^*, y_n \rangle \leq f(y_n) - f(x_n) + \langle x_n^*, x_n \rangle \leq \varepsilon\delta + \varepsilon\delta + \varepsilon\delta,$$

a contradiction. □

Corollary 3.5.7 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex. Then Gâteaux and Fréchet differentiability agree at every point.*

Proof. From Propositions 3.5.5 and 3.5.6. □

The next corollary shows a remarkable regularity property of the convex functions.

Corollary 3.5.8 *Let $f \in \Gamma(X)$ be Fréchet differentiable on an open convex set C. Then $f \in C^1(C)$.*

Proof. The *function* $f'(\cdot)$ is norm-norm continuous on C, being norm-norm upper semicontinuous as a multifunction. □

Corollary 3.5.9 *Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be convex and Gâteaux differentiable. Then $f \in C^1(\mathbb{R}^n)$.*

Proof. From Corollaries 3.5.7 and 3.5.8. □

Proposition 3.5.10 *Let $f \in \Gamma(X)$ be continuous at $x \in X$. If there exists a selection h of ∂f norm-weak* continuous (norm-norm continuous) at x, then f is Gâteaux (Fréchet) differentiable at x.*

Proof. Let us start with Gâteaux differentiability. For every $y \in Y$,

$$\langle h(x), y - x\rangle \le f(y) - f(x), \quad \langle h(y), x - y\rangle \le f(x) - f(y),$$

from which

$$0 \le f(y) - f(x) - \langle h(x), y - x\rangle \le \langle h(y) - h(x), y - x\rangle. \tag{3.8}$$

Setting $y = x + tz$, for small $t > 0$, and dividing by t, we get

$$0 \le \frac{f(x + tz) - f(x)}{t} - \langle h(x), z\rangle \le \langle h(x + tz) - h(x), z\rangle.$$

Letting $t \to 0^+$, and using the fact that h is norm-weak* continuous,

$$0 \le f'(x; z) - \langle h(x), z\rangle \le 0.$$

From (3.8) we also deduce

$$0 \le f(y) - f(x) - \langle h(x), y - x\rangle \le \|h(x) - h(y)\|\|x - y\|,$$

whence f is Fréchet differentiable provided h is norm-norm continuous. □

The next result extends to the subdifferential a well-known property of differentiable convex functions.

Definition 3.5.11 An operator $F\colon X \to X^*$ is said to be *monotone* if $\forall x, y \in X$, $\forall x^* \in F(x)$, $\forall y^* \in F(y)$,

$$\langle x^* - y^*, x - y\rangle \ge 0.$$

Proposition 3.5.12 *Let $f \in \Gamma(X)$. Then ∂f is a monotone operator.*

Proof. From

$$\langle x^*, y - x\rangle \le f(y) - f(x), \quad \langle y^*, x - y\rangle \le f(x) - f(y),$$

we get the result by addition. □

Proposition 3.5.12 can be refined in an interesting way.

Definition 3.5.13 A monotone operator $F\colon X \to X^*$ is said to be *maximal monotone* if $\forall y \in X$, $\forall y^* \notin F(y)$ there are $x \in X$, $x^* \in F(x)$ such that

$$\langle y^* - x^*, y - x\rangle < 0.$$

In other words, the graph of F is maximal in the class of the graph of monotone operators. We see now that the subdifferential is a maximal monotone operator.

Theorem 3.5.14 *Let $f\colon X \to \mathbb{R}$ be continuous and convex. Then ∂f is a maximal monotone operator.*

Proof. The geometric property of being maximal monotone does not change if we make a rotation and a translation of the graph of ∂f in $X \times X^*$. Thus we can suppose that $0 \notin \partial f(0)$ and we must find $x, x^* \in \partial f(x)$ such that $\langle x^*, x\rangle < 0$. As 0 is not a minimum point for f, there is $z \in X$ such that $f(0) > f(z)$. This implies that there exists $\bar{t} \in (0, 1]$ such that the directional derivative $f'(\bar{t}z; z) < 0$. Setting $x = \bar{t}z$, then $f'(x; x) < 0$. As $\partial f(x) \neq \emptyset$, if $x^* \in \partial f(x)$, then by Proposition 3.2.12 we get $\langle x^*, x\rangle < 0$. □

The above result holds for every function f in $\Gamma(X)$, but the proof in the general case is much more delicate. The idea of the proof is the same, but the nontrivial point, unless f is real valued, is to find, referring to the above proof, z and $\bar{t}$ such that $f'(\bar{t}z; z) < 0$. One way to prove it relies on a variational principle, as we shall see later (see Proposition 4.2.14).

3.6 Twice differentiable functions

In the previous section we have considered the subdifferential multifunction ∂f, and its continuity properties, relating them to some regularity of the convex function f. In this section, we define an additional regularity requirement for a multifunction, when when applied to the subdifferential of f, provides "second order regularity" for the function f. Let us start with two definitions.

Definition 3.6.1 Let X be a Banach space and $f \in \Gamma(X)$. Suppose $\bar{x} \in \operatorname{int}\operatorname{dom} f$. The subdifferential ∂f is said to be *Lipschitz stable* at $\bar{x}$ if $\partial f(\bar{x}) = \{\bar{p}\}$ and there are $\varepsilon > 0$, $K > 0$ such that

$$\|p - \bar{p}\| \le K\|x - \bar{x}\|,$$

provided $\|x - \bar{x}\| < \varepsilon, p \in \partial f(x)$.

Definition 3.6.2 Let X be a Banach space and $f \in \Gamma(X)$. Suppose $\bar{x} \in \operatorname{int}\operatorname{dom} f$. We say that ∂f is *Fréchet differentiable* at $\bar{x}$ if $\partial f(\bar{x}) = \{\bar{p}\}$ and there is a linear operator $T\colon X \to X^*$ such that

$$\lim_{x\to\bar{x}} \frac{\|p - \bar{p} - T(x-\bar{x})\|}{\|x-\bar{x}\|} = 0, \tag{3.9}$$

provided $p \in \partial f(x)$.

Definition 3.6.3 Let X be a Banach space and $f \in \Gamma(X)$. Suppose $\bar{x} \in \operatorname{int}\operatorname{dom} f$. We say that f is *twice Fréchet differentiable* at $\bar{x}$ if $\partial f(\bar{x}) = \bar{p}$ and there is a quadratic form $Q(x) := \langle Ax, x\rangle$ ($A\colon X \to X^*$ linear bounded operator) such that

$$\lim_{x\to\bar{x}} \frac{f(x) - \langle \bar{p}, x-\bar{x}\rangle - (1/2)Q(x-\bar{x})}{\|x-\bar{x}\|^2} = 0. \tag{3.10}$$

The following lemma shows that if two convex functions are close on a given bounded set and one of them is convex and the other is regular, the subdifferential of the convex function can be controlled (in a smaller set) by the derivative of the regular one, another nice property of convex functions.

Lemma 3.6.4 *Let $f\colon X \to (-\infty, \infty]$ be convex. Let $\delta, a > 0$, let $g\colon B(0;a) \to \mathbb{R}$ be a Fréchet differentiable function and suppose $|f(x) - g(x)| \le \delta$ for $x \in B(0;a)$. Let $0 < r < R \le a$, let x be such that $\|x\| \le r$ and $x^* \in \partial f(x)$. Then*

$$d\big(x^*, \operatorname{co}\{g'(B(x;R-r))\}\big) \le \frac{2\delta}{R-r}.$$

If g is convex, we also have

$$d\big(x^*, \operatorname{co}\{\partial g(B(0;R))\}\big) \le \frac{2\delta}{R-r}.$$

Proof. Without loss of generality we can suppose $x^* = 0$. Let α be such that $\alpha < \|y^*\|$ for all $y^* \in \operatorname{co}\{g'(B(x;R-r))\}$. Then there exists d, with $\|d\| = 1$, such that $\langle -y^*, d\rangle > \alpha$ for all $y^* \in \operatorname{co}\{g'(B(x,R-r))\}$. We have

$$\begin{aligned}\delta &\ge f\,(x + (R-r)d) - g\,(x + (R-r)d)\\ &\ge f(x) - g(x) - (g(x + (R-r)d) - g(x))\,.\end{aligned}$$

There is an $s \in (0, R-r)$ such that

$$\langle (R-r)g'(x+sd), d\rangle = g(x + (R-r)d) - g(x).$$

Thus

$$2\delta \ge (R-r)\langle -g'(x+sd), d\rangle \ge (R-r)\alpha.$$

Then

$$\alpha \le \frac{2\delta}{R-r},$$

and this ends the proof of the first claim. About the second one, let d be such that $\|d\| = 1$ and $\langle -y^*, d\rangle > \alpha$ for all $y^* \in \text{co}\{\partial g(B(0;R))\}$. Let $z^* \in \partial g(x + (R-r)d)$. Then

$$2\delta \ge g(x) - g(x + (R-r)d) \ge (R-r)\langle -z^*, d\rangle \ge (R-r)\alpha,$$

and we conclude as before. □

Remark 3.6.5 The above result can be refined in a sharp way by using the Ekeland variational principle, as we shall see in Lemma 4.2.18.

We are ready for our first result, which appears to be very natural, since it states that the variation of the function minus its linear approximation is of quadratic order if and only if the variation of its subdifferential is of the first order (thus extending in a natural way well-known properties of smooth functions).

Proposition 3.6.6 *Let* $\{\bar p\} = \partial f(\bar x)$. *Then the following two statements are equivalent:*

(i) ∂f *is Lipschitz stable at* $\bar x$*;*

(ii) *There are* $k > 0$ *and a neighborhood* $W \ni \bar x$ *such that*

$$|f(x) - f(\bar x) - \langle \bar p, x\rangle| \le k(\|x - \bar x\|)^2,$$

for all $x \in W$.

Proof. First, let us observe that we can suppose, without loss of generality,

$$\bar x = 0, \quad f(\bar x) = 0, \quad \bar p = 0,$$

by possibly considering the function

$$\hat f(x) = f(x + \bar x) - f(\bar x) - \langle \bar p, x\rangle.$$

In this case observe that $h(x) \ge 0$, $\forall x \in X$. Let us prove that (i) implies (ii). Let $H, K > 0$ be such that, if $\|x\| \le H$, $p \in \partial f(x)$, then

$$\|p\| \le K\|x\|.$$

Since

$$0 = f(0) \ge f(x) + \langle p, -x\rangle,$$

we have

$$f(x) \le \|p\|\|x\| \le K\|x\|^2.$$

We now prove that (ii) implies (i). Suppose there are $a, K > 0$ such that

$$|f(x)| \leq K\|x\|^2,$$

if $\|x\| \leq a$. Now take x with $r := \|x\| \leq (a/2)$. We have then

$$|f(x)| \leq Kr^2.$$

We now apply Lemma 3.6.4 to f and to the zero function, with a, r as above, $R = 2r$ and $\delta = Kr^2$. We then get

$$\|p\| \leq 2Kr = 2K\|x\|,$$

provided $\|x\| \leq (a/2)$. □

The following result connects Fréchet differentiability of ∂f with twice Fréchet differentiability of f. This result too is quite natural.

Proposition 3.6.7 *Let $\bar{p} \in \partial f(\bar{x})$. Then the following two statements are equivalent:*

(i) *∂f is Fréchet differentiable at $\bar{x}$;*
(ii) *f is twice Fréchet differentiable at $\bar{x}$.*

Proof. As in the previous proposition, we can suppose

$$\bar{x} = 0, \quad f(\bar{x}) = 0, \quad \bar{p} = 0.$$

Let us show that (i) implies (ii). Assume there is an operator T as in (3.9), and let Q be the quadratic function associated to it: $Q(u) = \frac{1}{2}\langle Tu, u\rangle$. Setting $h(s) = f(sx)$ we have that

$$f(x)(-f(0) = 0) = h(1) - h(0) = \int_0^1 h'(s)\, ds = \int_0^1 f'(sx; x)\, ds.$$

Now, remembering that $f'(sx; x) = \sup_{p \in \partial f(sx)} \langle p, x\rangle$ (see Theorem 3.2.14), we then have

$$f(x) - \frac{1}{2}Q(x) = \int_0^1 \Big[\sup_{p \in \partial f(sx)} \langle p, x\rangle - s\langle Tx, x\rangle\Big]\, ds,$$

from which we get

$$|f(x) - \frac{1}{2}Q(x)| \leq \int_0^1 \sup_{p \in \partial f(sx)} |\langle p - Tsx, x\rangle|\, ds;$$

from this, remembering (3.9), we easily get (3.10). The proof that (ii) implies (i) relies again on Lemma 3.6.4. There is a quadratic function Q of the form $Q(x) = \langle Tx, x\rangle$, such that there are $a, \varepsilon > 0$ with

$$|f(x) - \frac{1}{2}Q(x)| \leq \varepsilon\|x\|^2,$$

if $\|x\| \leq a$. Now take x such that $r := \|x\| \leq \frac{a}{2}$. We have then

$$|f(x) - \frac{1}{2}Q(x)| \leq \varepsilon r^2.$$

We apply Lemma 3.6.4 to f and to the function $\frac{1}{2}Q$, with a, r as above, $R = r(1 + \sqrt{\varepsilon})$ and $\delta = \varepsilon r^2$. We then get

$$\mathrm{d}(q, \mathrm{co}\{T(B(x, \sqrt{\varepsilon}r))\} \leq \frac{2\varepsilon r^2}{\sqrt{\varepsilon}r},$$

provided $\|x\| \leq \frac{a}{2}$. But then

$$\|p - Tx\| \leq 2\sqrt{\varepsilon}\|x\| + \|T\|\sqrt{\varepsilon}\|x\|,$$

and from this we easily get (3.10). □

3.7 The approximate subdifferential

There are both theoretical and practical reasons to define the concept of approximate subdifferential. On the one hand, the (exact) subdifferential does not exist at each point of $\mathrm{dom}\, f$. On the other hand, it is also difficult to evaluate. To partly overcome these difficulties the notion of approximate subdifferential is introduced.

Definition 3.7.1 Let $\varepsilon \geq 0$ and $f\colon X \to (-\infty, \infty]$. Then $x^* \in X^*$ is said to be an *ε-subgradient* of f at x_0 if

$$f(x) \geq f(x_0) + \langle x^*, x - x_0 \rangle - \varepsilon.$$

The *ε-subdifferential* of f at x, denoted by $\partial_\varepsilon f(x)$, is the set of the ε-subgradients of f at x.

Clearly, the case $\varepsilon = 0$ recovers the definition of the (exact) subdifferential. Moreover,

$$\partial f(x) = \bigcap_{\varepsilon > 0} \partial_\varepsilon f(x).$$

Here is a first result.

Theorem 3.7.2 *Let $f \in \Gamma(X)$, $x \in \mathrm{dom}\, f$. Then $\emptyset \neq \partial_\varepsilon f(x)$ is a weak* closed and convex set, $\forall \varepsilon > 0$. Furthermore,*

$$\partial_{\lambda\alpha + (1-\lambda)\beta} f(x) \supset \lambda \partial_\alpha f(x) + (1 - \lambda)\partial_\beta f(x),$$

for every $\alpha, \beta > 0$, for every $\lambda \in [0, 1]$.

Proof. To prove that $\partial_\varepsilon f(x) \neq \emptyset$, one exploits the usual separation argument of Lemma 2.2.16, by separating $(x, f(x) - \varepsilon)$ from $\mathrm{epi}\, f$; proving the other claims is straightforward. □

We provide two examples.

Example 3.7.3

$$f(x) = \begin{cases} -2\sqrt{x} & \text{if } x \geq 0, \\ \infty & \text{otherwise.} \end{cases}$$

It is not hard to see that for $\varepsilon > 0$, the ε-subdifferential of f at the origin is the half line $(-\infty, -\frac{1}{\varepsilon}]$, an unbounded set (not surprising, see Remark 3.2.9). On the other hand, the subdifferential of f at the origin is empty.

Example 3.7.4 Let $f(x) = |x|$. Then

$$\partial_\varepsilon f(x) = \begin{cases} [-1, -1 - \frac{\varepsilon}{x}] & \text{if } x < -\frac{\varepsilon}{2}, \\ [-1, 1] & \text{if } -\frac{\varepsilon}{2} \leq x \leq \frac{\varepsilon}{2}, \\ [1 - \frac{\varepsilon}{x}, 1] & \text{if } x > \frac{\varepsilon}{2}. \end{cases}$$

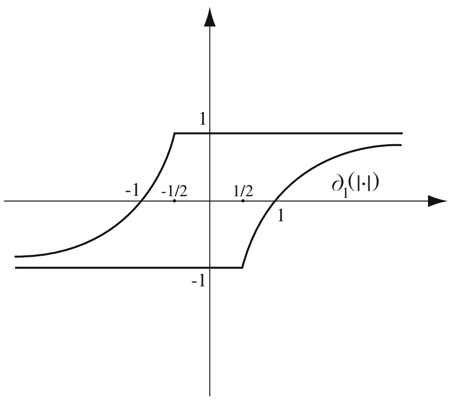


Figure 3.5. The approximate subdifferential $\partial_1(|\cdot|)(0)$.

The following result is easy and provides useful information.

Theorem 3.7.5 *Let $f \in \Gamma(X)$. Then $0^* \in \partial_\varepsilon f(x_0)$ if and only if*

$$\inf f \geq f(x_0) - \varepsilon.$$

Thus, whenever an algorithm is used to minimize a convex function, if we look for an ε-solution, it is enough that $0 \in \partial_\varepsilon f(x)$, a much weaker condition than $0 \in \partial f(x)$.

We now see an important connection between the ε-subdifferential and the directional derivatives (compare the result with Theorem 3.2.14).

Proposition 3.7.6 *Let $f \in \Gamma(X)$, $x \in \operatorname{dom} f$. Then, $\forall d \in X$,*

$$f'(x;d) = \lim_{\varepsilon \to 0^+} \sup\{\langle x^*, d\rangle : x^* \in \partial_\varepsilon f(x)\}.$$

Proof. Observe at first that, for monotonicity reasons, the limit in the above formula always exists. Now, let $\varepsilon > 0$ and $d \in X$; then, $\forall t > 0$, $\forall x^* \in \partial_\varepsilon f(x)$,

$$\frac{f(x+td) - f(x) + \varepsilon}{t} \geq \langle x^*, d\rangle.$$

Setting $t = \sqrt{\varepsilon}$, we get

$$\frac{f(x+\sqrt{\varepsilon}d) - f(x) + \varepsilon}{\sqrt{\varepsilon}} \geq \sup\{\langle x^*, d\rangle : x^* \in \partial_\varepsilon f(x)\}.$$

Taking the limit in the formula above,

$$f'(x;d) \geq \lim_{\varepsilon \to 0^+} \sup\{\langle x^*, d\rangle : x^* \in \partial_\varepsilon f(x)\},$$

which shows one inequality. To get the opposite one, it is useful to appeal again to a separation argument. Let $\alpha < f'(x;d)$ and observe that for $0 \leq t \leq 1$,

$$f(x+td) \geq f(x) + t\alpha.$$

Consider the line segment

$$S = \{(x, f(x) - \varepsilon) + t(d, \alpha) : 0 \leq t \leq 1\}.$$

S is a compact convex set disjoint from $\operatorname{epi} f$. Thus there are $y^* \in X^*$, $r \in \mathbb{R}$ such that

$$\langle y^*, y\rangle + rf(y) > \langle y^*, x + td\rangle + r(f(x) - \varepsilon + t\alpha),$$

$\forall y \in \operatorname{dom} f$, $\forall t \in [0,1]$. As usual, $r > 0$. Dividing by r and setting $x^* = -\frac{y^*}{r}$, we get

$$\langle x^*, d\rangle \geq \alpha - \varepsilon,$$

(with the choice of $y = x, t = 1$), and if $v \in X$ is such that $x + v \in \operatorname{dom} f$, setting $y = x + v$ and $t = 0$,

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

which means $x^* \in \partial_\varepsilon f(x)$. The last two facts provide

$$\sup\{\langle x^*, d\rangle : x^* \in \partial_\varepsilon f(x)\} \geq \alpha - \varepsilon,$$

and this ends the proof. □

We state, without proof, a result on the sum of approximate subdifferentials. To get an equality in the stated formula, one needs to add conditions as, for instance, $\operatorname{int} \operatorname{dom} f \cap \operatorname{int} \operatorname{dom} g \neq \emptyset$.

Proposition 3.7.7 *Let $\varepsilon \geq 0$ and $x \in \operatorname{dom} f \cap \operatorname{dom} g$. Then*

$$\partial_\varepsilon (f+g)(x) \supset \cup\{\partial_\sigma f(x) + \partial_\delta g(x) : 0 \leq \sigma, 0 \leq \delta, \sigma + \delta \leq \varepsilon\}.$$