An Invitation to Convex Functions Theory

Constantin P. Niculescu[1]

[1] Partially supported by Grant CEX05-D11-36.

Contents

Introduction

Convexity is a simple and natural notion which plays an important role both in pure and applied mathematics. However, the recognition of this subject as one that deserves to be studied in its own is generally traced to J. L. W. V. Jensen [27], [28]. During the 20th Century an intense research activity was done and significant results were obtained related to the theory of convex optimization and the isoperimetric problems. In fact, convex functions have two basic properties, that make them widely used in theoretical and applied mathematics:

a) their maximum is attained on the boundary of their domain of definition;

b) a strictly convex function admits at most one minimum.

Our presentation here aims to be a thorough introduction to the contemporary convex functions theory. It covers a large variety of subjects, from the convex calculus to the variational approach of partial differential equations.

The reader is only assumed to know elementary analysis and have some acquaintance with functional analysis. This can be covered from many textbooks such as [31].

Formally, our presentation is an extended version of our talk given at the *Symposium on Order Structures in Functional Analysis* (Bucharest, October 28, 2005). In this respect we acknowledge the financial support of Grant CEX05-D11-36, that made possible our participation.

At the moment there are available many books on convex functions theory commenting on different aspects of this vast subject. We cite here: J. M. Borwein and A. S. Lewis [4], I. Ekeland and R. Temam [13], L. Hörmander [26], J. E. Pečarić, F. Proschan and Y. C. Tong [37], R. R. Phelps [39], R. T. Rockafellar [42], C. Villani [44], and R. Webster [45]. In our joint book with L.-E. Persson [35] we tried to cover some other recent topics such as the Brunn–Minkowski inequality and the different theories concerning the convex-like functions.

Chapter 1

Background on Convex Sets

The natural domain for a convex function is a convex set. That is why we shall start by recalling some basic facts on convex sets, which should prove useful for understanding the general concept of convexity.

1.1 Convex Sets

All ambient linear spaces are assumed to be real.

A subset C of a linear space E is said to be *convex* if it contains the line segment

$$[x, y] = \{(1-\lambda)x + \lambda y : \lambda \in [0,1]\}$$

connecting any of its points x and y.

For example, convex sets in $\mathbb{R}^2$ include line segments, half-spaces, open or closed triangles, or open discs (plus any part of the boundary).

Many other examples can be obtained by considering the following operation with sets:

$$\lambda A + \mu B = \{\lambda x + \mu y : x \in A, y \in B\},$$

for $A, B \subset E$ and $\lambda, \mu \in \mathbb{R}$. One can prove easily that $\lambda A + \mu B$ is convex, provided that A and B are convex and $\lambda, \mu \geq 0$.

A subset A of E is said to be *affine* if it contains the whole line through any two of its points. Algebraically, this means

$$x, y \in A \text{ and } \lambda \in \mathbb{R} \;\text{ imply }\; (1-\lambda)x + \lambda y \in A.$$

Clearly, any affine subset is also convex (but the converse is not true). It is important to notice that any affine subset A is just the translate of a (unique) linear subspace L (and all translates of a linear space represent affine sets). In fact, for every $a \in A$, the translate

$$L = A - a$$

is a linear space and it is clear that $A = L + a$. For the uniqueness part, notice that if L and M are linear subspaces of E and $a, b \in E$ verify

$$L + a = M + b,$$

then necessarily $L = M$ and $a - b \in L$.

This remark allows us to introduce the concept of *dimension* for an affine set (as the dimension of the linear subspace of which it is a translate).

Given a finite family $x_1, \ldots, x_n$ of points in E, an *affine combination* of them is any point of the form

$$x = \sum_{k=1}^{n} \lambda_k x_k$$

where $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$, and $\sum_{k=1}^{n} \lambda_k = 1$. If, in addition, $\lambda_1, \ldots, \lambda_n \geq 0$, then x is called a *convex combination* (of $x_1, \ldots, x_n$).

1.1.1. Lemma. *A subset C of E is convex (respectively affine) if and only if it contains every convex (respectively affine) combination of points of C.*

Proof. The sufficiency part is clear, while the necessity part can be proved by mathematical induction. ■

Given a subset A of E, the intersection $\operatorname{co}(A)$ of all convex subsets containing A is convex and thus it is the smallest set of this nature containing A. We call it the *convex hull* of A. By using Lemma 1.1.1, one can verify easily that $\operatorname{co}(A)$ consists of all convex combinations of elements of A. The affine variant of this construction yields the *affine hull* of A, denoted $\operatorname{aff}(A)$. As a consequence we can introduce the concept of dimension for convex sets to be the dimension of their affine hulls.

1.1.2. Theorem (Carathéodory's theorem). *Suppose that A is a subset of a linear space E and its convex hull $\operatorname{co}(A)$ has dimension m. Then each point x of $\operatorname{co}(A)$ is the convex combination of at most $m+1$ points of A.*

Proof. Suppose that $x = \sum_{k=0}^{n} \lambda_k x_k$, where $x_k \in A$, $\lambda_k > 0$ and $\sum_{k=0}^{n} \lambda_k = 1$. If $n > m$, then the set $B = \{x_0, \ldots, x_n\}$ verifies

$$\dim(\operatorname{aff}(B)) \leq \dim(\operatorname{aff}(A)) = m \leq n - 1$$

and thus $\{x_1 - x_0, \ldots, x_n - x_0\}$ is a linearly dependent set. This gives us a set of real numbers $\mu_0, \ldots, \mu_n$, not all 0, such that $\sum_{k=0}^{n} \mu_k x_k = 0$ and $\sum_{k=0}^{n} \mu_k = 0$. Choose $t > 0$ for which $\upsilon_k = \lambda_k - t\mu_k \geq 0$ for $k = 0, \ldots, n$ and $\upsilon_j = 0$ for some index j. This allows us to reduce the number of terms in the representation of x. Indeed,

$$x = \sum_{k=0}^{n} \lambda_k x_k = \sum_{k=0}^{n} (\upsilon_k + t\mu_k) x_k = \sum_{k \neq j} \upsilon_k x_k,$$

and $\sum_{k \neq j} \upsilon_k = \sum_{k=0}^{n} \upsilon_k = \sum_{k=0}^{n} (\lambda_k - t\mu_k) = \sum_{k=0}^{n} \lambda_k = 1$. ■

The sets of the form $C = \mathrm{co}(\{x_0, \ldots, x_n\})$ are usually called *polytopes*. If $x_1 - x_0, \ldots, x_n - x_0$ are linearly independent, then C is called an *n-simplex* (with vertices $x_0, \ldots, x_n$); in this case, $\dim C = n$. Any point x in an n-simplex C has a unique representation $x = \sum_{k=0}^{n} \lambda_k x_k$, as a convex combination. In this case, the numbers $\lambda_0, \ldots, \lambda_n$ are called the *barycentric coordinates* of x.

An important class of convex sets are the convex cones. A *convex cone* in E is a subset C with the following two properties:

$$C + C \subset C$$
$$\lambda C \subset C \quad \text{for all } \lambda > 0.$$

Interesting examples are:

- $\mathbb{R}^n_+ = \{(x_1, \ldots, x_n) \in \mathbb{R}^n : x_1, \ldots, x_n \geq 0\}$, the nonnegative orthant;
- $\mathbb{R}^n_{++} = \{(x_1, \ldots, x_n) \in \mathbb{R}^n : x_1, \ldots, x_n > 0\}$;
- $\mathrm{Sym}^+(n, \mathbb{R})$, the set of all positive matrices A of $\mathrm{M}_n(\mathbb{R})$, that is,
$$\langle Ax, x\rangle \geq 0 \quad \text{for all } x \in \mathbb{R}^n;$$
- $\mathrm{Sym}^{++}(n, \mathbb{R})$, the set of all strictly positive matrices A of $\mathrm{M}_n(\mathbb{R})$, that is,
$$\langle Ax, x\rangle > 0 \quad \text{for all } x \in \mathbb{R}^n,\ x \neq 0.$$

The cones C containing the origin are important because of the ordering they induce:

$$x \leq y \quad \text{if and only if } y - x \in C.$$

So far we have not used any topology; only the linear properties of the space E have played a role.

Suppose now that E is a linear normed space. The following two results relate convexity and topology:

1.1.3. Lemma. *If U is a convex set in a linear normed space, then its interior $\mathrm{int}\, U$ and its closure $\overline{U}$ are convex as well.*

Proof. For example, if $x, y \in \mathrm{int}\, U$, and $\lambda \in (0, 1)$, then

$$\lambda x + (1 - \lambda)y + u = \lambda(x + u) + (1 - \lambda)(y + u) \in U$$

for all u in a suitable ball $B_\varepsilon(0)$. This shows that $\mathrm{int}\, U$ is a convex set. Now let $x, y \in \overline{U}$. Then there exist sequences $(x_k)_k$ and $(y_k)_k$ in U, converging to x and y respectively. This yields $\lambda x + (1 - \lambda)y = \lim_{k\to\infty}[\lambda x_k + (1 - \lambda)y_k] \in \overline{U}$ for all $\lambda \in [0, 1]$, that is, $\overline{U}$ is convex as well. ■

Notice that affine sets in $\mathbb{R}^n$ are closed because finite dimensional subspaces are always closed.

1.1.4. Lemma. *If U is an open set in a linear normed space E, then its convex hull is open. If E is finite dimensional and K is a compact set, then its convex hull is compact.*

Proof. For the first assertion, let $x = \sum_{k=0}^{m} \lambda_k x_k$ be a convex combination of elements of the open set U. Then

$$x + u = \sum_{k=0}^{m} \lambda_k (x_k + u) \quad \text{for all } u \in E$$

and since U is open it follows that $x_k + u \in U$ for all k, provided that $\|u\|$ is small enough. Consequently, $x + u \in \mathrm{co}\,(U)$ for u in a ball $B_\varepsilon(0)$.

We pass now to the second assertion. Clearly, we may assume that $E = \mathbb{R}^n$. Then consider the map defined by

$$f(\lambda_0, \ldots, \lambda_n, x_0, \ldots, x_n) = \sum_{k=0}^{n} \lambda_k x_k,$$

where $\lambda_0, \ldots, \lambda_n \in [0,1]$, $\sum_{k=0}^{n} \lambda_k = 1$, and $x_0, \ldots, x_n \in K$. Since f is continuous and its domain of definition is a compact space, so is the range of f. According to Carathéodory's theorem, the range of f is precisely $\mathrm{co}\,(K)$, and this ends the proof. ∎

While working with a convex subset A of $\mathbb{R}^n$, the natural space containing it is often $\mathrm{aff}\,(A)$, not $\mathbb{R}^n$, which may be far too large. For example, if $\dim A = k < n$, then A has empty interior. We can talk more meaningfully about the topological notions of interior and boundary by using the notions of relative interior and relative boundary. If A is a convex subset of $\mathbb{R}^n$, the *relative interior* of A, denoted $\mathrm{ri}\,(A)$, is the interior of A relative to $\mathrm{aff}\,(A)$. That is, $a \in \mathrm{ri}\,(A)$ if and only if there is an $\varepsilon > 0$ such that $B_\varepsilon(a) \cap \mathrm{aff}\,(A) \subset A$. We define the *relative boundary* of A, denoted $\mathrm{rbd}\,(A)$, as $\mathrm{rbd}\,(A) = \bar{A} \setminus \mathrm{ri}\,(A)$. These notions are important in optimization theory; see J. M. Borwein and A. S. Lewis [4].

It is well known that all norms on $\mathbb{R}^n$ give rise to the same topology. All (nonempty) open convex subsets of $\mathbb{R}^n$ are homeomorphic. For example, if B is the open unit ball of the Euclidean space $\mathbb{R}^n$, then the mapping $x \to x/(1 - \|x\|^2)$ provides a homeomorphism between B and $\mathbb{R}^n$.

1.2 The Orthogonal Projection

In any normed linear space E we can speak about the *distance* from a point $u \in E$ to a subset $A \subset E$. This is defined by the formula

$$d(u, A) = \inf\{\|u - a\| : a \in A\}$$

and represents a numerical indicator of how well u can be approximated by the elements of A. When $E = \mathbb{R}^3$ and A is the x-y plane, the Pythagorean theorem shows that $d(u, A)$ is precisely the distance between u and its orthogonal projection on that plane.

This remark has a notable generalization which will be presented in what follows.

1.2.1. Theorem. *Let C be a nonempty closed convex subset of a Hilbert space H (particularly, of the Euclidean space $\mathbb{R}^n$). Then for each $x \in H$ there is a unique point $P_C(x)$ of C such that*

$$d(x, C) = \|x - P_C(x)\|.$$

We call $P_C(x)$ the *orthogonal projection* of x onto C (or the nearest point of C to x).

Proof. The existence of $P_C(x)$ follows from the definition of the distance from a point to a set and the special geometry of the ambient space. In fact, any sequence $(y_n)_n$ in C such that $\|x - y_n\| \to \alpha = d(x, C)$ is a Cauchy sequence. This is a consequence of the following identity,

$$\|y_m - y_n\|^2 + 4\left\|x - \frac{y_m + y_n}{2}\right\|^2 = 2(\|x - y_m\|^2 + \|x - y_n\|^2)$$

(motivated by the parallelogram law), and the definition of α as an infimum; notice that $\|x - \frac{y_m+y_n}{2}\| \geq \alpha$, which forces $\limsup_{m,n\to\infty} \|y_m - y_n\|^2 = 0$.

Since H is complete, there must exist a point $y \in C$ at which $(y_n)_n$ converges. Then necessarily $d(x, y) = d(x, C)$. The uniqueness of y with this property follows again from the parallelogram law. If y' is another point of C such that $d(x, y') = d(x, C)$ then

$$\|y - y'\|^2 + 4\left\|x - \frac{y + y'}{2}\right\|^2 = 2(\|x - y\|^2 + \|x - y'\|^2)$$

which gives us $\|y - y'\|^2 \leq 0$, a contradiction since it was assumed that the points y and y' are distinct. ■

The map $P_C \colon x \to P_C(x)$, from H into itself, is called the *orthogonal projection* associated to C. Clearly,

$$P_C(x) \in C \quad \text{for every } x \in H$$

and

$$P_C(x) = x \quad \text{if and only if } x \in C.$$

In particular,

$$P_C^2 = P_C.$$

P_C is also *monotone*, that is,

$$\langle P_C(x) - P_C(y), x - y\rangle \geq 0 \quad \text{for all } x, y \in H. \tag{1.1}$$

This follows by adding the inequalities

$$\|x - P_C(x)\|^2 \leq \|x - P_C(y)\|^2 \text{ and } \|y - P_C(y)\|^2 \leq \|y - P_C(x)\|^2$$

after replacing the norm by the inner product.

If C is a closed subspace of the Hilbert space H, then P_C is a linear self-adjoint projection and $x - P_C(x)$ is orthogonal on each element of C. This fact is basic for the entire theory of orthogonal decompositions.

It is important to reformulate Theorem 1.2.1 in the framework of approximation theory. Suppose that C is a nonempty closed subset in a real linear normed space E. We define the *set of best approximation* from $x \in E$ to C as the set $\mathcal{P}_C(x)$ of all points in C closest to x, that is,

$$\mathcal{P}_C(x) = \{z \in C : d(x, C) = \|x - z\|\}.$$

We say that C is a *Chebyshev set* if $\mathcal{P}_C(x)$ is a singleton for all $x \in E$, and a *proximinal set* if all the sets $\mathcal{P}_C(x)$ are nonempty. Theorem 1.2.1 asserts that all nonempty closed convex sets in a Hilbert space are Chebyshev sets. There is an analogue of this theorem valid for the spaces $L^p(\mu)$ ($1 < p < \infty$, $p \neq 2$), saying that all such sets are proximinal. See the end of Section 2.3.

Clearly, the Chebyshev sets are closed.

The following result is a partial converse to Theorem 1.2.1:

1.2.2. Theorem (L. N. H. Bunt). *Every Chebyshev subset of $\mathbb{R}^n$ is convex.*

See R. Webster [45, pp. 362–365] for a proof based on Brouwer's fixed point theorem. Proofs based on the differentiability properties of the function $d_C : x \to d(x, C)$, are available in the paper by J.-B. Hiriart-Urruty [23], and in the monograph by L. Hörmander [26, pp. 62–63].

V. Klee raised the question whether Theorem 1.2.2 is valid for all real Hilbert spaces. The answer is known to be positive for all Chebyshev sets C such that the map d_C^2 is differentiable. See [23] for details (and an account of Klee's problem).

Outside Hilbert spaces, Klee's problem has a negative answer. For example, let $\ell^\infty(2, \mathbb{R})$ be the space $\mathbb{R}^2$ endowed with the sup norm, $\|(x_1, x_2)\| = \sup\{|x_1|, |x_2|\}$, and let C be the set of all vectors (x_1, x_2) such that $x_2 \geq x_1 \geq 0$. Then C is a nonconvex Chebyshev set.

1.3 The Hahn–Banach Extension Theorem

The Hahn–Banach theorem is a deep result in functional analysis which provides important consequences to convex function theory. We recall it here for the convenience of the reader.

Throughout, E will denote a real linear space.

A functional $p : E \to \mathbb{R}$ is *subadditive* if $p(x+y) \leq p(x)+p(y)$ for all $x, y \in E$; p is *positively homogeneous* if $p(\lambda x) = \lambda p(x)$ for each $\lambda \geq 0$ and each x in E; p is *sublinear* if it has both the above properties. A sublinear functional p is a seminorm if $p(\lambda x) = |\lambda| p(x)$ for all scalars. Finally, a seminorm p is a norm if

$$p(x) = 0 \implies x = 0.$$

If p is a sublinear functional, then $p(0) = 0$ and $-p(-x) \leq p(x)$. If p is a seminorm, then $p(x) \geq 0$ for all x in E and $\{x : p(x) = 0\}$ is a linear subspace of E.

1.3.1. Theorem (The Hahn–Banach theorem). *Let p be a sublinear functional on E, let E_0 be a linear subspace of E, and let $f_0 \colon E_0 \to \mathbb{R}$ be a linear functional dominated by p, that is, $f_0(x) \leq p(x)$ for all $x \in E_0$. Then f_0 has a linear extension f to E which is also dominated by p.*

This is an application of Zorn's lemma. See [10] for details.

1.3.2. Corollary. *If p is a sublinear functional on a real linear space E, then for every element $x_0 \in E$ there exists a linear functional $f \colon E \to \mathbb{R}$ such that $f(x_0) = p(x_0)$ and $f(x) \leq p(x)$ for all x in E.*

Proof. Take $E_0 = \{\lambda x_0 : \lambda \in \mathbb{R}\}$ and $f_0(\lambda x_0) = \lambda p(x_0)$ in Theorem 1.3.1. ■

The continuity of a linear functional on a topological linear space E means that it is bounded in a neighborhood of the origin. We shall denote by E' the *dual space* of E that is, the space of all continuous linear functionals on E.

In the context of normed linear spaces, the remark above allows us to define the norm of a continuous linear functional $f \colon E \to \mathbb{R}$ by the formula

$$\|f\| = \sup_{\|x\| \leq 1} |f(x)|.$$

With respect to this norm, the dual space of a normed linear space is always complete.

It is worth noting the following variant of Theorem 1.3.1 in the context of real normed linear spaces:

1.3.3. Theorem. *Let E_0 be a linear subspace of the normed linear space E, and let $f_0 \colon E_0 \to \mathbb{R}$ be a continuous linear functional. Then f_0 has a continuous linear extension f to E, with $\|f\| = \|f_0\|$.*

1.3.4. Corollary. *If E is a normed linear space, then for each $x_0 \in E$ with $x_0 \neq 0$ there exists a continuous linear functional $f \colon E \to \mathbb{R}$ such that $f(x_0) = \|x_0\|$ and $\|f\| = 1$.*

1.3.5. Corollary. *If E is a normed linear space and x is an element of E such that $f(x) = 0$ for all f in the dual space of E, then $x = 0$.*

The *weak topology* on E is the locally convex topology associated to the family of seminorms

$$p_F(x) = \sup\{|f(x)| : f \in F\},$$

where F runs over all nonempty finite subsets of E'. A sequence $(x_n)_n$ converges to x in the weak topology (abbreviated, $x_n \xrightarrow{w} x$) if and only if $f(x_n) \to f(x)$ for every $f \in E'$. When $E = \mathbb{R}^n$ this is the coordinate-wise convergence and agrees with the norm convergence. In general, the norm function is only *weakly lower semicontinuous*, that is,

$$x_n \xrightarrow{w} x \implies \|x\| \leq \liminf_{n \to \infty} \|x_n\|.$$

By Corollary 1.3.5 it follows that E' *separates* E in the sense that

$$x, y \in E \quad \text{and} \quad f(x) = f(y) \quad \text{for all } f \in E' \Longrightarrow x = y.$$

As a consequence we infer that the weak topology is separated (equivalently, Hausdorff).

For E' we can speak of the normed topology, of the weak topology (associated to $E'' = (E')'$) and also of the *weak-star topology*, which is associated to the family of seminorms p_F defined as above, with the difference that F runs over all nonempty finite subsets of E. The weak-star topology on E' is separated.

A net $(f_i)_{i\in I}$ (over some directed set I) converges to f in the weak-star topology (abbreviated, $f_i \overset{weak^\star}{\to} f$) if and only if $f_i(x) \to f(x)$ for all $x \in E$.

1.3.6. Theorem (The Banach–Alaoglu theorem). *If E is a normed linear space, then the closed unit ball of its dual space is compact in the weak-star topology. Consequently, each net of points of this ball has a converging subnet.*

See [10, p. 47] for details.

When E is a separable normed linear space, the closed unit ball of E' is also a metrizable space in the weak-star topology (and in this case dealing with sequences suffices as well). We come to the separability situation very often, by replacing E with a subspace generated by a suitable sequence of elements.

1.3.7. Remark. According to the Banach–Alaoglu theorem, if E is a normed linear space, then each weak-star closed subset of the closed unit ball of the dual of E is weak-star compact. This is a big source of compact convex sets in mathematics. For example, so is the set $\mathrm{Prob}\,(X)$, of all Borel probability measures on a compact Hausdorff space X. These are the regular σ-additive measures μ on the Borel subsets of X with $\mu(X) = 1$. The Riesz–Kakutani representation theorem (see [22, p. 177]) allows us to identify $\mathrm{Prob}\,(X)$ with the following weak-star closed subset of norm-1 functionals of $C(X)'$:

$$K = \{L : L \in C(X)',\ L(1) = 1 = \|L\|\}.$$

Notice that K consists of *positive functionals*, that is,

$$f \in C(X), \quad f \geq 0 \text{ implies } L(f) \geq 0.$$

In fact, if the range of f is included in $[0, 2r]$, then $\|f - r\| \leq r$, so that $r \geq |L(f - r)| = |L(f) - r|$, that is, $L(f) \in [0, 2r]$.

Corollary 1.3.4 yields an important canonical embedding of each normed linear space E into its second dual E'':

$$J_E \colon E \to E'', \quad J_E(x)(x') = x'(x).$$

One can easily show that J_E is a linear isometry.

A Banach space E is said to be *reflexive* if J_E is onto (that is, if E is isometric with its second dual through J_E). Besides the finite dimensional Banach spaces, other examples of reflexive Banach spaces are Hilbert spaces and the spaces $L^p(\mu)$ for $1 < p < \infty$. One can easily prove the following permanence properties:

(R1) every closed subspace of a reflexive space is reflexive;

(R2) the dual of a reflexive space is also a reflexive space;

(R3) reflexivity preserves under renorming by an equivalent norm.

Property (R3) is a consequence of the following characterization of reflexivity:

1.3.8. Theorem (The Eberlein–Šmulyan theorem). *A Banach space E is reflexive if and only if every bounded sequence of elements of E admits a weakly converging subsequence.*

Proof. The necessity part follows from the Banach–Alaoglu theorem (Theorem 1.3.6). In fact, we may restrict ourselves to the case where E is also separable. The sufficiency part follows from the remark that J_E maps the closed unit ball of E into a weak–star dense (and also weak-star closed) subset of the closed unit ball of E''. Full details are available in books such as those by H. W. Alt [2], J. B. Conway [8] or M. M. Day [10]. ■

1.4 Hyperplanes and Separation Theorems

The notion of a hyperplane represents a natural generalization of the notion of a line in $\mathbb{R}^2$ or a plane in $\mathbb{R}^3$. Hyperplanes are useful to split the whole space into two pieces (called half-spaces).

A *hyperplane* in a real linear space E is any set of constancy of a nonzero linear functional. In other words, a hyperplane is a set of the form

$$H = \{x \in E : h(x) = \alpha\}, \tag{1.2}$$

where $h\colon E \to \mathbb{R}$ is a suitable nonzero linear functional and α is a suitable scalar. In this case the sets

$$\{x \in E : h(x) \leq \alpha\} \quad \text{and} \quad \{x \in E : h(x) \geq \alpha\}$$

are called the *half-spaces* determined by H. We say that H *separates* two sets U and V if they lie in opposite half-spaces (and *strictly separates* U and V if one set is contained in $\{x \in E : h(x) < \alpha\}$ and the other in $\{x \in E : h(x) \geq \alpha\}$).

When E is a real normed linear space (or, more generally, a locally convex Hausdorff space) and the functional h which appears in the representation formula (1.2) is continuous (that is, when h belongs to the dual space E') we say that the corresponding hyperplane H is *closed*. In the context of $\mathbb{R}^n$, all linear functionals are continuous and thus all hyperplanes are closed. In fact, any linear functional $h\colon \mathbb{R}^n \to \mathbb{R}$ has the form $h(x) = \langle x, z\rangle$, for some $z \in \mathbb{R}^n$ (uniquely determined by h).

Some authors define the hyperplanes as the maximal proper affine subsets H of E. Here *proper* means different from E. One can prove that the hyperplanes are precisely the translates of codimension-1 linear subspaces, and this explains the agreement of the two definitions.

The following results on the separation of convex sets by closed hyperplanes are part of a very general theory that can be found in [10].

1.4.1. Theorem (Separation theorem). *Let U and V be two convex sets in a real locally convex Hausdorff space E, with $\operatorname{int} U \neq \emptyset$ and $V \cap \operatorname{int} U = \emptyset$. Then there exists a closed hyperplane that separates U and V.*

1.4.2. Theorem (Strong separation theorem). *Let K and C be two disjoint nonempty convex sets in a real locally convex Hausdorff space E, with K compact and C closed. Then there exists a closed hyperplane that separates strictly K and C.*

The special case of the last result when K is a singleton is known as the *basic separation theorem.*

Proof of Theorem 1.4.2 in the case where $E = \mathbb{R}^n$: We start by noticing that the distance

$$d = \inf\{\|x - y\| : x \in K,\ y \in C\}$$

is attained for a pair $x_0 \in K$, $y_0 \in C$. The hyperplane through x_0, orthogonal to the linear segment $[x_0, y_0]$, determined by x_0 and y_0, has the equation $\langle y_0 - x_0, z - x_0\rangle = 0$. Fix arbitrarily a point $x \in K$. Then $\langle y_0 - x_0, z - x_0\rangle \leq 0$ for every point $z \in [x_0, x]$ (and thus for every $z \in K$). Consequently, every hyperplane through any point inside the segment $[x_0, y_0]$, orthogonal to this segment, separates strictly K and C. ■

We are now in a position to infer the finite dimensional case of Theorem 1.4.1 from Theorem 1.4.2. In fact, it suffices to assume that both sets U and V are closed. Then choose a point $x_0 \in \operatorname{int} U$ and apply the preceding result to V and to the compact set

$$K_n = \{x_0 + (1 - 1/n)(x - x_0) : x \in U\} \cap \overline{B}_n(0)$$

for $n \in \mathbb{N}^*$. This gives us a sequence of unit vectors u_n and numbers α_n such that $\langle u_n, x\rangle \leq \alpha_n$ for $x \in K_n$ and $\langle u_n, y\rangle \geq \alpha_n$ for $y \in V$. As $(u_n)_n$ and $(\alpha_n)_n$ are bounded, they admit converging subsequences, say to u and α respectively. Now it is easy to conclude that $H = \{z : \langle u, z\rangle = \alpha\}$ is the desired separation hyperplane.

The *closed convex hull* of a subset A of a locally convex Hausdorff space E is the smallest closed convex set $\overline{\mathrm{co}}\,(A)$ containing A (that is, the intersection of all closed convex sets containing A). From Theorem 1.4.2 we can infer the following result on the support of closed convex sets:

1.4.3. Proposition. *If A is a nonempty subset of a real locally convex Hausdorff space E, then the closed convex hull $\overline{\mathrm{co}}\,(A)$ is the intersection of all the closed half-spaces containing A. Equivalently,*

$$\overline{\mathrm{co}}\,(A) = \bigcap_{f \in E'} \{x : f(x) \leq \sup_{y \in A} f(y)\}.$$

This proposition implies:

1.4.4. Corollary. *In a real locally convex Hausdorff space E, the closed convex sets and the weakly closed convex sets are the same.*

Let C be a nonempty subset of $\mathbb{R}^n$. The *polar* set of C, is the set

$$C^\circ = \{x \in \mathbb{R}^n : \langle u, x\rangle \leq 1 \quad \text{for every } u \in C\}.$$

Clearly, C° is a closed convex set containing 0 and $C \subset D$ implies $D^\circ \subset C^\circ$. By the basic separation theorem we infer that

$$C^{\circ\circ} = \overline{\text{co}}\,(C \cup \{0\})$$

(which is known as the bipolar theorem).

Next we introduce the notion of a *supporting hyperplane* to a convex set U in a normed linear space E.

1.4.5. Definition. We say that the hyperplane H *supports* U at a point a in U if $a \in H$ and U is contained in one of the half-spaces determined by H.

Theorem 1.4.1 assures the existence of a supporting hyperplane to any convex set U at a boundary point, provided that U has nonempty interior.

When $E = \mathbb{R}^n$, the existence of a supporting hyperplane of U at a boundary point a will mean the existence of a vector $z \in \mathbb{R}^n$ and of a real number α such that

$$\langle a, z\rangle = \alpha \;\text{ and }\; \langle x, z\rangle \leq \alpha \;\text{ for all } x \in U.$$

A direct argument for the existence of a supporting hyperplane in the finite dimensional case is as follows: We may assume that U is closed, by replacing U with $\overline{U}$. Choose a point $x_0 \in S_1(a) = \{x : \|x - a\| = 1\}$ such that

$$d(x_0, U) = \sup\{d(x, U) : x \in S_1(a)\},$$

that is, x_0 is the farthest point from U. The point a is the point of U closest to x_0 and we may conclude that the hyperplane $H = \{z : \langle x_0 - a, z - a\rangle = 0\}$ supports U at a.

1.5 The Krein–Milman Theorem

This section is devoted to a discussion on the geometry of convex sets.

1.5.1. Definition. Let U be a convex subset of a linear space E. A point z in U is an *extreme point* if it is not an interior point of any linear segment in U, that is, if there do not exist distinct points $x, y \in U$ and numbers $\lambda \in (0, 1)$ such that

$$z = (1 - \lambda)x + \lambda y.$$

The extreme points of a triangle are its vertices. More generally, every polytope $A = \text{co}\,\{a_0, \ldots, a_m\}$ has finitely many extreme points, and they are among the points $a_0, \ldots, a_m$.

All boundary points of a disc $\overline{D}_R(0) = \{(x,y) : x^2 + y^2 \leq R^2\}$ are extreme points; this is an expression of the rotundity of discs. The closed upper half-plane $y \geq 0$ in $\mathbb{R}^2$ has no extreme point.

The extreme points are the landmarks of compact convex sets in $\mathbb{R}^n$:

1.5.2. Theorem (H. Minkowski). *Every nonempty convex and compact subset K of $\mathbb{R}^n$ is the convex hull of its extreme points.*

Proof. We use induction on the dimension m of K. If $m = 0$ or $m = 1$, that is, when K is a point or a closed segment, the above statement is obvious. Assume the theorem is true for all compact convex sets of dimension at most $m \leq n-1$. Consider now a compact convex set K whose dimension is $m+1$ and embed it into a linear subspace E of dimension $m+1$.

If z is a boundary point of K, then we can choose a supporting hyperplane $H \subset E$ for K through z. The set $K \cap H$ is compact and convex and its dimension is less or equal to m. By the induction hypothesis, z is a convex combination of extreme points of $K \cap H$. Or, any extreme point e of $K \cap H$ is also an extreme point of K. In fact, letting $H = \{t \in E : \varphi(t) = \alpha\}$, we may assume that K is included in the half-space $\varphi(t) \leq \alpha$. If $e = (1-\lambda)x + \lambda y$ with $x \neq y$ in K and $\lambda \in (0,1)$, then necessarily $\varphi(x) = \varphi(y) = \alpha$, that is, x and y should be in $K \cap H$, in contradiction with the choice of e.

If z is an interior point of K, then each line through z intersects K in a segment whose endpoints belong to the boundary of K. Consequently, z is a convex combination of boundary points that in turn are convex combinations of extreme points. This ends the proof. ∎

The result of Theorem 1.5.2 can be made more precise: *every point in a compact convex subset K of $\mathbb{R}^n$ is the convex combination of at most $n+1$ extreme points*. See Theorem 1.1.2.

Theorem 1.5.2 admits a remarkable generalization to the setting of locally convex spaces.

1.5.3. Theorem. *Let E be a real locally convex Hausdorff space and K be a nonempty compact convex subset of E. If U is an open convex subset of K such that* $\operatorname{Ext} K \subset U$, *then $U = K$.*

Proof. Suppose that $U \neq K$ and consider the family $\mathcal{U}$ of all open convex sets in K which are not equal to K. By Zorn's lemma, each set $U \in \mathcal{U}$ is contained in a maximal element V of $\mathcal{U}$.

For each $x \in K$ and $t \in [0,1]$, let $\varphi_{x,t} \colon K \to K$ be the continuous map defined by $\varphi_{x,t}(y) = ty + (1-t)x$.

Assuming $x \in V$ and $t \in [0,1)$, we shall show that $\varphi_{x,t}^{-1}(V)$ is an open convex set which contains V properly, hence $\varphi_{x,t}^{-1}(V) = K$. In fact, this is clear when $t = 0$. If $t \in (0,1)$, then $\varphi_{x,t}$ is a homeomorphism and $\varphi_{x,t}^{-1}(V)$ is an open convex set in K. Moreover,

$$\varphi_{x,t}(\overline{V}) \subset V,$$

which yields $\overline{V} \subset \varphi_{x,t}^{-1}(V)$, hence $\varphi_{x,t}^{-1}(V) = K$ by the maximality of V. Therefore $\varphi_{x,t}(K) \subset V$. For any open convex set W in K the intersection $V \cap W$ is also open and convex, and the maximality of V yields that either $V \cup W = V$ or $V \cup W = K$. In conclusion $K \backslash V$ is precisely a singleton $\{e\}$. But such a point is necessarily an extreme point of K, which is a contradiction. ■

1.5.4. Corollary (Krein–Milman theorem). *Let K be a nonempty compact convex subset of a real locally convex Hausdorff space E. Then K is the closed convex hull of* $\operatorname{Ext} K$.

Proof. By Theorem 1.4.2, the set $L = \overline{\mathrm{co}}\,(\operatorname{Ext} K)$ is the intersection of all open convex sets containing L. If U is an open subset of K and $U \supset L$, then $U \supset \operatorname{Ext} K$. Hence $U = K$ and $L = K$. ■

The above proof of the Krein–Milman theorem yields the existence of extreme points as a consequence of the formula $K = \overline{\mathrm{co}}\,(\operatorname{Ext} K)$. However this can be checked directly. Call a subset A of K *extremal* if it is closed, nonempty and verifies the following property:

$$x, y \in K \;\text{ and }\; (1-\lambda)x + \lambda y \in A \;\text{ for some } \lambda \in (0,1) \implies x, y \in A.$$

By Zorn's lemma we can choose a minimal extremal subset, say S. We show that S is a singleton (which yields an extreme point of K). In fact, if S contains more than one point, the separation Theorem 1.4.2 proves the existence of a functional $f \in E'$ which is not constant on S. But in this case the set

$$S_0 = \{x \in S : f(x) = \sup_{y \in S} f(y)\}$$

will contradict the minimality of S. Now the formula $K = \overline{\mathrm{co}}\,(\operatorname{Ext} K)$ can easily be proved by noticing that the inclusion $\overline{\mathrm{co}}\,(\operatorname{Ext} K) \subset K$ cannot be strict.

It is interesting to note the following converse to Theorem 1.5.4:

1.5.5. Theorem (D. P. Milman). *Suppose that K is a compact convex set (in a locally convex Hausdorff space E) and C is a subset of K such that K is the closed convex hull of C. Then the extreme points of K are contained in the closure of C.*

Coming back to Theorem 1.5.2, the fact that every point x of a compact convex set K in $\mathbb{R}^n$ is a convex combination of extreme points of K,

$$x = \sum_{k=1}^{m} \lambda_k x_k,$$

can be reformulated as an integral representation,

$$f(x) = \sum_{k=1}^{m} \lambda_k f(x_k) = \int_{\operatorname{Ext} K} f \, d\mu \tag{1.3}$$

for all $f \in (\mathbb{R}^n)'$. Here $\mu = \sum_{k=1}^{m} \lambda_k \delta_{x_k}$ is a convex combination of Dirac measures δ_{x_k} and thus μ itself is a Borel probability measure on $\operatorname{Ext} K$.

The integral representation (1.3) can be extended to all Borel probability measures μ on a compact convex set K (in a locally convex Hausdorff space E). We shall need some definitions.

Given a Borel probability measure μ on K, and a Borel subset $S \subset K$, we say that μ *is concentrated on* S if $\mu(K \backslash S) = 0$. For example, a Dirac measure δ_x is concentrated on x.

A point $x \in K$ is said to be the *barycenter* of μ provided that

$$f(x) = \int_K f \, d\mu \quad \text{for all } f \in E'.$$

Since the functionals separate the points of E, the point x is uniquely determined by μ. With this preparation, we can reformulate the Krein–Milman theorem as follows:

1.5.6. Theorem. *Every point of a compact convex subset* K (of a locally convex Hausdorff space E), *is the barycenter of a Borel probability measure on* K, *which is supported by the closure of the extreme points of* K.

H. Bauer pointed out that the extremal points of K are precisely the points $x \in K$ for which the only Borel probability measure μ which admits x as a barycenter is δ_x. See [39, p. 6]. This fact together with Theorem 1.5.6 yields D. P. Milman's aforementioned converse of the Krein–Milman theorem. For an alternative argument see [10, pp. 103–104].

Theorem 1.5.6 led G. Choquet [6] to his theory on integral representation for elements of a closed convex cone. See [39] for details.

Chapter 2

Convex Functions on Banach Spaces

A central theme of applied mathematics is optimization, which involves minimizing or maximizing various quantities. In addition to the first and second derivative tests of one-variable calculus there is the powerful technique of Lagrange multipliers in several variables. The aim of this chapter is to develop analogues of these tests within the framework of convexity, which provides a valuable alternative to differentiability. Many applications in economics, business and other areas involve convex functions.

2.1 Convex Functions at First Glance

Convex functions are real-valued functions defined on convex sets.

In what follows U will be a convex set in a real linear space E.

2.1.1. Definition. A function $f\colon U \to \mathbb{R}$ is said to be *convex* if

$$f((1-\lambda)x + \lambda y) \leq (1-\lambda)f(x) + \lambda f(y) \tag{2.1}$$

for all $x, y \in U$ and all $\lambda \in [0,1]$.

The other related notions such as *concave function*, *affine function*, and *strictly convex function* can be introduced in a standard manner.

By mathematical induction we can extend the basic inequality (2.1) to the case of arbitrary convex combinations. We shall refer to this as the *discrete case of Jensen's inequality.*

Convexity, in the case of several variables, is equivalent with convexity on each line segment included in the domain of definition:

2.1.2. Proposition. *A function $f\colon U \to \mathbb{R}$ is convex if and only if for every two points x and y in U the function*

$$\varphi\colon [0,1] \to \mathbb{R}, \quad \varphi(t) = f((1-t)x + ty)$$

is convex.

Notice that convexity of functions in the several variables case means more than convexity in each variable separately; think of the case of the function $f(x,y) = xy$, $(x,y) \in \mathbb{R}^2$, which is not convex, though convex in each variable.

Some simple examples of strictly convex functions on $\mathbb{R}^n$ are as follows:

- $f(x_1,\ldots,x_n) = \sum_{k=1}^n \varphi(x_k)$, where φ is a strictly convex function on $\mathbb{R}$.
- $f(x_1,\ldots,x_n) = \sum_{i<j} c_{ij}(x_i - x_j)^2$, where the coefficients c_{ij} are positive.
- The distance function $d_U \colon \mathbb{R}^n \to \mathbb{R}$, $d_U(x) = d(x,U)$, associated to a nonempty convex set U in $\mathbb{R}^n$.

An important example of a concave function is given by

$$f\colon \operatorname{Sym}^{++}(n,\mathbb{R}) \to \mathbb{R}, \quad f(A) = \log(\det A).$$

Notice that $\operatorname{Sym}^{++}(n,\mathbb{R})$ is an open convex set in $\mathrm{M}_n(\mathbb{R})$. In order to prove that f is concave we have to recall the formula

$$\int_{\mathbb{R}^n} e^{-\langle Ax,x\rangle}\, dx = \pi^{n/2}/\sqrt{\det A}$$

for all A in $\operatorname{Sym}^{++}(n,\mathbb{R})$. Then, for all $A,B \in \operatorname{Sym}^{++}(n,\mathbb{R})$ and all $\alpha \in (0,1)$ we have

$$\int_{\mathbb{R}^n} e^{-\langle [\alpha A+(1-\alpha)B]x,x\rangle}\, dx \le \left(\int_{\mathbb{R}^n} e^{-\langle Ax,x\rangle}\, dx\right)^{\alpha} \left(\int_{\mathbb{R}^n} e^{-\langle Bx,x\rangle}\, dx\right)^{1-\alpha},$$

by the Rogers–Hölder inequality. This yields the log-concavity of the function det:

$$\det(\alpha A + (1-\alpha)B) \ge (\det A)^{\alpha}(\det B)^{1-\alpha}.$$

It is worth to extend this last formula to the setting of positive matrices (e.g., by using perturbations of the form $A + \varepsilon I$ and $B + \varepsilon I$).

Remarks (Getting new examples from old ones). i) Suppose that $\varphi_1,\ldots,\varphi_n$ are convex functions defined on the same convex set D in $\mathbb{R}^n$ and $f\colon \mathbb{R}^n \to \mathbb{R}$ is a nondecreasing convex function. Then $F(x) = f(\varphi_1(x),\ldots,\varphi_n(x))$ is convex on D. Here "nondecreasing" means nondecreasing in each variable (when the others are kept fixed).

ii) The limit of any pointwise converging sequence of convex functions is a convex function.

iii) Let $(f_\alpha)_\alpha$ be a family of convex functions defined on the same convex set U, such that $f(x) = \sup_\alpha f_\alpha(x) < \infty$ for all $x \in U$. Then f is convex.

We shall next discuss several connections between convex functions and convex sets.

By definition, the *epigraph* of a function $f\colon U \to \mathbb{R}$ is the set

$$\operatorname{epi}(f) = \{(x, y) : x \in U,\ y \in \mathbb{R} \text{ and } f(x) \leq y\}.$$

It is easy to verify that $f\colon U \to \mathbb{R}$ is convex if and only if $\operatorname{epi}(f)$ is convex in $E \times \mathbb{R}$. This shows that the *theory of convex functions can be subordinated to the theory of convex sets*.

A practical implication is the existence of supporting hyperplanes for convex functions. To make this more precise, we shall pass to the topological context, where U is an open convex set in a linear normed space E and $f\colon U \to \mathbb{R}$ is a continuous convex function. In this case, $\operatorname{epi}(f)$ has a *nonempty* interior in $E \times \mathbb{R}$ and every point $(a, f(a))$ is a boundary point for $\operatorname{epi}(f)$. According to Theorem 1.4.1, there is a closed hyperplane H in $E \times \mathbb{R}$ that contains $(a, f(a))$ and $\operatorname{epi}(f)$ is contained in one of the half-spaces determined by H. We call this a *supporting hyperplane* to f at a.

The closed hyperplanes H are associated to nonzero continuous linear functionals on $E \times \mathbb{R}$ and the dual space of $E \times \mathbb{R}$ is constituted of all pairs (h, λ), where h is any continuous linear functional on E and λ is any real number. Consequently a supporting hyperplane to f at a is determined by a pair (h, λ) and a real number α such that

$$h(a) + \lambda f(a) = \alpha$$

and

$$h(x) + \lambda y \geq \alpha \text{ for all } y \geq f(x) \quad \text{and all } x \in U.$$

Notice that $\lambda \neq 0$, since otherwise $h(x) \geq h(a)$ for x in a ball $B_r(a)$, which forces $h = 0$. A moment's reflection shows that actually $\lambda > 0$ and thus we are led to the existence of a continuous linear functional h such that

$$f(x) \geq f(a) + h(x - a) \quad \text{for every } x \in U.$$

We call h a *support* of f at a. Supports are instrumental in defining the concept of subdifferential. See Section 2.4.

We pass now to another connection between convex functions and convex sets.

Given a function $f\colon U \to \mathbb{R}$ and a scalar α, the *sublevel set* L_α of f at height α is the set

$$L_\alpha = \{x \in U : f(x) \leq \alpha\}.$$

2.1.3. Lemma. *Each sublevel set of a convex function is a convex set.*

The property of Lemma 2.1.3 characterizes the quasiconvex functions. Recall that a function $f\colon U \to \mathbb{R}$ defined on a convex set U is said to be *quasiconvex* if

$$f((1 - \lambda)x + \lambda y) \leq \sup\{f(x), f(y)\}$$

for all $x, y \in U$ and all $\lambda \in [0, 1]$.

Convex functions exhibit a series of nice properties related to maxima and minima, which make them important in theoretical and applied mathematics.

2.1.4. Theorem. *Assume that U is a convex subset of a normed linear space E. Then any local minimum of a convex function $f: U \to \mathbb{R}$ is also a global minimum. Moreover, the set of global minimizers of f is convex.*

If f is strictly convex in a neighborhood of a minimum point, then the minimum point is unique.

Proof. If a is a local minimum, then for each $x \in U$ there is an $\varepsilon > 0$ such that

$$\begin{aligned} f(a) &\leq f(a + \varepsilon(x - a)) = f((1 - \varepsilon)a + \varepsilon x) \\ &\leq (1 - \varepsilon)f(a) + \varepsilon f(x). \end{aligned} \tag{2.2}$$

This yields $f(a) \leq f(x)$, so a is a global minimum. If f is strictly convex in a neighborhood of a, then the last inequality in (2.2) is strict and the conclusion becomes $f(x) > f(a)$ for all $x \in U$, $x \neq a$. The second assertion is a consequence of Lemma 2.1.3. ■

The following result gives us a useful condition for the existence of a global minimum:

2.1.5. Theorem. *Assume that U is an unbounded closed convex set in $\mathbb{R}^n$ and $f: U \to \mathbb{R}$ is a continuous convex function whose sublevel sets are bounded. Then f has a global minimum.*

Proof. Notice that all sublevel sets L_α of f are bounded and closed (and thus compact in $\mathbb{R}^n$). Then every sequence of elements in a sublevel set has a converging subsequence and this yields immediately the existence of global minimizers. ■

Under the assumptions of Theorem 2.1.5, the condition on boundedness of sublevel sets is equivalent with the following growth condition:

$$\liminf_{\|x\| \to \infty} \frac{f(x)}{\|x\|} > 0. \tag{2.3}$$

The sufficiency part is clear. For the necessity part, reason by *reductio ad absurdum* and choose a sequence $(x_k)_k$ in U such that $\|x_k\| \to \infty$ and $f(x_k) \leq \|x_k\|/k$. Since the level sets are supposed to be bounded we have $\|x_k\|/k \to \infty$, and this leads to a contradiction. Indeed, for every $x \in U$ the sequence

$$x + \frac{k}{\|x_k\|}\,(x_k - x)$$

is unbounded though lies in some sublevel set $L_{f(x)+\varepsilon}$, with $\varepsilon > 0$.

The functions which verify the condition (2.3) are said to be *coercive*. Clearly, coercivity implies

$$\lim_{\|x\| \to \infty} f(x) = \infty.$$

Convex functions attain their maxima at the boundary:

2.1.6. Theorem (The maximum principle). *If f is a convex function on a convex subset U of a normed linear space E and attains a global maximum at an interior point of U, then f is constant.*

Proof. Assume that f is not constant and attains a global maximum at the point $a \in \operatorname{int} U$. Choose $x \in U$ such that $f(x) < f(a)$ and $\varepsilon \in (0,1)$ such that $y = a + \varepsilon(a - x) \in U$. Then $a = y/(1+\varepsilon) + \varepsilon x/(1+\varepsilon)$, which yields a contradiction since

$$f(a) \leq \frac{1}{1+\varepsilon} f(y) + \frac{\varepsilon}{1+\varepsilon} f(x) < \frac{1}{1+\varepsilon} f(a) + \frac{\varepsilon}{1+\varepsilon} f(a) = f(a). \quad \blacksquare$$

We end this section with an important consequence of Theorem 2.1.6.

2.1.7. Theorem. *If f is a continuous convex function on a compact convex subset K of $\mathbb{R}^n$, then f attains a global maximum at an extreme point.*

Proof. Assume that f attains its global maximum at $a \in K$. By Theorem 2.1.6, the point a can be represented as a convex combination of extreme points, say $a = \sum_{k=1}^{m} \lambda_k e_k$. Then $f(a) \leq \sum_{k=1}^{m} \lambda_k f(e_k) \leq \sup_k f(e_k)$, which forces $f(a) = f(e_k)$ for some k. $\blacksquare$

For functions defined on n-dimensional intervals $[a_1, b_1] \times \cdots \times [a_n, b_n]$ in $\mathbb{R}^n$, Theorem 2.1.7 extends to the case of continuous functions which are convex in each variable (when the others are kept fixed). This fact can be proved by one-variable means (taking into account that convex functions of one real variable attain their supremum at the end points). Here is an example: Find the maximum of the function

$$f(a,b,c) = \left[3(a^5 + b^7 \sin \frac{\pi a}{2} + c) - 2(bc + ca + ab)\right]$$

for $a, b, c \in [0,1]$. The answer is 4, by noticing that

$$f(a,b,c) \leq \sup[3(a+b+c) - 2(bc+ca+ab)] = 4.$$

In the infinite dimensional setting, it is difficult to state fairly general results on maximum-attaining. Besides, the deep results of Banach space theory appears to be crucial in answering questions which at first glance may look simple. Here is an example. By the Eberlein–Šmulyan theorem it follows that each continuous linear functional on a reflexive Banach space E achieves its norm on the unit ball. Surprisingly, these are the *only* Banach spaces for which the norm-attaining phenomenon occurs. This was proved by R. C. James (see [10, p. 63]).

2.2 Continuity of Convex Functions

It is well known that a convex function defined on an open interval is continuous. In what follows the same is true for convex functions defined on open convex sets in $\mathbb{R}^n$. The basic remark refers to a local property of convex functions.

2.2.1. Lemma. *Every convex function f defined on an open convex set U in $\mathbb{R}^n$ is locally bounded* (*that is*, *each $a \in U$ has a neighborhood on which f is bounded*).

Proof. For $a \in U$ arbitrarily fixed, choose a cube K in U, centered at a, with vertices $v_1, \ldots, v_{2^n}$. Clearly, K is a neighborhood of a. Every $x \in K$ is a convex combination of vertices and thus

$$f(x) = f(\sum_{k=1}^{2^n} \lambda_k v_k) \leq M = \sup_{1 \leq k \leq 2^n} f(v_k),$$

so f is bounded above on K. By the symmetry of K, for every $x \in K$ there is a $y \in K$ such that $a = (x+y)/2$. Then $f(a) \leq (f(x) + f(y))/2$, which yields $f(x) \geq 2f(a) - f(y) \geq 2f(a) - M$, and the proof is complete. ■

2.2.2. Proposition. *Let f be a convex function on an open convex set U in $\mathbb{R}^n$. Then f is locally Lipschitz. In particular, f is continuous on U.*

According to a classical theorem due to Rademacher (see [14]), we can infer from Proposition 2.2.2 that every convex function on an open convex set U in $\mathbb{R}^n$ is almost everywhere differentiable. A direct proof will be given in Section 2.5 (see Theorem 2.5.3).

Proof. According to the preceding lemma, given $a \in U$, we may find a ball $B_{2r}(a) \subset U$ on which f is bounded above, say by M. For $x \neq y$ in $B_r(a)$, put $z = y + (r/\alpha)(y - x)$, where $\alpha = \|y - x\|$. Clearly, $z \in B_{2r}(a)$. As

$$y = \frac{r}{r+\alpha}\, x + \frac{\alpha}{r+\alpha}\, z,$$

from the convexity of f we infer that

$$f(y) \leq \frac{r}{r+\alpha} f(x) + \frac{\alpha}{r+\alpha} f(z).$$

Then

$$\begin{aligned} f(y) - f(x) &\leq \frac{\alpha}{r+\alpha}[f(z) - f(x)] \\ &\leq \frac{\alpha}{r}\,[f(z) - f(x)] \leq \frac{2M}{r}\,\|y - x\| \end{aligned}$$

and the proof ends by interchanging the roles of x and y. ■

2.2.3. Corollary. *Let f be a convex function defined on a convex set A in $\mathbb{R}^n$. Then f is Lipschitz on each compact convex subset of* $\operatorname{ri}(A)$ (*and thus f is continuous on* $\operatorname{ri}(A)$).

Proof. Clearly, we may assume that $\operatorname{aff}(A) = \mathbb{R}^n$. In this case, $\operatorname{ri}(A) = \operatorname{int}(A)$ and Proposition 2.2.2 applies. ■

The infinite dimensional analogue of Proposition 2.2.2 is as follows:

2.2.4. Proposition. *Let f be a convex function on an open convex set U in a normed linear space. If f is bounded above in a neighborhood of one point of U, then f is locally Lipschitz on U. In particular, f is a continuous function.*

The proof is similar with that of Proposition 2.2.2, with the difference that the role of Lemma 2.2.1 is taken by the following lemma:

2.2.5. Lemma. *Let f be a convex function on an open convex set U in a normed linear space. If f is bounded above in a neighborhood of one point of U, then f is locally bounded on U.*

Proof. Suppose that f is bounded above by M on a ball $B_r(a)$. Let $x \in U$ and choose $\rho > 1$ such that $z = a + \rho(x - a) \in U$. If $\lambda = 1/\rho$, then

$$V = \{v : v = (1 - \lambda)y + \lambda z,\ y \in B_r(a)\}$$

is a neighborhood of $x = (1 - \lambda)a + \lambda z$, with radius $(1 - \lambda)r$. Moreover, for $v \in V$ we have

$$f(v) \leq (1 - \lambda)f(y) + \lambda f(z) \leq (1 - \lambda)M + \lambda f(z).$$

To show that f is bounded below in the same neighborhood, choose arbitrarily $v \in V$ and notice that $2x - v \in V$. Consequently, $f(x) \leq f(v)/2 + f(2x - v)/2$, which yields $f(v) \geq 2f(x) - f(2x - v) \geq 2f(x) - M$. ■

A convex function on an infinite dimensional Banach space E is not necessarily continuous. Actually, one can prove that the only Banach spaces E such that every convex function $f : E \to \mathbb{R}$ is continuous are the finite dimensional ones. This is a consequence of the well-known fact that the norm and the weak topology agree only in the finite dimensional case. See [10, Lemma 1, p. 45].

In applications it is often useful to consider extended real-valued functions, defined on a real linear space E.

2.2.6. Definition. A function $f : E \to \overline{\mathbb{R}}$ is said to be *convex* if its epigraph,

$$\operatorname{epi}(f) = \{(x, y) : x \in E,\ y \in \mathbb{R} \text{ and } f(x) \leq y\}$$

is a convex subset of $E \times \mathbb{R}$.

The *effective domain* of a convex function $f : E \to \overline{\mathbb{R}}$ is the set

$$\operatorname{dom} f = \{x : f(x) < \infty\}.$$

Clearly, this is a convex set. Most of the time we shall deal with *proper convex functions*, that is, with convex functions $f : E \to \mathbb{R} \cup \{\infty\}$ which are not identically ∞. In their case, the property of convexity can be reformulated in more familiar terms,

$$f((1 - \lambda)x + \lambda y) \leq (1 - \lambda)f(x) + \lambda f(y)$$

for all $x, y \in E$ and all $\lambda \in [0, 1]$ for which the right hand side is finite.

If U is a convex subset of E, then every convex function $f : U \to \mathbb{R}$ extends to a proper convex function $\widetilde{f}$ on E, letting $\widetilde{f}(x) = \infty$ for $x \in E \backslash U$. Another basic example is related to the indicator function. The *indicator function* of a nonempty subset A is defined by the formula

$$\delta_A(x) = \begin{cases} 0 & \text{if } x \in A \\ \infty & \text{if } x \in E \backslash A. \end{cases}$$

Clearly, A is convex if and only if δ_A is a proper convex function.

The sublevel sets of a proper convex function $f\colon E \to \mathbb{R} \cup \{\infty\}$ are convex sets. A discussion of the topological nature of the sublevel sets needs the framework of lower semicontinuity.

2.2.7. Definition. An extended real-valued function f defined on a Hausdorff topological space X is called *lower semicontinuous* if

$$f(x) = \liminf_{y \to x} f(y) \quad \text{for all } x \in X.$$

In the same framework, a function g is called *upper semicontinuous* if $-g$ is lower semicontinuous.

The lower semicontinuous functions are precisely the functions for which all sublevel sets are closed. An important remark is that the supremum of any family of lower semicontinuous proper convex functions is a function of the same nature.

If the effective domain of a proper convex function is closed and f is continuous relative to $\operatorname{dom} f$, then f is lower semicontinuous. However, f can be lower semicontinuous without its effective domain being closed. The following function,

$$\varphi(x, y) = \begin{cases} y^2/2x & \text{if } x > 0, \\ \alpha & \text{if } x = y = 0, \\ \infty & \text{otherwise}, \end{cases}$$

is illustrative on what can happen at the boundary points of the effective domain. In fact, f is a proper convex function for each $\alpha \in [0, \infty]$. All points of its effective domain are points of continuity except the origin, where the limit does not exist. The function φ is lower semicontinuous for $\alpha = 0$.

2.3 Positively Homogeneous Functions

Many of the functions which arise naturally in convex analysis are real-valued functions f defined on a convex cone C in $\mathbb{R}^n$ (often $\mathbb{R}^n$ itself) that satisfy the relation

$$f(\lambda x) = \lambda f(x) \quad \text{for all } x \in C \text{ and all } \lambda > 0.$$

Such functions are called *positively homogeneous*. An important example is the norm mapping $\|\cdot\|$, which is defined on the whole space $\mathbb{R}^n$.

2.3.1. Lemma. *Let f be a positively homogeneous function defined on a convex cone C in $\mathbb{R}^n$. Then f is convex if and only if f is subadditive.*

Proof. Suppose that f is convex and $x, y \in C$. Then

$$\frac{1}{2} f(x + y) = f\Big(\frac{x + y}{2}\Big) \leq \frac{1}{2}\left(f(x) + f(y)\right)$$

and so $f(x + y) \leq f(x) + f(y)$.

Conversely, suppose that f is subadditive. Then

$$f((1-\lambda)x+\lambda y) \leq f((1-\lambda)x)+f(\lambda y) = (1-\lambda)f(x)+\lambda f(y)$$

for all $x, y \in C$ and $\lambda \in [0,1]$, which shows that f is convex. ■

2.3.2 Lemma. *Let f be a nonnegative positively homogeneous function defined on a convex cone C in $\mathbb{R}^n$ such that the sublevel set $\{x \in C : f(x) \leq 1\}$ is convex. Then f is a convex function.*

Proof. According to Lemma 2.3.1, it suffices to show that f is subadditive. For that, let $x, y \in C$ and choose scalars α and β such that $\alpha > f(x)$, and $\beta > f(y)$. Since f is nonnegative and positively homogeneous, $f(x/\alpha) \leq 1$ and $f(y/\beta) \leq 1$. Thus x/α and y/β both lie in the sublevel set of f at height 1. The assumed convexity of this sublevel set shows that

$$\frac{1}{\alpha+\beta}\, f(x+y) = f\Big(\frac{x+y}{\alpha+\beta}\Big) = f\Big(\frac{\alpha}{\alpha+\beta}\cdot\frac{x}{\alpha}+\frac{\beta}{\alpha+\beta}\cdot\frac{y}{\beta}\Big) \leq 1,$$

that is, $f(x+y) \leq \alpha+\beta$ whenever $\alpha > f(x)$, $\beta > f(y)$. Hence $f(x+y) \leq f(x)+f(y)$, which shows that f is subadditive. ■

A sample of how the last lemma yields the convexity of some functions is as follows. Let $p \geq 1$ and consider the function f given on the nonnegative orthant $\mathbb{R}^n_+$ by the formula

$$f(x_1, \ldots, x_n) = (x_1^p + \cdots + x_n^p)^{1/p}.$$

Clearly, f is nonnegative and positively homogeneous, and f^p is convex as a sum of convex functions. Hence the sublevel set

$$\{x \in X : f(x) \leq 1\} = \{x \in X : f^p(x) \leq 1\}$$

is convex and this implies that f is a convex function. By Lemma 2.3.1 we conclude that f is subadditive, a fact which is equivalent with the Minkowski inequality.

The *support function* of a nonempty compact convex set C in $\mathbb{R}^n$ is defined by

$$h(u) = \sup_{x \in C}\langle x, u\rangle, \quad u \in \mathbb{R}^n.$$

If $\|u\| = 1$, the set $H_\alpha = \{x \in \mathbb{R}^n : \langle x, u\rangle = \alpha\}$ describes a family of parallel hyperplanes, each having u as a normal vector; $\alpha = h(u)$ represents the value for which each H_α supports C and C is contained in the half-space H^-_α.

Notice that the support function is positively homogeneous and convex. Moreover

$$C = \{x \in \mathbb{R}^n : \langle x, u\rangle \leq h(u) \text{ for all } u \in \mathbb{R}^n\},$$

which shows that C is the intersection of all half-spaces that contain it.

Conversely, let $h\colon \mathbb{R}^n \to \mathbb{R}$ be a positively homogeneous convex function. Then

$$C = \{x \in \mathbb{R}^n : \langle x, u\rangle \le h(u) \text{ for every } u \in \mathbb{R}^n\}$$

is nonempty, compact and convex set whose support function is h.

The notion of a support function can be attached to any nonempty convex set C in $\mathbb{R}^n$. See the end of Section 2.4.

The Minkowski functional (also called the *gauge function*) associated to a nonempty subset C of $\mathbb{R}^n$ is the function

$$p_C\colon \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}, \quad p_C(x) = \inf\{\lambda > 0 : x \in \lambda C\},$$

with the convention $\inf \emptyset = \infty$. Suppose that C is a closed convex set which contains the origin. Then:

(i) p_C is a positively homogeneous convex function.

(ii) The Minkowski functional of C is the support function of the polar set C°, and the Minkowski functional of C° is the support function of C.

(iii) C° is bounded if and only if $0 \in \operatorname{int} C$ (so by the bipolar theorem C is bounded if and only if $0 \in \operatorname{int} C^\circ$). As a consequence, the Minkowski functional of C is real-valued if $0 \in \operatorname{int} C$.

It is well known Jensen's inequality in the context of finite measure spaces. Recently, P. Roselli and M. Willem [43] proved an extension of this inequality for all measure spaces, under the assumption that the convex function under attention is positively homogeneous and continuous. The basic ingredient in their proof is the following result:

2.3.3. Lemma. *Suppose that $J\colon \mathbb{R}^2_+ \to \mathbb{R}$ is a positively homogeneous continuous function. Then the following assertions are equivalent*:

(i) *J is convex*;

(ii) *$\varphi = J(1,t)$ is convex*;

(iii) *there exists a subset $G \subset \mathbb{R}^2$ such that*

$$J(u,v) = \sup\{au + bv : (a,b) \in G\}.$$

Proof. Clearly, (i) $\Rightarrow$ (ii) and (iii) $\Rightarrow$ (i). For (ii) $\Rightarrow$ (iii) notice that $J(u,v) = uJ(1, v/u)$ if $u > 0$ and $J(u,v) = vJ(0,1)$ if $u = 0$. Or,

$$\varphi(t) = \sup\{a + bt : (a,b) \in G\}$$

where $G = \{(\varphi(s) - sb, b) : b \in \partial\varphi(s),\ s \in \mathbb{R}\}$. ∎

2.3.4. Theorem (Roselli–Willem theorem). *Let $J\colon \mathbb{R}^2_+ \to \mathbb{R}$ be a positively homogeneous continuous convex function. Then for every measure space (X, Ω, μ) and every μ-integrable function $f\colon X \to \mathbb{R}^2_+$ for which $J \circ f$ is also μ-integrable, we have the inequality*

$$J\left(\int_X f\, d\mu\right) \le \int_X J \circ f\, d\mu. \tag{2.4}$$

The role of $\mathbb{R}_+^2$ can be taken by every cone in $\mathbb{R}_+^n$.

Proof. Put $f = (f_1, f_2)$. According to Lemma 2.3.3, and Lebesgue's dominated convergence theorem,

$$\begin{aligned}\int_X J \circ f \, d\mu &= \int_X \sup_{(a,b)\in G} (af_1 + bf_2)\, d\mu \\ &\geq \sup_{(a,b)\in G} \Big(a \int_X f_1 \, d\mu + b \int_X f_2 \, d\mu \Big) = J\Big(\int_X f \, d\mu \Big). \quad \blacksquare\end{aligned}$$

The particular case where $f(x) = (|u(x)|^p, |v(x)|^p)$ and

$$J(u, v) = (u^{1/p} + v^{1/p})^p \quad (p \in \mathbb{R},\ p \neq 0)$$

gives us a very general version of Minkowski's inequality:

2.3.5. Theorem. *For $p \in (-\infty, 0) \cup [1, \infty)$ and $f, g \in L^p(\mu)$ we have*

$$\|f + g\|_{L^p} \leq \|f\|_{L^p} + \|g\|_{L^p}, \tag{2.5}$$

while for $0 < p < 1$ the inequality works in the reverse sense,

$$\|f + g\|_{L^p} \geq \|f\|_{L^p} + \|g\|_{L^p}. \tag{2.6}$$

If f is not 0 almost everywhere, then we have equality if and only if $g = \lambda f$ almost everywhere, for some $\lambda \geq 0$.

Proof. In fact $J(1, t) = (1 + t^{1/p})^p$ is strictly convex for $0 < p < 1$ and strictly concave for $p \in (-\infty, 0) \cup (1, \infty)$. Then apply Theorem 2.3.4 above. $\blacksquare$

Another application of Theorem 2.3.4 is given by Hanner's inequalities.

2.3.6. Theorem (Hanner's inequalities). *If $f, g \in L^p(\mu)$ and $2 \leq p < \infty$, then*

$$\|f + g\|_{L^p}^p + \|f - g\|_{L^p}^p \leq (\|f\|_{L^p} + \|g\|_{L^p})^p + \big| \|f\|_{L^p} - \|g\|_{L^p} \big|^p,$$

equivalently (by making the replacements $f \to f + g$ and $g \to f - g$),

$$(\|f + g\|_{L^p} + \|f - g\|_{L^p})^p + \big| \|f + g\|_{L^p} - \|f - g\|_{L^p} \big|^p \geq 2^p (\|f\|_{L^p}^p + \|g\|_{L^p}^p).$$

If $1 < p \leq 2$, the above inequalities are reversed.

Proof. Apply Theorem 2.3.4 for $f(x) = (|u(x)|^p, |v(x)|^p)$ and $J(u, v) = (u^{1/p} + v^{1/p})^p + |u^{1/p} - v^{1/p}|^p$. $\blacksquare$

By using the inequalities of Hanner as a substitute for parallelogram's law we can infer (as in Theorem 1.2.1) that all nonempty closed convex subsets in a space $L^p(\mu)$ $(1 < p < \infty)$ are proximinal.

2.4 The Subdifferential

As already noted in Section 2.1, if f is a convex function (on an open convex subset U of a normed linear space E), then f has a supporting hyperplane at each point $a \in U$. This means the existence of a continuous linear functional h on E (the support of f at a) such that

$$f(x) \geq f(a) + h(x - a) \quad \text{for all } x \in U. \tag{2.7}$$

The set $\partial f(a)$ of all such functionals h constitutes the *subdifferential* of f at the point a.

2.4.1. Theorem. *Suppose that U is an open convex set in a normed linear space E. Then a function $f \colon U \to \mathbb{R}$ is convex if and only if $\partial f(a) \neq \emptyset$ at all $a \in U$.*

When E is $\mathbb{R}^n$ (or, more generally, a Hilbert space), all such h can be uniquely represented as

$$h(x) = \langle x, z \rangle \quad \text{for } x \in E.$$

In this case the inequality (2.7) becomes

$$f(x) \geq f(a) + \langle x - a, z \rangle \quad \text{for all } x \in U \tag{2.8}$$

and the subdifferential $\partial f(a)$ will be meant as the set of all such vectors z (usually called *subgradients*).

The subdifferential can be related to the directional derivative. Let f be a real-valued function defined on an open subset U of a Banach space E. The *one-sided directional derivatives of f at $a \in U$ relative to v* are defined to be the limits

$$f'_+(a; v) = \lim_{t \to 0+} \frac{f(a + tv) - f(a)}{t}$$

and

$$f'_-(a; v) = \lim_{t \to 0-} \frac{f(a + tv) - f(a)}{t}.$$

If both directional derivatives $f'_+(a; v)$ and $f'_-(a; v)$ exist and they are equal, we shall call their common value the *directional derivative of f at a, relative to v* (also denoted $f'(a; v)$). Notice that the one-sided directional derivatives of a convex function $f \colon U \to \mathbb{R}$ are positively homogeneous and subadditive (as a function of v). In fact, for $a \in U$, $u, v \in E$ and $t > 0$ small enough, we have

$$\frac{f(a + t(u + v)) - f(a)}{t} \leq \frac{f(a + 2tu) - f(a)}{2t} + \frac{f(a + 2tv) - f(a)}{2t}$$

which yields $f'_+(a; u + v) \leq f'_+(a; u) + f'_+(a; v)$.

Taking into account the formula

$$f'_+(a; v) = -f'_-(a; -v),$$

we infer that the directional derivatives (when they exist) are linear.

The directional derivatives relative to the vectors of the canonical basis of $\mathbb{R}^n$ are nothing but the partial derivatives.

If f is convex, then for each pair $(a,v) \in U \times E$ there exists an interval $(-\varepsilon, \varepsilon)$ on which the function $t \to f(a+tv)$ is well-defined and convex. Taking into account the one real variable case, it follows that every convex function admits one-sided directional derivatives at any point and that

$$f'_+(a;v) \geq f'_-(a;v).$$

As $f'_-(a;v) = -f'_+(a;-v)$, the above discussion yields the following result:

2.4.2. Lemma. *Suppose that f is a convex function defined on an open convex subset U of $\mathbb{R}^n$. Then $z \in \partial f(a)$ if and only if $f'_+(a;v) \geq \langle z, v\rangle$ for all $v \in \mathbb{R}^n$.*

In the finite dimensional case, $\partial f(a)$ is a singleton precisely when f has a directional derivative $f'(a;v)$ relative to any v. In that case, $\partial f(a)$ consists of the mapping $v \to f'(a;v)$. See Theorem 2.5.2.

If $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is a lower semicontinuous proper convex function, we say that $z \in \mathbb{R}^n$ is a *subgradient* of f at $a \in \operatorname{dom} f$ if

$$f(x) \geq f(a) + \langle x-a, z\rangle \quad \text{for all } x \in \mathbb{R}^n. \tag{2.9}$$

We call the set $\partial f(a)$, of all subgradients of f at a, the *subdifferential* of f (at the point a).

A derivative is a local property, while the subgradient definition (2.8) describes a global property. An illustration of this idea is the following remark: *for any lower semicontinuous proper convex function $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$, the point a is a global minimizer of f if and only if*

$$0 \in \partial f(a).$$

Remarks (The subdifferential calculus).

(i) Suppose that f, f_1, f_2 are convex functions on $\mathbb{R}^n$ and $a \in \mathbb{R}^n$. By Lemma 2.4.2, we know that

$$f'_+(a;v) = \sup\{\langle z, v\rangle : z \in \partial f(a)\} \quad \text{for all } v \in \mathbb{R}^n.$$

Let λ_1 and λ_2 be two positive numbers. Then

$$\partial(\lambda_1 f_1 + \lambda_2 f_2)(a) = \lambda_1 \partial f_1(a) + \lambda_2 \partial f_2(a).$$

In the general setting of proper convex functions, only the inclusion $\supset$ works. The equality needs additional assumptions, for example, the existence of a common point in the convex sets $\operatorname{ri}(\operatorname{dom} f_k)$ for $k = 1, \dots, m$. See [42, p. 223].

(ii) Let f be a proper convex function on $\mathbb{R}^n$ and let A be a linear transformation from $\mathbb{R}^m$ to $\mathbb{R}^n$. Then

$$\partial(f \circ A)(x) \supset A^* \partial f(Ax).$$

The equality needs additional assumptions. For example, it works when the range of A contains a point of $\operatorname{ri}(\operatorname{dom} f)$. See [42, p. 225].

(iii) (Subdifferential of a max-function). Suppose that $f_1, \ldots, f_m$ are convex functions on $\mathbb{R}^n$ and set

$$f = \max\{f_1, \ldots, f_m\}$$

and

$$J(a) = \{j : f_j(a) = f(a)\}$$

for $a \in \mathbb{R}^n$. Then $\partial f(a) = \operatorname{co}\{\partial f_j(a) : j \in J(a)\}$.

The *subdifferential* of f is defined as the set-valued map ∂f which associates to each $x \in \mathbb{R}^n$ the subset $\partial f(x) \subset \mathbb{R}^n$. Equivalently, ∂f may be seen as a graph in $\mathbb{R}^n \times \mathbb{R}^n$. Given two set-valued maps $u, v \colon \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$, we define:

- *domain of* u, $\operatorname{dom} u = \{x : u(x) \neq \emptyset\}$;
- *graph of* u, $\operatorname{graph} u = \{(x, y) : y \in u(x)\}$;
- *inverse of* u, $u^{-1}(y) = \{x : y \in u(x)\}$;
- $u \subset v$ if the graph of u is contained in the graph of v.

2.4.3. Definition. A set-valued map $u \colon \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$ is said to be *monotone* if it verifies

$$\langle x_1 - x_2, y_1 - y_2 \rangle \geq 0$$

for all $x_1, x_2 \in \mathbb{R}^n$ and all $y_1 \in u(x_1)$, $y_2 \in u(x_2)$. A monotone function u is called *maximal monotone* when it is maximal with respect to inclusion in the class of monotone functions, that is, if the following implication holds:

$$v \supset u \text{ and } v \text{ monotone} \implies v = u.$$

According to Zorn's lemma, for each monotone function u there exists a maximal monotone function $\widetilde{u}$ which includes u.

The graph of any maximal monotone map $u \colon \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$ is closed and thus it verifies the following conditions of upper semicontinuity:

$$x_k \to x,\ y_k \to y, \text{ and } y_k \in u(x_k) \text{ for all } k \in \mathbb{N} \implies y \in u(x).$$

We shall prove the existence of a one-to-one correspondence between graphs of maximal monotone maps and graphs of nonexpansive functions. Recall that a function $h \colon \mathbb{R}^n \to \mathbb{R}^n$ is called *nonexpansive* if its Lipschitz constant verifies

$$\operatorname{Lip}(h) = \sup_{x \neq y} \frac{\|h(x) - h(y)\|}{\|x - y\|} \leq 1.$$

We shall need the following result concerning the extension of Lipschitz functions:

2.4.4. Theorem (M. D. Kirszbraun). *Suppose that A is a subset of $\mathbb{R}^n$ and $f \colon A \to \mathbb{R}^m$ is a Lipschitz function. Then there exists a Lipschitz function*

$\widetilde{f}\colon \mathbb{R}^n \to \mathbb{R}^m$ such that $\widetilde{f} = f$ on A and $\mathrm{Lip}\,(\widetilde{f}) = \mathrm{Lip}\,(f)$. *Moreover, we may choose $\widetilde{f}$ convex, when A and f are also convex.*

Proof. When $m = 1$, we may choose

$$\widetilde{f}(x) = \inf_{y\in A} \{f(y) + Lip\,(f)\cdot \|x - y\|\}.$$

In the general case, a direct application of this remark at the level of components of f leads to an extension $\widetilde{f}$ with $\mathrm{Lip}\,(\widetilde{f}) \leq \sqrt{m}\mathrm{Lip}\,(f)$. The existence of an extension with the same Lipschitz constant is described in [15, Section 2.10.43, p. 201]. ■

The aforementioned correspondence between graphs is realized by the *Cayley transform*, that is, by the linear isometry

$$\Phi\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^n, \quad \Phi(x, y) = \frac{1}{\sqrt{2}}(x + y, -x + y).$$

When $n = 1$, the Cayley transform represents a clockwise rotation of angle $\pi/4$. The precise statement of this correspondence is as follows:

2.4.5 Theorem (G. Minty [34]). *Let $u\colon \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$ be a maximal monotone map. Then $J = (I + u)^{-1}$ is defined on the whole $\mathbb{R}^n$ and $\Phi(\mathrm{graph}\, u)$ is the graph of a nonexpansive function $v\colon \mathbb{R}^n \to \mathbb{R}^n$, given by*

$$v(x) = x - \sqrt{2}\,(I + u)^{-1}(\sqrt{2}\,x). \tag{2.10}$$

Conversely, if $v\colon \mathbb{R}^n \to \mathbb{R}^n$ is a nonexpansive function, then the inverse image of graph v *under Φ is the graph of a maximal monotone function on $\mathbb{R}^n$.*

Here I denotes the identity map of $\mathbb{R}^n$.

Proof. Let u be a monotone map and let v be the set-valued function whose graph is $\Phi(\mathrm{graph}\, u)$. We shall show that v is nonexpansive in its domain (and thus single-valued). In fact, given $x \in \mathbb{R}^n$, we have

$$y \in v(x) \quad \text{if and only if} \quad \frac{x + y}{\sqrt{2}} \in u\left(\frac{x - y}{\sqrt{2}}\right) \tag{2.11}$$

and this yields $y \in x - \sqrt{2}\,(I + u)^{-1}(\sqrt{2}x)$ for all $y \in v(x)$.

Now, if $x_k \in \mathbb{R}^n$ and $y_k \in v(x_k)$ for $k = 1, 2$, we infer from (2.11) that

$$\langle (x_1 - y_1) - (x_2 - y_2), (x_1 + y_1) - (x_2 + y_2)\rangle \geq 0,$$

hence $\|y_1 - y_2\|^2 \leq \|x_1 - x_2\|^2$. This shows that v is indeed nonexpansive.

The same argument shows that Φ^{-1} maps graphs of nonexpansive functions into graphs of monotone functions.

Assuming that u is maximal monotone, we shall show that the domain of v is $\mathbb{R}^n$. In fact, if the contrary were true, we could apply Theorem 2.4.4 to extend v to a nonexpansive function $\widetilde{v}$ defined on the whole $\mathbb{R}^n$, and then $\Phi^{-1}(\mathrm{graph}\,\widetilde{v})$ provides a monotone extension of u, which contradicts the maximality of u. ■

2.4.6. Corollary. *Let $u\colon \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$ be a maximal monotone map. Then $J = (I+u)^{-1}$ is a nonexpansive map of $\mathbb{R}^n$ into itself.*

Proof. It is easy to see that $I + u$ (and thus $(I + u)^{-1}$) is monotone. By Theorem 2.4.5, the maximality of u yields the surjectivity of $I + u$, hence $\operatorname{dom}(I + u)^{-1} = \mathbb{R}^n$. In order to prove that $(I + u)^{-1}$ is also a nonexpansive function, let us consider points $x_k \in \mathbb{R}^n$ and $y_k \in u(x_k)$ (for $k = 1, 2$). Then

$$\begin{aligned} \|x_1 - x_2\|^2 &\le \langle x_1 - x_2, x_1 - x_2 + y_1 - y_2\rangle \\ &\le \|x_1 - x_2\| \cdot \|x_1 + y_1 - (x_2 + y_2)\|, \end{aligned} \tag{2.12}$$

which yields $\|x_1-x_2\| \le \|(x_1+y_1)-(x_2+y_2)\|$. Particularly, if $x_1+y_1 = x_2+y_2$, then $x_1 = x_2$, and this shows that $(I + u)^{-1}$ is single-valued. Consequently, $(I + u)^{-1}(x_k + y_k) = x_k$ for $k = 1, 2$ and thus (2.12) yields the nonexpansivity of $(I + u)^{-1}$. ■

An important class of maximal monotone maps is provided by the subdifferentials of convex functions.

2.4.7. Theorem. *If $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is a lower semicontinuous proper convex function, then ∂f is a maximal monotone function such that*

$$\operatorname{int}\operatorname{dom} f \subset \operatorname{dom}\partial f \subset \operatorname{dom} f.$$

Proof. The fact that ∂f is monotone follows from (2.9). According to Theorem 2.4.5, the maximality of ∂f is equivalent to the surjectivity of $\partial f + I$. To prove that $\partial f + I$ is onto, let us fix arbitrarily $y \in \mathbb{R}^n$, and choose $x \in \mathbb{R}^n$ as the unique minimizer of the coercive lower semicontinuous function

$$g\colon x \to f(x) + \frac{1}{2}\,\|x\|^2 - \langle x, y\rangle.$$

Then $0 \in \partial g(x)$, which yields $y \in \partial(f(x) + \|x\|^2/2) = (\partial f + I)(x)$. ■

One can prove (by examples) that both inclusions in Theorem 2.4.7 may be strict.

According to W. Fenchel [16], the *conjugate* (or the *Legendre transform*) of a function $f\colon \mathbb{R}^n \to \overline{\mathbb{R}}$ is the function $f^*\colon \mathbb{R}^n \to \overline{\mathbb{R}}$ defined by

$$f^*(y) = \sup_{x\in\mathbb{R}^n} [\langle x, y\rangle - f(x)].$$

The function f^* is always lower semicontinuous and convex, and, if the effective domain of f is nonempty, then f^* never takes the value $-\infty$. Clearly, $f \le g$ yields $f^* \ge g^*$ (and thus $f^{**} \le g^{**}$). Also, the following generalization of *Young's inequality* holds true: If f is a proper convex function then so is f^* and

$$f(x) + f^*(y) \ge \langle x, y\rangle \quad \text{for all } x, y \in \mathbb{R}^n.$$

Equality holds if and only if $\langle x, y\rangle \ge f(x)+f^*(y)$, equivalently, when $f(z) \ge f(x) + \langle y, z - x\rangle$ for all z (that is, when $y \in \partial f(x)$).

By Young's inequality we infer that

$$f(x) \geq \sup_{y \in \mathbb{R}^n} [\langle x, y\rangle - f^*(y)] = f^{**}(x) \quad \text{for all } x \in \mathbb{R}^n.$$

2.4.8. Theorem (Computing the Legendre transform). *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a strictly convex function of class C^1, such that $f(x)/\|x\| \to \infty$ as $\|x\| \to \infty$. Then*:

(i) *The map $x \to \nabla f(x)$ is a homeomorphism (from $\mathbb{R}^n$ onto itself)*;

(ii) $f^\star(y) = \langle y, (\nabla f)^{-1} y\rangle - f((\nabla f)^{-1} y)$ *for all* $y \in \mathbb{R}^n$;

(iii0 $f^\star$ *is a C^1 function and* $\nabla f^\star = (\nabla f)^{-1}$.

Proof. For every $x, y \in \mathbb{R}^n$, $x, y \neq 0$, the function $g(t) = f(x + ty)$ is strictly convex on $\mathbb{R}$ and thus $g'(1) - g'(0) = \langle \nabla f(x+y) - \nabla f(y), y\rangle > 0$. This shows that ∇f is one-to-one. Let $z \in \mathbb{R}^n$. Since $g(x) = f(x) - \langle x, z\rangle$ is coercive and C^1, it attains a global minimum at a point a for which $\nabla g(a) = \nabla f(a) - z = 0$. Hence ∇f is onto. The inequality $f(0) + \langle \nabla f(x), x\rangle \geq f(x)$ yields $\|\nabla f(x)\| \to \infty$ as $\|x\| \to \infty$. Therefore the inverse image under ∇f of every compact set is compact too, a fact which assures the continuity of $(\nabla f)^{-1}$. ■

If f is a convex function on $\mathbb{R}^n$, then $f = f^*$ if and only if $f(x) = \|x\|^2/2$.

One can prove that conjugacy induces a bijection between lower semicontinuous proper convex functions.

2.4.9. Theorem. *Suppose that $f \colon \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is a proper convex function. Then the following assertions are equivalent*:

(i) *f is lower semicontinuous*;

(ii) $f = f^{**}$;

(iii) *f is the pointwise supremum of the family of all affine functions h such that $h \leq f$.*

Proof. Clearly, (ii) ⇒ (i) and (ii) ⇒ (iii). Since any affine minorant h of f verifies $h = h^{**} \leq f^{**} \leq f$, it follows that (iii) ⇒ (ii). The implication (i) ⇒ (iii) can be proved easily by using the basic separation theorem. See [4, pp. 76–77].

Alternatively, we can show that (i) ⇒ (ii). If $x \in \operatorname{int}(\operatorname{dom} f)$, then $\partial f(x)$ is nonempty and for each $y \in \partial f(x)$ we have $\langle x, y\rangle = f(x) + f^*(y)$, hence $f(x) = \langle x, y\rangle - f^*(y) \leq f^{**}(x)$. In the general case, we may use an approximation argument. See the end of Section 2.5. ■

The Legendre transform allows us to attach the notion of *support function* to any nonempty convex set C in $\mathbb{R}^n$, by defining it as the conjugate of the indicator function of C. For example, the support function of $C = \{(x, y) \in \mathbb{R}^2 : x + y^2/2 \leq 0\}$ is $\delta_C^*(x, y) = y^2/2x$ if $x > 0$, $\delta_C^*(0, 0) = 0$ and $\delta_C^*(x, y) = \infty$ otherwise. Consequently, δ_C^* is a lower semicontinuous proper convex function.

2.5 Differentiability of Convex Functions

The problem of differentiability of a convex function defined on an open subset U of a Banach space E can be treated in the setting of Fréchet differentiability or in the more general setting of Gâteaux differentiability.

The *Fréchet differentiability* (or, simply, the *differentiability*) of f at a point a means the existence of a continuous linear functional $df(a)\colon E \to \mathbb{R}$ such that

$$\lim_{x\to a} \frac{|f(x) - f(a) - df(a)(x-a)|}{\|x-a\|} = 0.$$

Equivalently,

$$f(x) = f(a) + df(a)(x-a) + \omega(x)\|x-a\| \quad \text{for } x \in U,$$

where $\omega\colon U \to \mathbb{R}$ is a function such that $\omega(a) = \lim_{x\to a}\omega(x) = 0$. When $E = \mathbb{R}^n$, the functional $df(a)$ can be computed via the formula

$$df(a)(v) = \langle \nabla f(a), v\rangle,$$

where

$$\nabla f(a) = \sum_{k=1}^{n} \frac{\partial f}{\partial x_k}(a)\, e_k$$

represents the *gradient of f at a*. As usually, $e_1, \dots, e_n$ denotes the unit vector basis of $\mathbb{R}^n$.

A function $f\colon U \to \mathbb{R}$ is said to be *Gâteaux differentiable* at a point a if the directional derivative $f'(a;v)$ exists for every $v \in E$ and defines a continuous linear functional $f'(a)\colon v \to f'(a;v)$ on E. It is straightforward that differentiability implies Gâteaux differentiability and also the equality

$$f'(a) = df(a).$$

For convex functions on open subsets of $\mathbb{R}^n$, Gâteaux and Fréchet differentiability agree:

2.5.1. Theorem. *Suppose that a convex function f defined on an open convex set U in $\mathbb{R}^n$ possesses all its partial derivatives $\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n}$ at some point $a \in U$. Then f is differentiable at a.*

Proof. Since U is open, there is a $r > 0$ such that $B_r(a) \subset U$. We have to prove that the function

$$g(u) = f(a+u) - f(a) - \sum_{k=1}^{n} \frac{\partial f}{\partial x_k}(a) u_k$$

defined for all $u = (u_1, \dots, u_n)$ with $\|u\| < r$, verifies $\lim_{\|u\|\to 0} g(u)/\|u\| = 0$.

Clearly, the function g is convex. Then

$$0 = g(0) = g\Big(\frac{u + (-u)}{2}\Big) \le \frac{1}{2}\,(g(u) + g(-u)),$$

which yields $g(u) \geq -g(-u)$. On the other hand, for each u with $n\|u\| < r$, we have

$$\begin{aligned} g(u) &= g\Big(\frac{1}{n}\sum_{k=1}^{n} n u_k e_k\Big) \leq \frac{1}{n}\sum_{k=1}^{n} g(n u_k e_k) \\ &= \sum_{\{k:u_k\neq 0\}} u_k\,\frac{g(n u_k e_k)}{n u_k} \leq \|u\| \sum_{\{k:u_k\neq 0\}} \left|\frac{g(n u_k e_k)}{n u_k}\right|. \end{aligned}$$

Similarly,

$$g(-u) \leq \|u\| \sum_{\{k:u_k\neq 0\}} \left|\frac{g(-n u_k e_k)}{n u_k}\right|.$$

Then

$$-\|u\| \sum_{\{k:u_k\neq 0\}} \left|\frac{g(-n u_k e_k)}{n u_k}\right| \leq -g(-u) \leq g(u) \leq \|u\| \sum_{\{k:u_k\neq 0\}} \left|\frac{g(n u_k e_k)}{n u_k}\right|$$

and it remains to remark that $g(n u_k e_k)/(n u_k) \to 0$ as $u_k \to 0$. ■

The condition of differentiability is equivalent to the uniqueness of the support function:

2.5.2. Theorem. *Let f be a convex function defined on an open convex set U in $\mathbb{R}^n$. Then f is differentiable at a if and only if f has a unique support at a.*

Proof. Suppose that $f'(a;v)$ exists for every v. If $h\colon E \to \mathbb{R}$ is a support of f at a, then

$$f(a+\varepsilon v) - f(a) \geq \varepsilon h(v)$$

for sufficiently small $\varepsilon > 0$, which yields $f'(a;v) \geq h(v)$. Replacing v by $-v$, and taking into account that the directional derivative is linear in v, we obtain

$$-f'(a;v) = f'(a;-v) \geq -h(v)$$

from which we conclude that $h(v) = f'(a;v)$.

Suppose now that f has a unique support h at a and choose a number λ such that

$$-f'_+(a,-e_1) \leq \lambda \leq f'_+(a;e_1).$$

Then the line L in $\mathbb{R}^{n+1}$ given by $t \to (a+te_1, f(a)+\lambda t)$ meets the epigraph of f at $(a, f(a))$. Since $f(a+te_1) \geq f(a) + \lambda t$ as long as $a + te_1 \in U$, the line L does not meet the interior of the epigraph of f. By the Hahn–Banach theorem we infer the existence of a supporting hyperplane to the epigraph of f at $(a, f(a))$ which contains L. The uniqueness of the support of f at a shows that this hyperplane must be the graph of h. Then

$$h(a+te_1) = f(a) + \lambda t = h(a) + \lambda t$$

for all $t \in \mathbb{R}$, so that by the choice of λ we get $-f'_+(a,-e_1) = f'_+(a;e_1)$. In other words we established the existence of $\partial f/\partial x_1$ at a. Similarly, one can prove the existence of all partial derivatives at a so, by Theorem 2.5.1, the function f is differentiable at a. ∎

In the context of several variables, the set of points where a convex function is not differentiable can be uncountable, though still negligible:

2.5.3. Theorem. *Suppose that f is a convex function on an open subset U of $\mathbb{R}^n$. Then f is differentiable almost everywhere in U.*

Proof. Consider first the case when U is also bounded. According to Theorem 2.5.1 we must show that each of the sets

$$E_k = \{x \in U : \frac{\partial f}{\partial x_k}(x) \text{ does not exist}\}$$

is Lebesgue negligible. The measurability of E_k is a consequence of the fact that the limit of a pointwise converging sequence of measurable functions is measurable too. In fact, the formula

$$f'_+(x, e_k) = \lim_{j \to \infty} \frac{f(x + e_k/j) - f(x)}{1/j}$$

motivates the measurability of one-sided directional derivative $f'_+(x, e_k)$ and a similar argument applies for $f'_-(x, e_k)$. Consequently the set

$$E_k = \{x \in U : f'_+(x, e_k) - f'_-(x, e_k) > 0\}$$

is measurable. Being bounded, it is also integrable. By Fubini's theorem,

$$\begin{aligned} m(E_k) &= \int_{\mathbb{R}^n} \chi_{E_k} \, dx \\ &= \int_{\mathbb{R}} \cdots \left(\int_{\mathbb{R}} \chi_{E_k} \, dx_i \right) dx_1 \cdots dx_{i-1} \, dx_{i+1} \cdots dx_n \end{aligned}$$

and the interior integral is zero since f is convex as a function of x_i (and thus differentiable except at an enumerable set of points).

If U is arbitrary, the argument above shows that all the sets $E_k \cap B_n(0)$ are negligible. Or, $E_k = \bigcup_{n=1}^{\infty} (E_k \cap B_n(0))$ and a countable union of negligible sets is negligible too. ∎

The function $f(x, y) = \sup\{x, 0\}$ is convex on $\mathbb{R}^2$ and nondifferentiable at the points of y-axis (which constitutes an uncountable set).

The coincidence of Gâteaux and Fréchet differentiability is no longer true in the context of infinite dimensional spaces.

2.5.4. Theorem. *Let E be a Banach space such that for each continuous convex function $f\colon E \to \mathbb{R}$, every point of Gâteaux differentiability is also a point of Fréchet differentiability. Then E is finite dimensional.*

The proof we present here is due to J. M. Borwein and A. S. Lewis [4], and depends on a deep result in Banach space theory:

2.5.5. Theorem (The Josephson–Nissenzweig theorem; [29], [36]). *If E is a Banach space such that*

$$x_n' \to 0 \text{ in the weak-star topology of } E' \text{ implies } \|x_n'\| \to 0,$$

then E is finite dimensional.

Proof of Theorem 2.5.5. Consider a sequence $(x_n')_n$ of norm-1 functionals in E' and a sequence $(\alpha_n)_n$ of real numbers such that $\alpha_n \downarrow 0$. Then the function

$$f(x) = \sup_n [\langle x, x_n' \rangle - \alpha_n]$$

is convex and continuous and, moreover,

$$f \text{ is Gâteaux differentiable at } 0 \iff x_n'(x) \to 0 \text{ for all } x \in E$$
$$f \text{ is Fréchet differentiable at } 0 \iff \|x_n'\| \to 0.$$

The proof ends by applying the Josephson–Nissenzweig theorem. ■

Convolution by smooth functions provides us with a powerful technique for approximating locally integrable functions by C^∞ functions. Particularly, this applies to the convex functions.

Let φ be a *mollifier*, that is, a nonnegative function in $C_c^\infty(\mathbb{R}^n)$ such that

$$\int_{\mathbb{R}^n} \varphi \, dx = 1 \quad \text{and} \quad \operatorname{supp} \varphi \subset \overline{B}_1(0).$$

The standard example of such a function is given by

$$\varphi(x) = \begin{cases} C \exp(-1/(1 - \|x\|^2)) & \text{if } \|x\| < 1, \\ 0 & \text{if } \|x\| \geq 1, \end{cases}$$

where C is chosen such that $\int_{\mathbb{R}^n} \varphi \, dx = 1$. Each mollifier φ gives rise to an one-parameter family of nonnegative functions

$$\varphi_\varepsilon(x) = \frac{1}{\varepsilon^n} \varphi\Big(\frac{x}{\varepsilon}\Big), \quad \varepsilon > 0$$

with similar properties:

$$\varphi_\varepsilon \in C_c^\infty(\mathbb{R}^n), \quad \operatorname{supp} \varphi_\varepsilon \subset \overline{B}_\varepsilon(0) \quad \text{and} \quad \int_{\mathbb{R}^n} \varphi_\varepsilon \, dx = 1.$$

The following lemma is standard and available in many places. For example, see [14, pp. 122–125] or [46, pp. 22–23].

2.5.6. Lemma. *Suppose that $f \in L^1_{loc}(\mathbb{R}^n)$ and $(\varphi_\varepsilon)_{\varepsilon>0}$ is the one-parameter family of functions associated to a mollifier φ. Then:*

(i) *the functions*

$$f_\varepsilon = \varphi_\varepsilon * f$$

belong to $C^\infty(\mathbb{R}^n)$ *and*

$$D^\alpha f_\varepsilon = D^\alpha \varphi_\varepsilon * f$$

for every multi-index α;

(ii) $f_\varepsilon(x) \to f(x)$ *whenever* x *is a point of continuity of* f. *If* f *is continuous on an open subset* U, *then* f_ε *converges uniformly to* f *on each compact subset of* U;

(iii) *if* $f \in L^p(\mathbb{R}^n)$ *(for some* $p \in [1,\infty)$*), then* $f_\varepsilon \in L^p(\mathbb{R}^n)$, $\|f_\varepsilon\|_{L^p} \leq \|f\|_{L^p}$ *and* $\lim_{\varepsilon\to 0} \|f_\varepsilon - f\|_{L^p} = 0$;

(iv) *if* f *is a convex function on an open convex subset* U *of* $\mathbb{R}^n$, *then* f_ε *is convex too.*

Mollification allows us to prove that the convex functions on open convex subsets in $\mathbb{R}^n$ are locally Lipschitz. In fact, by Lemma 2.5.6 (iv) we can infer that any such function $f : B_r(a) \to \mathbb{R}$ verifies inequalities of the form

$$\begin{aligned}
\sup_{x\in \bar{B}_{r/2}(a)} |f(x)| &\leq \frac{C}{\operatorname{Vol}(B_r(a))} \int_{B_r(a)} |f(y)|\, dy \\
\operatorname*{ess\,sup}_{x\in \bar{B}_{r/2}(a)} |Df(x)| &\leq \frac{C}{r\operatorname{Vol}(B_r(a))} \int_{B_r(a)} |f(y)|\, dy,
\end{aligned}$$

where $C > 0$ is a constant. As a hint, consider first the case of C^2 functions. See [14], pp. 236-239, for details.

A nonlinear analogue of mollification is offered by the *infimal convolution*, which for two proper convex functions $f, g\colon E \to \mathbb{R}\cup\{\infty\}$ is defined by the formula

$$(f \odot g)(x) = \inf\{f(x-y) + g(y) : y \in E\};$$

the value $-\infty$ is allowed. If $(f \odot g)(x) > -\infty$ for all x, then $f \odot g$ is a proper convex function. For example, this happens when both functions f and g are nonnegative (or, more generally, when there exists an affine function $h\colon E \to \mathbb{R}$ such that $f \geq h$ and $g \geq h$).

Infimal convolution and addition are inverse each other under the action of Legendre transform:

(INF1) $(f \odot g)^* = f^* + g^*$;

(INF2) $(f+g)^* = f^* \odot g^*$ if the effective domain of f contains a point of continuity of g.

By computing the infimal convolution of the norm function and the indicator function of a nonempty convex set C, we get

$$(\|\cdot\| \odot \delta_C)(x) = \inf_{y\in C} \|x-y\| = d_C(x),$$

a fact which implies the convexity of the distance function.

A standard way to approximate from below a lower semicontinuous proper convex function $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is the *Moreau–Yosida approximation*:

$$\begin{aligned} f_\varepsilon(x) &= \Big(f \odot \frac{1}{2\varepsilon}\|\cdot\|^2\Big)(x) \\ &= \inf_{y\in\mathbb{R}^n}\Big\{f(y) + \frac{1}{2\varepsilon}\|x-y\|^2\Big\} \end{aligned}$$

for $x \in \mathbb{R}^n$ and $\varepsilon > 0$. The functions f_ε are well-defined and finite for all x because the function $y \to f(y) + \frac{1}{2\varepsilon}\|x-y\|^2$ is lower semicontinuous and also coercive (due to the existence of a support for f).

2.5.7. Lemma. *The Moreau–Yosida approximates f_ε are differentiable convex functions on $\mathbb{R}^n$ and $f_\varepsilon \to f$ as $\varepsilon \to 0$. Moreover, $\partial f_\varepsilon = (\varepsilon I + (\partial f)^{-1})^{-1}$ as set-valued maps.*

Proof. The first statement is straightforward. The proof of the second one may be found in [1] and [3]. ■

The Moreau-Yosida approximates can be used to complete the proof of the implication (i) ⇒ (ii) in Theorem 2.4.8. In fact,

$$f^{**}(x) \geq \liminf_{\varepsilon\to 0} f_\varepsilon^{**}(x) = \liminf_{\varepsilon\to 0} f_\varepsilon(x) = f(x).$$

The infimal convolution provides an efficient regularization procedure for (even degenerate) elliptic equations. This explains the *Lax formula*,

$$u(x,t) = \sup_{y\in\mathbb{R}^n}\Big\{v(y) - \frac{1}{2t}\|x-y\|^2\Big\},$$

for the solution of the *Hamilton-Jacobi equation*,

$$\begin{aligned} \frac{\partial u}{\partial t} - \frac{1}{2}\|\nabla u\|^2 = 0 &\quad \text{for } x\in\mathbb{R}^n,\ t>0 \\ u|_{t=0} = v &\quad \text{on } \mathbb{R}^n. \end{aligned}$$

See J. M. Lasry and P.-L. Lions [32] for details.

2.6 Recognizing Convex Functions

We start with the following variant of Theorem 2.4.1:

2.6.1. Theorem. *Suppose that f is defined on an open convex set U in a Banach space. If f is convex on U and Gâteaux differentiable at $a \in U$, then*

$$f(x) \geq f(a) + f'(a; x-a) \quad \textit{for every } x \in U. \tag{2.13}$$

If f is Gâteaux differentiable throughout U, then f is convex if and only if (2.13) holds for all $a \in U$. Moreover, f is strictly convex if and only if the inequality is strict for $x \neq a$.

On intervals, a differentiable function is convex if and only if its derivative is nondecreasing. The higher dimensional analogue of this fact is as follows:

2.6.2. Theorem. *Suppose that f is Gâteaux differentiable on the open convex set U in a Banach space. Then f is convex if and only if*

$$f'(x; x-y) \geq f'(y; x-y) \tag{2.14}$$

for all $x, y \in U$.

The variant of this result for strictly convex functions asks the above inequality to be strict for $x \neq y$ in U.

Proof. If f is convex, then for x and y in U and $0 < t < 1$ we have

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

so by letting $t \to 0+$ we obtain $f'(y; x-y) \leq f(x) - f(y)$. Interchanging x and y, we also have $f'(x; y-x) \leq f(y) - f(x)$. Adding, we arrive at (2.14).

Suppose now that (2.14) holds. Let $x, y \in U$ and consider the function $g(\lambda) = f((1-\lambda)x + \lambda y)$, $\lambda \in [0,1]$. One can easily verify that

$$\lambda_1 \leq \lambda_2 \text{ implies } g'(\lambda_1) \leq g'(\lambda_2)$$

which shows that g is convex. Then

$$\begin{aligned} f((1-\lambda)x + \lambda y) &= g(\lambda) = g(\lambda \cdot 1 + (1-\lambda) \cdot 0) \\ &\leq \lambda g(1) + (1-\lambda) g(0) \\ &= (1-\lambda) f(x) + \lambda f(y). \quad \blacksquare \end{aligned}$$

When the ambient space is $\mathbb{R}^n$, then the inequality (2.14) becomes

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq 0. \tag{2.15}$$

In this context, a function $F\colon U \to \mathbb{R}^n$ is said to be *nondecreasing* (respectively *increasing*) if it is the gradient of a convex (strictly convex) function.

Higher differentiability leads to other important criteria of convexity.

Suppose that $f\colon U \to \mathbb{R}$ is Gâteaux differentiable. We say that f is *twice Gâteaux differentiable* at $a \in U$ if the limit

$$f''(a; v, w) = \lim_{\lambda \to 0} \frac{f'(a + \lambda w, v) - f'(a; v)}{\lambda}$$

exists for all v, w in the ambient Banach space E. This gives rise to a map $f''(a)\colon (v, w) \mapsto f''(a; v, w)$, from $E \times E$ into $\mathbb{R}$, called the *second Gâteaux differential* of f at a. One can prove easily that this function is homogeneous in v and w, that is,

$$f''(a; \lambda v, \mu w) = \lambda \mu f''(a; v, w)$$

for all $\lambda, \mu \in \mathbb{R}$. Another immediate fact is as follows:

2.6.3. Lemma. *If $f\colon U \to \mathbb{R}$ is twice differentiable, then it is also twice Gâteaux differentiable and*

$$d^2f(a)(v,w) = f''(a;v,w) \tag{2.16}$$

for all $a \in U$ and $v, w \in E$.

Our next goal is to establish the analogue of Taylor's formula in the context of Gâteaux differentiability and to infer from it an important characterization of convexity under the presence of Gâteaux differentiability.

2.6.4. Theorem (Taylor's formula). *If f is twice Gâteaux differentiable at all points of the segment $[a, a+v]$ relative to the pair (v,v), then there exists a $\theta \in (0,1)$ such that*

$$f(a+v) = f(a) + f'(a;v) + \frac{1}{2}f''(a+\theta v; v, v). \tag{2.17}$$

Proof. Consider the function $g(t) = f(a+tv)$, for $t \in [0,1]$. Its derivative is

$$\begin{aligned} g'(t) &= \lim_{\varepsilon \to 0} \frac{g(t+\varepsilon) - g(t)}{\varepsilon} \\ &= \lim_{\varepsilon \to 0} \frac{f(a+tv+\varepsilon v) - f(a+tv)}{\varepsilon} = f'(a+tv;v) \end{aligned}$$

and similarly, $g''(t) = f''(a+tv;v,v)$. Then by the usual Taylor's formula we get a $\theta \in (0,1)$ such that

$$g(1) = g(0) + g'(0) + \frac{1}{2}\,g''(\theta),$$

which in turn yields the formula (2.17). ■

2.6.5. Corollary. *Suppose that f is twice Gâteaux differentiable on the open convex set U in a Banach space E and*

$$f''(a;v,v) \geq 0 \quad \text{for all } a \in U,\ v \in E. \tag{2.18}$$

Then f is convex on U. If the above inequality is strict for $v \neq 0$, then f is strictly convex.

Proof. In fact, by Taylor's formula we have

$$f(x) = f(a) + f'(a;x-a) + \frac{1}{2}\,f''(a+\theta(x-a); x-a, x-a)$$

for some $\theta \in (0,1)$, so by our hypothesis,

$$f(x) \geq f(a) + f'(a;x-a)$$

and the conclusion follows from Theorem 2.6.1. ■

When $E = \mathbb{R}^n$ and $f''(a; v, w)$ is bilinear, it is easy to check the equality

$$f''(a; v, w) = \langle (H_a f)v, w\rangle,$$

where

$$H_a f = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(a)\right)_{i,j=1}^{n}$$

is the *Hessian matrix* of f at a.

Corollary 2.6.5 shows that the positivity (strict positivity) of the Hessian matrix at all points of U guarantees the convexity (strict convexity) of f.

If $A \in \mathrm{M}_n(\mathbb{R})$ is a strictly positive matrix and $u \in \mathbb{R}^n$, then the function

$$f(x) = \frac{1}{2}\langle Ax, x\rangle - \langle x, u\rangle,$$

satisfies

$$f'(x; v) = \langle v, Ax\rangle - \langle v, u\rangle,$$
$$f''(x; v, w) = \langle Av, w\rangle = \langle v, Aw\rangle,$$

so by Corollary 2.6.5 it follows that f is strictly convex. By Theorem 2.2.5, f admits a global minimum a. According to Fermat's theorem (applied to the function $t \to f(a + tv)$), we infer that $f'(a; v) = 0$ for all v. This shows that a is the solution of the equation

$$Ax = u.$$

The above idea, to solve equations by finding the minimum of suitable functionals, is very useful in partial differential equations. This is detailed in the next chapter.

Corollary 2.6.5 is the source of many interesting inequalities. Here are two examples.

Consider the open set $U = \{(x, y, z) \in \mathbb{R}^3 : x, y > 0, xy > z^2\}$. Then U is convex and the differentiable function

$$f\colon U \to \mathbb{R}, \quad f(x, y, z) = \frac{1}{xy - z^2}$$

is strictly convex. As a consequence, we infer that

$$\frac{8}{(x_1 + x_2)(y_1 + y_2) - (z_1 + z_2)^2} < \frac{1}{x_1 y_1 - z_1^2} + \frac{1}{x_2 y_2 - z_2^2}$$

for every pair of distinct points (x_1, y_1, z_1) and (x_2, y_2, z_2) of the set U.

The fact that the function

$$f\colon [0, \infty)^n \to \mathbb{R}, \quad f(x_1, \ldots, x_n) = \sqrt[n]{x_1 \cdots x_n}$$

is concave yields the Minkowski's inequality for $p = 0$:

$$\sqrt[n]{(x_1 + y_1) \cdots (x_n + y_n)} \geq \sqrt[n]{x_1 \cdots x_n} + \sqrt[n]{y_1 \cdots y_n}$$

for all $x_1, \ldots, x_n, y_1, \ldots, y_n \geq 0$.

2.7 The Convex Programming Problem

The aim of this section is to discuss the problem of minimizing a convex function over a convex set defined by a system of convex inequalities. The main result is the equivalence of this problem to the so-called saddle-point problem. Assuming the differentiability of the functions concerned, the solution of the saddle-point problem is characterized by the Karush–Kuhn–Tucker conditions, which will be made explicit in Theorem 2.7.2 below.

In what follows $f, g_1, \ldots, g_m$ will denote convex functions on $\mathbb{R}^n$. The *convex programming problem* for these data is to minimize $f(x)$ over the convex set

$$X = \{x \in \mathbb{R}^n : x \geq 0,\ g_1(x) \leq 0, \ldots, g_m(x) \leq 0\}.$$

In optimization theory f represents a *cost*, which is minimized over the *feasible set* X.

A particular case is the standard *linear programming problem.* In this problem we seek to maximize a linear function

$$L(x) = -\langle x, c\rangle = -\sum_{k=1}^{n} c_k x_k$$

subject to the constraints

$$x \geq 0 \quad \text{and} \quad Ax \leq b.$$

Here $A = (a_{ij})_{i,j} \in \mathrm{M}_n(\mathbb{R})$ and $b, c \in \mathbb{R}^n$. Notice that this problem can be easily converted into a minimization problem, by replacing L by $-L$. According to Theorem 2.1.7, L attains its global maximum at an extreme point of the convex set $\{x : x \geq 0, Ax \leq b\}$. This point can be found by the *simplex algorithm* of G. B. Dantzig. See [41] for details.

The linear programming problem has many practical applications in allocation of resources. For example, in the *diet problem*, we seek for the minimum of $\sum_{k=1}^{n} c_k x_k$ subject to

$$\sum_{j=1}^{n} a_{ij} x_j \geq b_i \text{ (for } i = 1, \ldots, n) \text{ and } x_j \geq 0 \text{ for } j = 1, \ldots, n.$$

Here

a_{ij} represents the amount of nutrient i in one unit of food j;
b_i represents the minimum daily requirement of nutrient i;
c_k is the cost of one unit of food k;
x_k is the amount of food k in a daily diet.

We pass now to the solution of the convex programming problem. As in the case of any constrained extremal problem, one can apply the method of Lagrange multipliers in order to eliminate the constraints (at the cost of increasing the number of variables). The *Lagrangian function* associated with the convex programming problem is the function

$$F(x, y) = f(x) + y_1 g_1(x) + \cdots + y_m g_m(x)$$

of $n + m$ real variables $x_1, \ldots, x_n, y_1, \ldots, y_m$ (the components of x and respectively of y). A *saddle point* of F is any point (x^0, y^0) of $\mathbb{R}^n \times \mathbb{R}^m$ such that

$$x^0 \geq 0, \quad y^0 \geq 0$$

and

$$F(x^0, y) \leq F(x^0, y^0) \leq F(x, y^0)$$

for all $x \geq 0$, $y \geq 0$. The saddle points of F will provide solutions to the convex programming problem that generates F:

2.7.1. Theorem. *Let (x^0, y^0) be a saddle point of the Lagrangian function F. Then x^0 is a solution to the convex programming problem and*

$$f(x^0) = F(x^0, y^0).$$

Proof. The condition $F(x^0, y) \leq F(x^0, y^0)$ yields

$$y_1 g_1(x^0) + \cdots + y_m g_m(x^0) \leq y_1^0 g_1(x^0) + \cdots + y_m^0 g_m(x^0).$$

By keeping $y_2, \ldots, y_m$ fixed and taking the limit as $y_1 \to \infty$ we infer that $g_1(x^0) \leq 0$. Similarly, $g_2(x^0) \leq 0, \ldots, g_m(x^0) \leq 0$. Thus x^0 belongs to the feasible set X.

From $F(x^0, 0) \leq F(x^0, y^0)$ and the definition of X we infer

$$0 \leq y_1^0 g_1(x^0) + \cdots + y_m^0 g_m(x^0) \leq 0,$$

that is, $y_1^0 g_1(x^0) + \cdots + y_m^0 g_m(x^0) = 0$. Then $f(x^0) = F(x^0, y^0)$. Since $F(x^0, y^0) \leq F(x, y^0)$ for all $x \geq 0$, we have

$$f(x^0) \leq f(x) + y_1^0 g_1(x) + \cdots + y_m^0 g_m(x) \leq f(x)$$

for all $x \geq 0$, which shows that x^0 is a solution to the convex programming problem. ■

2.7.2. Theorem (The Karush–Kuhn–Tucker conditions). *Suppose that the convex functions $f, g_1, \ldots, g_m$ are differentiable functions on $\mathbb{R}^n$. Then (x^0, y^0) is a saddle point of the Lagrangian function F if and only if*

$$x^0 \geq 0, \tag{2.19}$$

$$\frac{\partial F}{\partial x_k}(x^0, y^0) \geq 0, \quad \textit{for } k = 1, \ldots, n, \tag{2.20}$$

$$\frac{\partial F}{\partial x_k}(x^0, y^0) = 0 \quad \textit{whenever } x_k^0 > 0, \tag{2.21}$$

and

$$y^0 \geq 0, \tag{2.22}$$

$$\frac{\partial F}{\partial y_j}(x^0, y^0) = g_j(x^0) \leq 0, \quad \textit{for } j = 1, \ldots, m, \tag{2.23}$$

$$\frac{\partial F}{\partial y_j}(x^0, y^0) = 0 \quad \textit{whenever } y_j^0 > 0. \tag{2.24}$$

Proof. If (x^0, y^0) is a saddle point of F, then (2.19) and (2.22) are clearly fulfilled. Also,

$$F(x^0 + te_k, y^0) \geq F(x^0, y^0) \quad \text{for all } t \geq -x_k^0.$$

If $x_k^0 = 0$, then

$$\frac{\partial F}{\partial x_k}(x^0, y^0) = \lim_{t \to 0+} \frac{F(x^0 + te_k, y^0) - F(x^0, y^0)}{t} \geq 0.$$

If $x_k^0 > 0$, then $\frac{\partial F}{\partial x_k}(x^0, y^0) = 0$ by Fermat's theorem. In a similar way one can prove (2.23) and (2.24).

Suppose now that the conditions (2.19)–(2.24) are satisfied. As $F(x, y^0)$ is a differentiable convex function of x (being a linear combination, with positive coefficients, of such functions), it verifies the assumptions of Theorem 2.6.1. Taking into account the conditions (2.19)–(2.21), we are led to

$$\begin{aligned}
F(x, y^0) &\geq F(x^0, y^0) + \langle x - x^0, \nabla_x F(x^0, y^0) \rangle \\
&= F(x^0, y^0) + \sum_{k=1}^{n} (x_k - x_k^0) \frac{\partial F}{\partial x_k}(x^0, y^0) \\
&= F(x^0, y^0) + \sum_{k=1}^{n} x_k \frac{\partial F}{\partial x_k}(x^0, y^0) \geq F(x^0, y^0)
\end{aligned}$$

for all $x \geq 0$. On the other hand, by (2.23)–(2.24), for $y \geq 0$, we have

$$\begin{aligned}
F(x^0, y) &= F(x^0, y^0) + \sum_{j=1}^{m} (y_j - y_j^0)\, g_j(x^0) \\
&= F(x^0, y^0) + \sum_{j=1}^{m} y_j g_j(x^0) \\
&\leq F(x^0, y^0).
\end{aligned}$$

Consequently, (x^0, y^0) is a saddle point of F. ■

We shall illustrate Theorem 2.7.2 by the following example:

minimize $(x_1 - 2)^2 + (x_2 + 1)^2$ subject to $0 \leq x_1 \leq 1$ and $0 \leq x_2 \leq 2$.

Here $f(x_1, x_2) = (x_1 - 2)^2 + (x_2 + 1)^2$, $g_1(x_1, x_2) = x_1 - 1$ and $g_2(x_1, x_2) = x_2 - 2$. The Lagrangian function attached to this problem is

$$F(x_1, x_2, y_1, y_2) = (x_1 - 2)^2 + (x_2 + 1)^2 + y_1(x_1 - 1) + y_2(x_2 - 2)$$

and the Karush–Kuhn–Tucker conditions give us the equations

$$\begin{cases} x_1(2x_1 - 4 + y_1) = 0, \\ x_2(2x_2 + 2 + y_2) = 0, \\ y_1(x_1 - 1) = 0, \\ y_2(x_2 - 2) = 0, \end{cases} \tag{2.25}$$

and the inequalities

$$\begin{cases} 2x_1 - 4 + y_1 \geq 0, \\ 2x_2 + 2 + y_2 \geq 0, \\ 0 \leq x_1 \leq 1 \text{ and } 0 \leq x_2 \leq 2, \\ y_1, y_2 \geq 0. \end{cases} \tag{2.26}$$

The system of equations (2.25) admits 9 solutions:

$$(1, 0, 2, 0),\ (1, 2, 2, -6),\ (1, -1, 2, 0),\ (0, 0, 0, 0),\ (2, 0, 0, 0),\ (0, -1, 0, 0),$$
$$(2, -1, 0, 0),\ (0, 0, 0, -1),\ (2, 0, 0, -1),$$

of which only $(1, 0, 2, 0)$ verifies also the inequalities (2.26). Consequently,

$$\inf_{\substack{0 \leq x_1 \leq 1 \\ 0 \leq x_2 \leq 2}} f(x_1, x_2) = f(1, 0) = 2.$$

The Karush–Kuhn–Tucker conditions in the nondifferentiable setting are based on the subdifferential calculus. See [4] and [42] for details.

We next indicate a fairly general situation when the convex programming problem is equivalent to the saddle-point problem. For this we shall need the following technical result, known as Farkas' lemma:

2.7.3. Lemma. *Let $f_1, \ldots, f_m$ be convex functions defined on a nonempty convex set Y in $\mathbb{R}^n$. Then either there exists y in Y such that*

$$f_1(y) < 0, \ldots, f_m(y) < 0,$$

or there exist nonnegative numbers $a_1, \ldots, a_m$, not all zero, such that

$$a_1 f_1(y) + \cdots + a_m f_m(y) \geq 0 \quad \text{for all } y \in Y.$$

Proof. Assume that the first alternative does not work and consider the set

$$C = \Big\{(t_1, \ldots, t_m) \in \mathbb{R}^m : \text{there is } y \in Y \text{ with } f_k(y) < t_k \\ \text{for all } k = 1, \ldots, m\Big\}.$$

Then C is an open convex set that does not contain the origin of $\mathbb{R}^m$. According to Theorem 1.4.1, C and the origin can be separated by a closed hyperplane, that is, there exist scalars $a_1, \ldots, a_m$ not all zero, such that for all $y \in Y$ and all $\varepsilon_1, \ldots, \varepsilon_m > 0$,

$$a_1(f_1(y) + \varepsilon_1) + \cdots + a_m(f_m(y) + \varepsilon_m) \geq 0. \tag{2.27}$$

Keeping $\varepsilon_2, \ldots, \varepsilon_m$ fixed and letting $\varepsilon_1 \to \infty$, we infer that $a_1 \geq 0$. Similarly, $a_2 \geq 0, \ldots, a_m \geq 0$. Letting $\varepsilon_1 \to 0, \ldots, \varepsilon_m \to 0$ in (2.27) we conclude that $a_1 f_1(y) + \cdots + a_m f_m(y) \geq 0$ for all y in Y. ■

2.7.4. Theorem (Slater's condition). *Suppose that x^0 is a solution of the convex programming problem. If there exists $x^* \geq 0$ such that*

$$g_1(x^*) < 0, \ldots, g_m(x^*) < 0,$$

then one can find a y^0 in $\mathbb{R}^m$ for which (x^0, y^0) is a saddle point of the associated Lagrangian function F.

Proof. By Lemma 2.7.3, applied to the functions $g_1, \ldots, g_m, f - f(x^0)$ and the set $Y = \mathbb{R}^n_+$, we can find $a_1, \ldots, a_m, a_0 \geq 0$, not all zero, such that

$$a_1 g_1(x) + \cdots + a_m g_m(x) + a_0(f(x) - f(x^0)) \geq 0 \tag{2.28}$$

for all $x \geq 0$. A moment's reflection shows that $a_0 > 0$. Put $y_j^0 = a_j / a_0$ and $y^0 = (y_1^0, \ldots, y_m^0)$. By (2.28) we infer that

$$f(x^0) \leq f(x) + \sum_{j=1}^{m} y_j^0 g_j(x) = F(x, y^0)$$

for all $x \geq 0$. Particularly, for $x = x^0$, this yields

$$f(x^0) \leq f(x^0) + \sum_{j=1}^{m} y_j^0 g_j(x^0) \leq f(x^0)$$

that is, $\sum_{j=1}^{m} y_j^0 g_j(x^0) = 0$, whence

$$F(x^0, y^0) = f(x^0) \leq F(x, y^0)$$

for all $x \geq 0$. On the other hand, for $y \geq 0$ we have

$$F(x^0, y^0) = f(x^0) \geq f(x^0) + \sum_{j=1}^{m} y_j g_j(x^0) = F(x^0, y),$$

so that (x^0, y^0) is a saddle point. ■

We end this section with a nice geometric application of quadratic programming, which was noted by J. Franklin [17], in his beautiful introduction to mathematical methods of economics. It is about a problem of J. Sylvester, requiring the least circle which contains a given set of points in the plane.

Suppose the given points are $a_1, \dots, a_m$. They lie inside the circle of center x and radius r if

$$\|a_k - x\|^2 \leq r^2 \quad \text{for } k = 1, \dots, m. \tag{2.29}$$

We want to find x and r so as to minimize r. Letting

$$x_0 = \frac{1}{2}(r^2 - \|x\|^2),$$

we can replace the quadratic constraints (2.29) by linear ones,

$$x_0 + \langle a_k, x \rangle \geq b_k \quad \text{for } k = 1, \dots, m.$$

Here $b_k = \|a_k\|^2/2$. In this way, Sylvester's problem becomes a problem of quadratic programming,

$$\text{minimize } (2x_0 + x_1^2 + x_2^2),$$

subject to the m linear inequalities

$$x_0 + a_{k1}x_1 + a_{k2}x_2 \geq b_k \quad (k = 1, \dots, m).$$

Chapter 3

The Variational Approach of PDE

The aim of this chapter is to illustrate a number of problems in partial differential equations (PDE) which can be solved by seeking a global minimum of suitable convex functionals. This idea goes back to advanced calculus. See the comments at the end of Section 2.6.

3.1 The Minimum of Convex Functionals

The main criterion for the existence and uniqueness of global minimum of convex functions is actually a far reaching generalization of the orthogonal projection:

3.1.1. Theorem. *Let C be a closed convex set in a reflexive Banach space V and let $J\colon C \to \mathbb{R}$ be a convex function such that:*

(a) *J is weakly lower semicontinuous, that is,*

$$u_n \to u \text{ weakly in } V \quad \text{implies} \quad J(u) \leq \liminf_{n\to\infty} J(u_n);$$

(b) *Either C is bounded, or $\lim_{\|u\|\to\infty} J(u) = \infty$.*

Then J admits at least one global minimum and the points of global minimum constitute a convex set.

If, moreover, J is strictly convex, then there is a unique global minimum.

Proof. Put

$$m = \inf_{u\in C} J(u).$$

Clearly, $m < \infty$, and there exists a sequence $(u_n)_n$ of elements in C such that $J(u_n) \to m$. By our hypotheses, the sequence $(u_n)_n$ is bounded, so by Theorem 1.3.8, we may assume (replacing $(u_n)_n$ by a subsequence) that it is also weakly converging to an element u in C. Here we used the fact that C is weakly closed (which is a consequence of Corollary 1.4.4). Then

$$m \le J(u) \le \liminf_{n\to\infty} J(u_n) = m,$$

and thus u is a global minimum. The remainder of the proof is left to the reader as an exercise. ■

In the differentiable case we state the following useful version of Theorem 3.1.1:

3.1.2. Theorem. *Let V be a reflexive Banach space and let $J\colon V \to \mathbb{R}$ be a Gâteaux differentiable convex functional with the following properties*:

(a) *For each $u \in V$, the map $J'(u)\colon v \to J'(u;v)$ is an element of V'*;

(b) $\lim_{\|u\|\to\infty} J(u) = \infty$.

Then J admits at least one global minimum and the points of global minimum are precisely the points u such that

$$J'(u;v) = 0 \quad \text{for all } v \in V.$$

If, moreover, J is strictly convex, then there is a unique global minimum.

Proof. First notice that J is weakly lower semicontinuous. In fact, by Theorem 2.6.1,

$$J(u_n) \ge J(u) + J'(u; u_n - u)$$

for all n, while $J'(u;u_n - u) = J'(u)(u_n - u) \to 0$ by our hypotheses. Hence, according to Theorem 3.1.1, J admits global minima.

If u is a global minimum, then for each $v \in V$ there is a $\delta > 0$ such that

$$\frac{J(u+\varepsilon v) - J(u)}{\varepsilon} \ge 0 \quad \text{whenever } |\varepsilon| < \delta.$$

This yields $J'(u;v) \ge 0$. Replacing v by $-v$, we obtain

$$-J'(u;v) = J'(u;-v) \ge 0,$$

and thus $J'(u;v) = 0$. Conversely, if $J'(u;v) = 0$ for all $v \in V$, then by Theorem 2.6.1 we get

$$J(v) \ge J(u) + J'(u, v-u) = J(u),$$

that is, u is a global minimum. ■

Typically, Theorem 3.1.1 applies to functionals of the form

$$J(u) = \frac{1}{2}\|u - w\|^2 + \varphi(u), \quad u \in V,$$

where V is an L^p-space with $p \in (1,\infty)$, w is an arbitrary fixed element of V and $\varphi\colon V \to \mathbb{R}$ is a weakly lower semicontinuous convex function. Theorem 3.1.2 covers a large range of well-behaved convex functionals, with important consequences to the problem of existence of solutions of partial differential equations:

3.1.3. Corollary. *Let Ω be a nonempty open set in $\mathbb{R}^n$ and let $p > 1$. Consider a function $g \in C^1(\mathbb{R})$ which verifies the following properties*:

(a) $g(0) = 0$ *and* $g(t) \geq \alpha |t|^p$ *for a suitable constant* $\alpha > 0$;

(b) *The derivative* g' *is increasing and* $|g'(t)| \leq \beta |t|^{p-1}$ *for a suitable constant* $\beta > 0$. *Then the linear space* $V = L^p(\Omega) \cap L^2(\Omega)$ *is reflexive when endowed with the norm*

$$\|u\|_V = \|u\|_{L^p} + \|u\|_{L^2},$$

and for all $f \in L^2(\Omega)$ *the functional*

$$J(u) = \int_\Omega g(u(x))\,dx + \frac{1}{2}\int_\Omega |u(x)|^2\,dx + \int_\Omega f(x)u(x)\,dx, \quad u \in V$$

is convex and Gâteaux differentiable with

$$J'(u;v) = \int_\Omega g'(u(x))v(x)\,dx + \int_\Omega u(x)v(x)\,dx + \int_\Omega f(x)v(x)\,dx.$$

Moreover, J *admits a unique global minimum* $\bar{u}$, *which is the solution of the equation*

$$J'(u;v) = 0 \quad \textit{for all } v \in V.$$

Proof. V is a closed subspace of $L^2(\Omega)$ and thus it is a reflexive space. Then notice that

$$\begin{aligned} |g(t)| &= |g(t) - g(0)| \\ &= \left| \int_0^t g'(s)\,ds \right| \leq \frac{\beta}{p}\,|t|^p, \end{aligned}$$

from which it follows easily that J is well defined. Letting

$$J_1(u) = \int_\Omega g(u(x))\,dx,$$

by Lagrange's mean value theorem,

$$\begin{aligned} J_1(u + tv) &= \int_\Omega g(u(x) + tv(x))\,dx \\ &= \int_\Omega g(u(x))\,dx + t\int_\Omega g'(u(x) + \tau(x)v(x))v(x)\,dx, \end{aligned}$$

where $0 < \tau(x) < t$ for all x, provided that $t > 0$. Then

$$\frac{J_1(u + tv) - J_1(u)}{t} = \int_\Omega g'(u(x) + \tau(x)v(x))v(x)\,dx,$$

and letting $t \to 0+$ we get the desired formula for $J'(u;v)$.

Again by Lagrange's mean value theorem, and the fact that g' is increasing, we have

$$\begin{aligned} J_1(v) &= J_1(u) + \int_\Omega [g'(u(x) + \tau(x)(v(x) - u(x))) \cdot (v(x) - u(x))]\, dx \\ &\geq J_1(u) + \int_\Omega g'(u(x)) \cdot (v(x) - u(x))\, dx \\ &= J_1(u) + J_1'(u, v - u), \end{aligned}$$

which shows that J_1 is convex. Then the functional J is the sum of a convex function and a strictly convex function.

Finally,

$$\begin{aligned} J(u) &\geq \alpha \int_\Omega |u(x)|^p\, dx + \frac{1}{2} \int_\Omega |u(x)|^2\, dx - \left| \int_\Omega f(x)u(x)\, dx \right| \\ &\geq \alpha \|u\|_{L^p}^p + \frac{1}{2} \|u\|_{L^2}^2 - \|f\|_{L^2} \|u\|_{L^2}, \end{aligned}$$

from which it follows that

$$\lim_{\|u\|_V \to \infty} J(u) = \infty,$$

and the conclusion follows from Theorem 3.1.2. ■

The result of Corollary 3.1.3 extends (with obvious changes) to the case where V is the space of all $u \in L^2(\Omega)$ such that $Au \in L^p(\Omega)$ for a given linear differential operator A. Also, we can consider finitely many functions g_k (verifying the conditions (a) and (b) for different exponents $p_k > 1$) and finitely many linear differential operators A_k. In that case we shall deal with the functional

$$J(u) = \sum_{k=1}^m \int_\Omega g_k(A_k u)\, dx + \frac{1}{2} \int_\Omega |u|^2\, dx + \int_\Omega f u\, dx,$$

defined on $V = \bigcap_{k=1}^m L^{p_k}(\Omega) \cap L^2(\Omega)$; V is reflexive when endowed with the norm

$$\|u\|_V = \sum_{k=1}^m \|A_k u\|_{L^{p_k}} + \|u\|_{L^2}.$$

3.2 Preliminaries on Sobolev Spaces

Some basic results on Sobolev spaces are recalled here for the convenience of the reader. The details are available from many sources, including [2], [14], [40] and [46].

Let Ω be a bounded open set in $\mathbb{R}^n$ with Lipschitz boundary $\partial\Omega$, and let m be a positive integer.

The Sobolev space $H^m(\Omega)$ consists of all functions $u \in L^2(\Omega)$ which admit *weak derivatives* $D^\alpha u$ in $L^2(\Omega)$, for all multi-indices α with $|\alpha| \leq m$. This means the existence of functions $v_\alpha \in L^2(\Omega)$ such that

$$\int_\Omega v_\alpha \cdot \varphi \, dx = (-1)^{|\alpha|} \int_\Omega u \cdot D^\alpha \varphi \, dx \tag{3.1}$$

for all φ in the space $C_c^\infty(\Omega)$ and all α with $|\alpha| \leq m$. Due to the denseness of $C_c^\infty(\Omega)$ in $L^2(\Omega)$, the functions v_α are uniquely defined by (3.1), and they are usually denoted as $D^\alpha u$.

One can prove easily that $H^m(\Omega)$ is a Hilbert space when endowed with the norm $\|\cdot\|_{H^m}$ associated to the inner product

$$\langle u, v\rangle_{H^m} = \sum_{|\alpha| \leq m} \int_\Omega D^\alpha u \cdot D^\alpha v \, dx.$$

Notice that $C^m(\overline{\Omega})$ is a dense subspace of $H^m(\Omega)$.

3.2.1. Theorem (The trace theorem). *There is a continuous linear operator*

$$\gamma = (\gamma_0, \ldots, \gamma_{m-1})\colon H^m(\Omega) \to L^2(\partial\Omega)^m$$

such that

$$\gamma_0 u = u|_{\partial\Omega}, \quad \gamma_1 u = \frac{\partial u}{\partial n}, \ldots, \gamma_{m-1} u = \frac{\partial^{m-1} u}{\partial n^{m-1}}$$

for all u in $C^m(\overline{\Omega})$.

The closure of $C_c^\infty(\Omega)$ in $H^m(\Omega)$ is the Sobolev space $H_0^m(\Omega)$. This space coincides with the kernel of the trace operator γ, indicated in Theorem 3.2.1.

On $H_0^1(\Omega)$, the norm $\|\cdot\|_{H^1}$ can be replaced by an equivalent norm,

$$\|u\|_{H_0^1} = \left(\int_\Omega \|\nabla u\|^2 \, dx\right)^{1/2}.$$

In fact, there exists a constant $c > 0$ such that

$$\|u\|_{H_0^1} \leq \|u\|_{H^1} \leq c\|u\|_{H_0^1} \quad \text{for all } u \in H_0^1(\Omega).$$

This is a consequence of a basic inequality in partial differential equations:

3.2.2. Theorem (Poincaré's inequality). *If Ω is a bounded open subset of $\mathbb{R}^n$, then there exists a constant $C > 0$ such that*

$$\|u\|_{L^2} \leq C \left(\int_\Omega \|\nabla u\|^2 \, dx\right)^{1/2}$$

for all $u \in H_0^1(\Omega)$.

Proof. Since $C_c^\infty(\Omega)$ is dense into $H_0^1(\Omega)$, it suffices to prove Poincaré's inequality for functions $u \in C_c^\infty(\Omega) \subset C_c^\infty(\mathbb{R}^n)$. The fact that Ω is bounded, yields two real numbers a and b such that

$$\Omega \subset \{x = (x', x_n) \in \mathbb{R}^{n-1} \times \mathbb{R} \mid a \leq x_n \leq b\}.$$

We have

$$u(x', x_n) = \int_a^{x_n} \frac{\partial u}{\partial x_n}(x', t)\, dt,$$

and an application of the Cauchy–Buniakovski–Schwarz inequality gives us

$$\begin{aligned} |u(x', x_n)|^2 &\leq (x_n - a) \int_a^{x_n} \left| \frac{\partial u}{\partial x_n}(x', t) \right|^2 dt \\ &\leq (x_n - a) \int_{\mathbb{R}} \left| \frac{\partial u}{\partial x_n}(x', t) \right|^2 dt. \end{aligned}$$

Then

$$\int_{\mathbb{R}^{n-1}} |u(x', t)|^2\, dx' \leq (x_n - a) \int_{\mathbb{R}^n} \left| \frac{\partial u}{\partial x_n}(x) \right|^2 dx,$$

which leads to

$$\int_{\mathbb{R}^n} |u(x)|^2\, dx = \int_a^b \int_{\mathbb{R}^{n-1}} |u(x', t)|^2\, dx' dt \leq \frac{(b-a)^2}{2} \int_{\mathbb{R}^n} \left| \frac{\partial u}{\partial x_n}(x) \right|^2 dx$$

and now the assertion of Theorem 3.2.2 is clear. ■

By Poincaré's inequality, the inclusion $H_0^m(\Omega) \subset H^m(\Omega)$ is strict whenever Ω is bounded. Notice that $H_0^m(\mathbb{R}^n) = H^m(\mathbb{R}^n)$, due to the possibility to approximate (via mollification) the functions in $H^m(\mathbb{R}^n)$ by functions in $C_c^\infty(\mathbb{R}^n)$.

3.3 Applications to Elliptic Boundary-Value Problems

In what follows we shall illustrate the role of the variational methods in solving some problems in partial differential equations. More advanced applications may be found in books like those by G. Duvaut and J.-L. Lions [12] and I. Ekeland and R. Temam [13].

3.3.1. Dirichlet Problems. Let Ω be a bounded open set in $\mathbb{R}^n$ and let $f \in C(\overline{\Omega})$. A function $u \in C^2(\Omega) \cap C(\overline{\Omega})$ is said to be a *classical solution* of the Dirichlet problem

$$\begin{cases} -\Delta u + u = f & \text{in } \Omega \\ u = 0 & \text{on } \partial\Omega\,, \end{cases} \tag{3.2}$$

provided that it satisfies the equation and the boundary condition pointwise.

If u is a classical solution to this problem then the equation $-\Delta u + u = f$ is equivalent to

$$\int_\Omega (-\Delta u + u) \cdot v \, dx = \int_\Omega f \cdot v \, dx \quad \text{for all } v \in H_0^1(\Omega).$$

By Green's formula,

$$\int_\Omega (-\Delta u + u) \cdot v \, dx = -\int_{\partial\Omega} \frac{\partial u}{\partial n} \cdot v \, dx + \int_\Omega u \cdot v \, dx + \sum_{k=1}^n \int_\Omega \frac{\partial u}{\partial x_k} \cdot \frac{\partial v}{\partial x_k} \, dx,$$

so that we arrive at the following restatement of (3.2):

$$\sum_{k=1}^n \int_\Omega \frac{\partial u}{\partial x_k} \cdot \frac{\partial v}{\partial x_k} \, dx + \int_\Omega u \cdot v \, dx = \int_\Omega f \cdot v \, dx \tag{3.3}$$

for all $v \in C_c^\infty(\Omega)$. It turns out that (3.3) makes sense for $u \in H_0^1(\Omega)$ and $f \in L^2(\Omega)$. We shall say that a function $u \in H_0^1(\Omega)$ is a *weak solution* for the Dirichlet problem (3.2) with $f \in L^2(\Omega)$ if it satisfies (3.3) for all $v \in H_0^1(\Omega)$.

The existence and uniqueness of the weak solution for the Dirichlet problem (3.2) follows from Theorem 3.1.2, applied to the functional

$$J(u) = \frac{1}{2} \|u\|_{H_0^1}^2 - \langle f, u \rangle_{L^2}, \quad u \in H_0^1(\Omega).$$

In fact, this functional is strictly convex and twice Gâteaux differentiable, with

$$\begin{aligned} J'(u; v) &= \langle u, v \rangle_{H_0^1} - \langle f, v \rangle_{L^2} \\ J''(u; v, w) &= \langle w, v \rangle_{H_0^1}. \end{aligned}$$

According to Theorem 3.1.2, the unique point of global minimum of J is the unique solution of the equation

$$J'(u; v) = 0 \quad \text{for all } v \in H_0^1(\Omega),$$

and clearly, the latter is equivalent with (3.3).

3.3.2. Neumann Problems. Let Ω be a bounded open set in $\mathbb{R}^n$ (with Lipschitz boundary) and let $f \in C(\overline{\Omega})$. A function $u \in C^2(\Omega) \cap C^1(\overline{\Omega})$ is said to be a *classical solution* of the Neumann problem

$$\begin{cases} -\Delta u + u = f & \text{in } \Omega \\ \dfrac{\partial u}{\partial n} = 0 & \text{on } \partial\Omega, \end{cases} \tag{3.4}$$

provided that it satisfies the equation and the boundary condition pointwise.

If u is a classical solution to this problem, then the equation $-\Delta u + u = f$ is equivalent to

$$\int_\Omega (-\Delta u + u) \cdot v \, dx = \int_\Omega f \cdot v \, dx \quad \text{for all } v \in H^1(\Omega\,),$$

and thus with

$$\sum_{k=1}^{n} \int_\Omega \frac{\partial u}{\partial x_k} \cdot \frac{\partial v}{\partial x_k} \, dx + \int_\Omega u \cdot v \, dx = \int_\Omega f \cdot v \, dx \quad \text{for all } v \in H^1(\Omega\,), \tag{3.5}$$

taking into account Green's formula and the boundary condition $\frac{\partial u}{\partial n} = 0$ on $\partial\Omega$. As in the case of Dirichlet problem, we can introduce a concept of a weak solution for the Neumann problem (3.4) with $f \in L^2(\Omega\,)$. We shall say that a function $u \in H^1(\Omega\,)$ is a *weak solution* for the problem (3.4) if it satisfies (3.5) for all $v \in H^1(\Omega\,)$.

The existence and uniqueness of the weak solution for the Neumann problem follows from Theorem 3.1.2, applied to the functional

$$J(u) = \frac{1}{2} \, \|u\|_{H^1}^2 - \langle f, u \rangle_{L^2}, \quad u \in H^1(\Omega\,).$$

The details are similar to the above case of Dirichlet problem.

Corollary 3.1.3, and its generalization to finite families of functions g, allow us to prove the existence and uniqueness of considerably more subtle Neumann problems such as

$$\begin{cases} -\Delta u + u + u^3 = f & \text{in } \Omega \\ \dfrac{\partial u}{\partial n} = 0 & \text{on } \partial\Omega\,, \end{cases} \tag{3.6}$$

where $f \in L^2(\Omega\,)$. This corresponds to the case where

$$\begin{gathered} g_1(t) = \cdots = g_n(t) = t^2/2, \quad g_{n+1}(t) = t^4/4, \\ A_k u = \partial u/\partial x_k \text{ for } k = 1, \ldots, n, \quad A_{n+1} u = u, \\ p_1 = \cdots = p_n = 2, \; p_{n+1} = 4, \end{gathered}$$

and

$$J(u) = \frac{1}{2} \, \|u\|_{H^1}^2 + \frac{1}{4} \, \|u\|_{L^4}^4 - \langle f, u \rangle_{L^2}, \quad u \in V = H^1(\Omega\,) \cap L^4(\Omega\,).$$

According to Corollary 3.1.3, there is a unique global minimum of J and this is done by the equation

$$J'(u; v) = 0 \quad \text{for all } v \in V,$$

that is, by

$$\sum_{k=1}^{n} \int_\Omega \frac{\partial u}{\partial x_k} \cdot \frac{\partial v}{\partial x_k} \, dx + \int_\Omega u \cdot v \, dx + \int_\Omega u^3 \cdot v \, dx = \int_\Omega f \cdot v \, dx$$

for all $v \in V$. Notice that the latter equation represents the weak form of (3.6).

The conditions under which weak solutions provide classical solutions are discussed in textbooks like that by M. Renardy and R. C. Rogers [40].

3.4 The Galerkin Method

It is important to give here an idea how the global minimum of convex functionals can be determined via numerical algorithms. For this, consider a reflexive real Banach space V, with Schauder basis $(e_k)_k$. This means that every $u \in V$ admits a unique representation

$$u = \sum_{k=1}^{\infty} c_k e_k$$

with $c_k \in \mathbb{R}$, the convergence being in the norm topology. As a consequence, for each $n \in \mathbb{N}^\star$ there is a linear projection

$$P_n \colon V \to V, \quad P_n u = \sum_{k=1}^{n} c_k e_k.$$

Since $P_n u \to u$ for every u, the Banach–Steinhaus theorem in functional analysis assures that $\sup \|P_n\| < \infty$.

Consider a functional $J \colon V \to \mathbb{R}$ which is twice Gâteaux differentiable and for each $u \in V$ there exist $\nabla J(u) \in V'$ and $H(u) \in L(V, V')$ such that

$$J'(u; v) = \langle \nabla J(u), v \rangle$$
$$J''(u; v, w) = \langle H(u)v, w \rangle$$

for all $u, v, w \in V$. In addition, we assume that $H(u)$ satisfies estimates of the form:

$$\begin{cases} |\langle H(u)v, w \rangle| \leq M\|v\| \, \|w\| \\ \langle H(u)v, v \rangle \geq \alpha \|v\|^2 \end{cases} \tag{3.7}$$

for all $u, v, w \in V$. Here M and α are positive constants.

By Taylor's formula, J is strictly convex and $\lim_{\|u\| \to \infty} J(u) = \infty$. Then J is weakly lower semicontinuous, so by Theorem 3.1.2 it admits a unique global minimum.

In the *Galerkin method*, the global minimum u of J is found by a finite dimensional approximation process. More precisely, one considers the restriction of J to $V_n = \text{Span}\,\{e_1, \ldots, e_n\}$ and one computes the global minimum u_n of this restriction by solving the equation

$$\langle \nabla J(u_n), v \rangle = 0 \quad \text{for all } v \in V_n.$$

The existence of u_n follows again from Theorem 3.1.2. Remarkably, these minimum points approximate the global minimum u in the following strong way:

3.4.1. Theorem. *We have*

$$\lim_{n \to \infty} \|u_n - u\| = 0.$$

Proof. Letting $v_n = P_n u$, we know that $v_n \to u$. By Taylor's formula, for each n there is a $\lambda_n \in (0,1)$ such that

$$J(v_n) = J(u) + \langle \nabla J(u), v_n - u \rangle + \frac{1}{2} \langle H(u + \lambda_n(v_n - u))(v_n - u), v_n - u \rangle.$$

Combining this with the first estimate in (3.7), we get $J(v_n) \to J(u)$. By the choice of u_n, it yields that

$$J(u) \leq J(u_n) \leq J(v_n),$$

so that $J(u_n) \to J(u)$ too. Also, $\sup J(u_n) < \infty$.

Since $\lim_{\|u\| \to \infty} J(u) = \infty$, we deduce that the sequence $(u_n)_n$ is norm bounded. According to Theorem 1.3.8, it follows that $(u_n)_n$ has a weakly converging subsequence, say $u_{k(n)} \xrightarrow{w} u'$. Since J is lower semicontinuous, we have

$$J(u') \leq \liminf_{n \to \infty} J(u_{k(n)}) \leq J(u),$$

from which it follows that $u' = u$ and $u_n \xrightarrow{w} u$. Again by Taylor's formula, for each n there is a $\mu_n \in (0,1)$ such that

$$J(u_n) = J(u) + \langle \nabla J(u), u_n - u \rangle + \frac{1}{2} \langle H(u + \mu_n(u_n - u))(u_n - u), u_n - u \rangle.$$

This relation, when combined with the second estimate in (3.7), leads to

$$\frac{2}{\alpha} \|u_n - u\|^2 \leq |J(u_n) - J(u)| + |\langle \nabla J(u), u_n - u \rangle|$$

and the conclusion of the theorem is now obvious. ■

Bibliography

[1] G. Alberti and L. Ambrosio, *A geometrical approach to monotone functions in* $\mathbb{R}^n$, Math. Z. **230** (1999), 259–316.

[2] H. W. Alt, *Lineare Funktionalanalysis*, 2 Auflage, Springer-Verlag, Berlin, 1992.

[3] V. Barbu and Th. Precupanu, *Convexity and Optimization in Banach Spaces*, Ed. Academiei, Bucharest, and D. Reidel Publishing Company, Dordrecht, 1986.

[4] J. M. Borwein and A. S. Lewis, *Convex Analysis and Nonlinear Optimization. Theory and Examples*, Springer-Verlag, Berlin, 2000.

[5] J. Céa, *Optimisation. Théorie et Algorithmes*, Dunod, Paris, 1971.

[6] G. Choquet, *Les cônes convexes faiblement complets dans l'analyse*, Proc. Intern. Congr. Mathematicians, Stockholm, 1962, 317–330.

[7] F. H. Clarke, *Optimization and Nonsmooth Analysis*, 2nd edition, Classics in Applied Mathematics, Vol. 5, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1990.

[8] J. B. Conway, *A Course in Functional Analysis*, 2nd edition, Springer-Verlag, Berlin, 1997.

[9] M. G. Crandall, H. Ishii and P.-L. Lions, *User's guide to viscosity solutions of second order partial differential equations*, Bull. Amer. Math. Soc. **27** (1992), 1–67.

[10] M. M. Day, *Normed linear spaces*, 3rd edition, Springer-Verlag, Berlin, 1973.

[11] J. Dieudonné, *Foundations of Modern Analysis*, Academic Press, 1960.

[12] G. Duvaut and J.-L. Lions, *Inequalities in Mechanics and Physics*, Springer-Verlag, Berlin, 1976.

[13] I. Ekeland and R. Temam, *Convex Analysis and Variational Problems*, North-Holland Publishing Company, Amsterdam, 1976.

[14] L. C. Evans and R. F. Gariepy, *Measure Theory and Fine Properties of Functions*, CRC Press, Boca Raton–New York–London–Tokyo, 1992.

[15] H. Federer, *Geometric Measure Theory*, Springer-Verlag, New York, 1969.

[16] W. Fenchel, *On conjugate convex functions*, Canad. J. Math. **1** (1949), 73–77.

[17] J. Franklin, *Mathematical Methods of Economics*, Amer. Math. Monthly, **90** (1983), 229-244.

[18] R. J. Gardner, *The Brunn–Minkowski inequality*, Bull. Amer. Math. Soc. **39** (2002), 355–405. See also http://www.ac.wwu.edu/~gardner/research.html

[19] P. M. Gruber, *Aspects of convexity and its applications*, Expo. Math. **2** (1984), 47–83.

[20] P. M. Gruber and J. M. Willis (eds.), *Handbook of convex geometry*, North-Holland, Amsterdam, 1993.

[21] G. H. Hardy, J. E. Littlewood and G. Pólya, *Inequalities*, 2nd edition, Cambridge Mathematical Library, 1952; reprinted 1988.

[22] E. Hewitt and K. Stromberg, *Real and Abstract Analysis*, Springer-Verlag, Berlin, 1965.

[23] J.-B. Hiriart-Urruty, *Ensembles de Tchebychev vs. ensembles convexes: l'état de la situation vu via l'analyse non lisse*, Ann. Sci. Math. Québec **22** (1998), 47–62.

[24] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms*, Springer-Verlag, Berlin, 1993.

[25] L. Hörmander, *Sur la fonction d'appui des ensembles convexes dans une espace localement convexe*, Ark. Mat. **3** (1954), 181–186.

[26] L. Hörmander, *Notions of Convexity*, Birkhäuser, Boston, 1994.

[27] J. L. W. V. Jensen, *Om konvexe Funktioner og Uligheder mellem Middelvaerdier*, Nyt. Tidsskr. Math. **16B** (1905), 49–69.

[28] J. L. W. V. Jensen, *Sur les fonctions convexes et les inegalités entre les valeurs moyennes*, Acta Math. **30** (1906), 175–193.

[29] B. Josefson, *Weak sequential convergence in the dual of a Banach space does not imply norm convergence*, Ark. Mat. **13** (1975), 79–89.

[30] V. Klee, *Some new results on smoothness and rotundity in normed linear spaces*, Math. Ann. **139** (1959), 51–63.

[31] S. Lang, *Analysis* II, Addison-Wesley Publ. Co., Reading, Massachusetts, 1969.

[32] J. M. Lasry and P.-L. Lions, *A remark on regularization in Hilbert spaces*, Israel J. Math. **55** (1986), 257–266.

[33] E. H. Lieb and M. Loss, *Analysis*, 2nd edition, Amer. Math. Soc., Providence, R.I., 2001.

[34] G. Minty, *Monotone nonlinear operators on a Hilbert space*, Duke Math. J. **29** (1962), 341–346.

[35] C. P. Niculescu and L.-E. Persson, *Convex Functions and Their Applications*, Springer-Verlag, 2006.

[36] A. Nissenzweig, *$w^\star$-sequential convergence*, Israel J. Math. **22** (1975), 266–272.

[37] J. E. Pečarić, F. Proschan and Y. C. Tong, *Convex functions, partial orderings and statistical applications*, Academic Press, New York, 1992.

[38] R. R. Phelps, *Convex Functions, Monotone Operators, and Differentiability*, 2nd edition, Lecture Notes in Math. **1364**, Springer-Verlag, Berlin, 1993.

[39] R. R. Phelps, *Lectures on Choquet's Theorem*, 2nd edition, Lecture Notes in Math. **1757**, Springer-Verlag, Berlin, 2001.

[40] M. Renardy and R. C. Rogers, *An Introduction to Partial Differential Equations*, Springer-Verlag, Berlin, 1993.

[41] A. W. Roberts and D. E. Varberg, *Convex Functions*, Academic Press, New York and London, 1973.

[42] R. T. Rockafellar, *Convex Analysis*, Princeton Mathematical Series **28**, Princeton Univ. Press, Princeton, New Jersey, 1970.

[43] P. Roselli and M. Willem, *A convexity inequality*, Amer. Math. Monthly **109** (2002), 64–70.

[44] C. Villani, *Topics in Optimal Transportation*, Graduate Studies in Math. **58**, Amer. Math. Soc., Providence, R.I., 2003.

[45] R. Webster, *Convexity*, Oxford Univ. Press, Oxford–New York–Tokyo, 1994.

[46] W. P. Ziemer, *Weakly Differentiable Functions, Sobolev Spaces and Functions of Bounded Variation*, Graduate Texts in Math. **120**, Springer-Verlag, Berlin, 1989.